



HAL
open science

Actes de la conférence SAGEO'2016 - Spatial Analysis and GEomatics

Karine Emsellem, Diego Moreno, Christine Voiron, Didier Josselin

► **To cite this version:**

Karine Emsellem, Diego Moreno, Christine Voiron, Didier Josselin. Actes de la conférence SAGEO'2016 - Spatial Analysis and GEomatics. Karine Emsellem; Didier Josselin; Diego Moreno Sierra; Christine Voiron. SAGEO'2016, Dec 2016, Nice, France. , 600 p., 2016, 2-910545-11-3. hal-01424734

HAL Id: hal-01424734

<https://hal.science/hal-01424734>

Submitted on 19 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACTES DE LA CONFÉRENCE



SAGEO 2016

SPATIAL ANALYSIS AND GEOMATICS

NICE, 6-9/12/2016



Avant-propos

SAGEO est la cordelette qui noue les domaines, les approches, les disciplines et les acteurs qui font la géomatique. Conférence internationale annuelle à l'acronyme explicite (Spatial Analysis and GEOmatics), portée par le GDR MAGIS (Méthodes et Applications pour la géomatique et l'information spatiale), et cette année par l'UMR ESPACE (Etude des Structures, des Processus d'Adaptation et des Changements de l'Espace), SAGEO a pour objectifs de

- Présenter les recherches récentes des domaines de la Géomatique, de l'Analyse Spatiale, et de l'Information Géographique
- Réunir les communautés de chercheurs venant de disciplines diverses, et permettre les échanges à un niveau national et international
- Permettre d'accroître les relations entre les chercheurs spécialistes de ces domaines, et les organismes publics et privés

La douzième édition de la conférence SAGEO s'est déroulée à Nice, du 6 au 9 décembre 2016. Elle a rassemblé plus de 130 participants. Les articles rassemblés dans ces actes témoignent à nouveau, à travers un programme riche et varié, de la pluralité des problématiques, mais aussi de l'actualité et des enjeux des questionnements traités. 28 articles longs ont été présentés lors de la conférence, et 26 posters ont été discutés. Ils ont été répartis dans les 6 sessions thématiques plénières :

- Géovisualisation, cartographie, exploration des dynamiques spatio-temporelles
- Infrastructures de données spatiales, observatoires, géographie participative, géoprospective
- Dynamiques urbaines, accessibilité
- Mobilités, flux, trajectoires, modélisation spatio-temporelle
- Risques, changement climatique
- Environnement, Traitement d'image, résolution spatiale

Trois conférenciers invités de renom ont partagé leurs expériences et réflexions sur quelques enjeux importants : Florence Sedès (Institut de Recherche en Informatique de Toulouse), Konstadinos G. Goulias (University of California Santa Barbara) et Noamen Rebai (Ecole nationale d'Ingénieurs de Tunis). Les résumés de leurs interventions ont été insérés dans ces actes.

Enfin, 5 ateliers de groupes-projets du GDR Magis ont discuté de leurs activités.

Nous tenons à remercier l'ensemble des membres du comité scientifique pour l'évaluation des soumissions qui permettent d'atteindre la qualité reconnue de cette manifestation. Nous remercions aussi tous les auteurs qui ont proposé leurs travaux à cette conférence. Enfin, sans l'efficacité du comité d'organisation, la préparation de SAGEO 2016 n'aurait pas été possible : merci à lui !

La prochaine édition de SAGEO 2017 aura lieu à Rouen du 6 au 9 novembre 2017. Au plaisir de vous y retrouver !

Le comité d'organisation SAGEO 2016

La direction du GDR Magis

ALI FRIHIDA, *France*
ANDRIEU JULIEN, *France*
BAAZAOUI HAJER, *Tunisie*
BATTON-HUBERT MIREILLE, *France*
BAVAUD FRANÇOIS, *Suisse*
BECK ELISE, *France*
BILLEN ROLAND, *Belgique*
BIMONTE SANDRO, *France*
BOUCELMA OMAR, *France*
BOUJU ALAIN, *France*
BOUVERET SYLVAIN, *France*
BROSSET DAVID, *France*
BUCHER BÉNÉDICTE, *France*
CHANET JEAN-PIERRE, *France*
CHARDONNEL SONIA, *France*
CLARAMUNT CHRISTOPHE, *France*
DAVOINE PAULE-ANNICK, *France*
DE RUFFRAY SOPHIE, *France*
DE SÈDE-MARCEAU MARIE-HÉLÈNE, *France*
DEL MONDO GÉRALDINE, *France*
DESJARDIN ÉRIC, *France*
DEVILLERS RODOLPHE, *France*
DEVOGELE THOMAS, *France*
DONNAY JEAN-PAUL, *Belgique*
DOUVINET JOHNNY, *France*
EMSELLEM KARINE, *France*
FAIZ SAMI, *Tunisie*
FERDINANDO SEMBOLONI, *Italie*
FEYT GRÉGOIRE, *France*
FUSCO GIOVANNI, *France*
GAIO MAURO, *France*
GANÇARSKI PIERRE, *France*
GENRE-GRANDPIERRE CYRILLE, *France*
GENSEL JÉRÔME, *France*
GESQUIÈRE GILLES, *France*
GOURMELON FRANÇOISE, *France*
GUILBERT ERIC, *France*
HADDAD MOHAMED RAMZI, *Tunisie*
HOARAU CHARLOTTE, *France*
JOLIVEAU THIERRY, *France*
JOSSELIN DIDIER, *France*
LARDON SYLVIE, *France*
LE BER FLORENCE, *France*
LEIBOVICI DIDIER, *Royaume-Uni*
LIBOUREL THERESE, *France*
LOUBIER JEAN-CHRISTOPHE, *France*
MALDONADO ERIC, *France*
MATHIAN HÉLÈNE, *France*
MORENO SIERRA DIEGO, *France*
MUSTIERE SÉBASTIEN, *France*
NOUCHER MATTHIEU, *France*
PINET FRANÇOIS, *France*
PUISSANT ANNE, *France*
RAY CYRIL, *France*
RUAS ANNE, *France*
SALLABERRY CHRISTIAN, *France*
SANDERS LENA, *France*
SEDES FLORENCE, *France*
SERRE DAMIEN, *France*
SERVIGNE SYLVIE, *France*
SERVIÈRES MYRIAM, *France*
SHEEREN DAVID, *France*
TACNET JEAN-MARC, *France*
THIBAUD RÉMY, *France*
TOURRE VINCENT, *France*
VILLANOVA-OLIVER MARLÈNE, *France*
VOIRON-CANICIO CHRISTINE, *France*
ZEITOUNI KARINE, *France*
ZIÉBELIN DANIELLE, *France*

COMITÉ D'ORGANISATION

BRACHET NATHALIE,
Assistant Ingénieur Édition et Communication, CNRS, Université d'Avignon et des pays du Vaucluse, ESPACE

DEGREVE D'AMICO SYLVIE,
Assistante de direction, Université de Nice Sophia-Antipolis, ESPACE

EMSELLEM KARINE,
Maître de Conférence en Géographie, Université de Nice Sophia-Antipolis, ESPACE

GIBELLO VÉRONIQUE,
Secrétaire Gestionnaire, Université de Nice Sophia-Antipolis, ESPACE

JOSELIN DIDIER,
Directeur de Recherches, CNRS, Université d'Avignon et des pays du Vaucluse, ESPACE

MORENO SIERRA DIEGO,
Maître de Conférence en Géographie, Université de Nice Sophia-Antipolis, ESPACE

VOIRON-CANICIO CHRISTINE,
Professeur en Géographie, Université de Nice Sophia-Antipolis, ESPACE

Et les étudiants du Master II GEOPRAD



BABON NORMAN,
DI VINCENZO CHRISTOPHE,
DIALLO MAMADOU,
PHILIPPON TÉRENCE,
POITEAUX QUENTIN,
PONTIUS AMÉLIE

• Collecting data is not enough. What stories do the data tell?

FLORENCE SEDÈS, PROFESSEURE D'INFORMATIQUE, IRIT, TOULOUSE

Qui est-elle ?

Florence Sèdes est professeure des universités. Elle enseigne l'informatique à l'université Toulouse 3 Paul Sabatier et effectue sa recherche à l'Institut de recherche en informatique de Toulouse (IRIT). Son domaine de recherche et ses projets relèvent des systèmes d'information et de la gestion des données (documents, multimédia, métadonnées, vidéosurveillance, sécurité, confidentialité, mobilité...).

Subject of the conference

Data analytics and management is the process of turning data into meaningful information that humans can use to make effective decisions. Raw data is only the starting point. The data need to be analyzed, turned into information and made accessible to users, and the data also needs to meet varying needs and profiles, and be understandable to different audiences. The value... of the data comes from the stories it tells...

• Two case studies of activity-travel behavior, space, and place

KONSTADINOS G. GOULIAS, PROFESSOR, UCSB

Who is he?

Konstadinos (Kostas) Goulias is a professor of transportation in the Geography Department at the University of California Santa Barbara and co-director of the GeoTrans laboratory. From 1991 to 2004, he was a professor of transportation in the Department of Civil Engineering at PennState where he was also director of different research units. He served as chair of the Traveler Behavior and Values Committee (now serving as emeritus member) and chair of the Task Force on Moving Activity-based Approaches to Practice both of the Transportation Research Board of the National Academies in the United States. He edited three books, authored and co-authored more than 270 papers and reports to sponsor, and serves on a variety of editorial and research boards, peer review panels, and committees. He is also co-founding editor (with Kouros Mohammadian) of the journal Transportation Letters. Kostas has a Laurea in Engineering degree (5 years and a thesis) from University of Calabria in Italy (1986), MS in Engineering from University of Michigan, Ann Arbor (1987), and Ph.D. from University of California at Davis (1991). Most of his research is in travel behaviour dynamics and microsimulation.

Subject of the conference

In this presentation I will review two case studies relating activity-travel patterns, space, and place. The first case study is from a recently developed large scale spatio-temporal simulator of activities and travel for Southern California (Los Angeles Metropolis). The simulator includes population synthesis that recreates the entire resident population in this region, provides locations for residences, workplaces, and schools for each person, estimates car ownership and type, and provides other key personal and household characteristics. A synthetic schedule generator recreates for each resident person in the simulated region a daily schedule of activities and travel that reflects intra-household activity coordination for a day. Key informant of these spatiotemporal patterns is accessibility at a fine spatial and temporal resolution. Accessibility here is used to capture the affordances (i.e., possibilities for action) of different locations. In this presentation I review the accessibility indicators and the role they play in determining patterns of behavior. The second case study, from Santa Barbara in California, is about place perception and its measurement using an online survey with ultimate objective building spatial choice models. This case study moves us away from just space representation and closer to a place representation (i.e., a representation that combines affordances with meanings) in behavioral models. One way to represent place is by measuring the attitudes and preferences of people about locations to tease out some of the unobserved variation in behavior. In this research, we examine place attitudes and the possible importance that such attitudes can have on destination choices. Using latent class cluster analysis we find groups of people with substantial differences in perception and preference as expected. In the conclusion of my presentation I discuss differences between the two approaches to include space and place in behavioral models and hint on some ideas we are exploring as next steps.

• **Approche topographique quantitative pour la modélisation et l'analyse spatiale :
Étude de cas en Tunisie**

NOAMEN REBAI, PROFESSEUR, ÉCOLE NATIONALE D'INGÉNIEURS DE TUNIS

Qui est-il ?

Noamen REBAI, Docteur de l'Observatoire de Paris 1992 est actuellement Professeur à l'École Nationale d'Ingénieurs de Tunis au Département de Génie Civil où il enseigne la topographie. De même il intervient dans le mastère Géoressources et Développement Durable, où il enseigne le module Traitement et modélisation des données en géosciences. Depuis 2001, il est coordonnateur du Mastère professionnel en Géomatique à l'ENIT. En tant que chercheur associé, il travaille dans le laboratoire Eau, Energie et Environnement et appartient à l'UR : Dynamique des bassins sédimentaires et structures géologiques.

Sujet de la conférence

Cette présentation s'articule sur le Modèle Numérique d'Élévation et ses attributs, avec le développement de trois aspects qui sont :

- 1- Effet de la topographie sur l'interpolation*
- 2- Évaluation de l'extraction automatique d'un RH à partir d'un MNE*
- 3- Nouvelle approche algorithmique pour la classification d'un RH*

Session 1

Géovisualisation, cartographie, exploration des dynamiques spatio-temporelles

Géovisualisation des dynamiques cycliques d'événements spatio-temporels

JACQUES GAUTIER, PAULE-ANNICK DAVOINE, CLAIRE CUNTY15

Géovisualisation de récits de catastrophes naturelles et de leurs impacts : application aux inondations ayant impacté le système ferroviaire

CÉCILE SAINT-MARC, PAULE-ANNICK DAVOINE, MARLÈNE VILLANOVA-OLIVER, CICELY PAMS CAPOCCIONI, DORINE CHENIER 30

L'approche cartographique de la représentation du mouvement spatial. L'exemple des flux commerciaux maritimes euro-méditerranéens

FRANCOISE BAHOKEN, CLAIRE LAGESSE, CÉSAR DUCRUET 44

Proposition méthodologique pour la visualisation et l'exploration des mobilités géographiques

ÉMILIE LEROND, JEAN-PHILIPPE ANTONI, OLIVIER KLEIN61

Pour une cartographie romanesque de Paris au XIX^e siècle. Proposition méthodologique

THIERRY JOLIVEAU, NOÉMIE BOEGLIN, MICHEL DEPEYRE, YVES-FRANÇOIS LE LAY..... 76

Session 2

Infrastructures de données spatiales, observatoires, géographie participative, géoprospective

Infrastructures de Données Géographiques et observatoires de recherche en environnement : un exemple de mise en œuvre

FRANCOISE GOURMELON, MATHIAS ROUAN, JEAN NABUCET 93

Intérêt de la simulation paysagère en modélisation géoprospective

CHRISTINE VOIRON-CANICIO, JEAN-CHRISTOPHE LOUBIER, DOMINIQUE GENOUD, DANIEL HUNACEK, FLORIAN SANT107

Un modèle et une plateforme pour l'évaluation et la comparaison de services de reconnaissances d'entités nommées spatiales

CHRISTOPHE MARQUESUZAÀ, PATRICK ETCHEVERRY, CHRISTIAN SALLABERRY.....123

Conception logique et physique des hiérarchies spatiales non strictes dans les entrepôts de données spatiales

IBTISAM FERRAHI, SANDRO BIMONTE, KAMEL BOUKHALFA..... 138

Modélisation des activités viticoles sous contraintes climatiques

CYRIL TISSOT, CÉLINE LE COQ, MATHIAS ROUAN, ETIENNE NEETHLING, HERVÉ QUÉNOL153

Un cadre d'analyse des Infrastructures de Données Géographiques pour interroger la mise en réseaux des acteurs et des outils

MATTHIEU NOUCHER, FRANÇOISE GOURMELON, ANTOINE LAUMOND, GUY MELANÇON, BRUNO PINAULT, ADELIN MAULPOIX, JULIE PIERSON, OLIVIER PISSOAT, MATHIAS ROUAN..... 169

Session 3

Dynamiques urbaines, accessibilité

<i>Analyser la dynamique intra-urbaine dans le temps long : une approche par la localisation des activités</i> LUCIE NAHASSIA.....	185
<i>RELOC : un modèle pour produire de l'accessibilité par la localisation plutôt que par la construction d'infrastructures</i> CYRILLE GENRE-GRANDPIERRE.....	199
<i>Une base de données pour modéliser l'urbanisation résidentielle en zone côtière</i> IWAN LE BERRE, ADELIN MAULPOIX, MARIUS THÉRIAULT, FRANÇOISE GOURMELON.....	215
<i>Urban Fabric with Bayesian LINCS. Empirical Evidence from the French Riviera</i> GIOVANNI FUSCO, ALESSANDRO ARALDI.....	229

Session 4

Mobilités, flux, trajectoires, modélisation spatio-temporelle

<i>Les réseaux de transport à grande vitesse en France : construction et apports d'un système d'information géo-historique</i> FRANÇOISE BAHOKEN, CHRISTOPHE MIMEUR, MARTIN KONING, THOMAS THEVENIN.....	247
<i>Formalisation spatio-temporelle de perturbations maritimes pour la simulation du réseau maritime</i> TANGUY MARTIN, ALDO NAPOLI.....	266
<i>Interaction Homme Machine pour l'orientation spatiale des malvoyants</i> SAID BOULAROUK, DIDIER JOSSELIN, EITAN ALTMAN.....	280
<i>Méthodologie de localisation de systèmes fixes d'identification de plaques minéralogiques sur le réseau routier</i> JEAN-PAUL DONNAY, MARIE TROTTA.....	295
<i>Modèle sémantique de trajectoire d'objet mobile</i> MOJDEH SOLTAN MOHAMMADI, THÉRÈSE LIBOUREL, ISABELLE MOUGENOT, CHRISTOPHE FAGOT.....	311
<i>Interrogation de données spatio-temporelles de géolocalisation indoor à partir de trajectoires hybrides</i> FRANCK JEVEME PANTA, FLORENCE SÈDES.....	326

Session 5

Risques, changement climatique

<i>Un indice pour caractériser et représenter la dangerosité des canicules en zone urbaine</i> LAURA PINSON, HA PHAM, ANNE RUAS.....	343
<i>Analyser des tweets géolocalisés pour explorer les réponses sociales face aux phénomènes météorologiques extrêmes : réflexions épistémologiques et verrous méthodologiques</i> CAMILLE CAVALIERE, PAULE-ANNICK DAVOINE, CÉLINE LUTOFF, ISABELLE RUIN.....	357
<i>Détection automatique de séquences sismiques par analyse spatio-temporelle de catalogues de sismicité</i> JENNY TRÉVISAN, LIONEL MATTÉO, ANTHONY SLADEN.....	371

Session 6

Environnement, traitement d'image, résolution spatiale

<i>Génération d'environnements artificiels pour la simulation spatiale d'arboviroses en milieu urbain : application à la dengue et au virus Zika</i> RENAUD MISSLIN, ERIC DAUDÉ.....	389
<i>Contribution de la segmentation d'images en géomorphologie pour spatialiser les processus hydro-pédologiques</i> JULIEN ANDRIEU.....	403

<i>Classification Rules for forest fires loss estimation through data mining techniques. Case study of the Kroumiria region (Northwestern Tunisia)</i> <i>AHMED TOUJANI, HAMMADI ACHOUR, SAMI FAIZ</i>	418
<i>Évolution altitudinale à 25 m de résolution spatiale de l'isotherme 0°C des températures moyennes mensuelles hivernales dans les Alpes du Sud françaises durant le XX^e siècle</i> <i>NICOLAS MARTIN</i>	432

Posters

<i>1. Analyses des cycles saisonniers de la végétation en lien avec les variations saisonnières des précipitations à partir des données de Télédétection basses résolutions spatiales NDVI (MODIS) et de précipitation (TRMM)</i> <i>GUY BOUSSOUGOU, YAO TÉLÈSPHORE BROU, PATRICK VALIMBA</i>	449
<i>2. Apport de l'imagerie hyperspectrale pour la planification</i> <i>CHRISTIANE WEBER</i>	454
<i>3. De l'intérêt de la morphologie mathématique pour la définition de la taille de la fenêtre de convolution lors de la réalisation d'une carte de densité</i> <i>HADRIEN FOUILLADE ORSINI</i>	463
<i>4. Démarche d'analyse de l'intégrité d'un système de localisation de navires</i> <i>CLÉMENT IPHAR, ALDO NAPOLI, CYRIL RAY</i>	468
<i>5. Distance discrète de Fréchet optimisée</i> <i>THOMAS DEVOGELE, MAXENCE ESNALUT, LAURENT ÉTIENNE</i>	472
<i>6. Dynamiques paysagères et dysfonctionnement (SLC-off) du satellite Landsat La morphologie mathématique comme outil d'interpolation de l'occupation du sol</i> <i>MATTHIEU VIGNAL</i>	476
<i>7. Étude du paramètre turbidité par télédétection (Landsat, et Spot5-Take5) : indicateur « d'aléa sanitaire » dans les eaux de surface en Afrique de l'Ouest</i> <i>ELODIE ROBERT, MANUELA GRIPPA, LAURENT KERGOAT, JEAN-MICHEL MARTINEZ, SYLVAIN PINET, TOM SOMDECOSTE, CINDY GOSSET, LAETITIA GAL</i>	480
<i>8. Évaluation de la sensibilité littorale à l'érosion en région PACA dans un contexte de changement climatique</i> <i>ASIMINA MAVROMATIDI, ÉLODIE BRICHE</i>	485
<i>9. Expérimentation de la cartographie collaborative dans la gouvernance des espaces verts Le cas de la ville d'Antibes Juan-les-Pins.</i> <i>CYNTHIA RONDELLI</i>	489
<i>10. FLOOD AR : utiliser des modèles 3D temporels interactifs pour la sensibilisation au risque d'inondation</i> <i>FLORENCE JACQUINOD, JÉRÉMY EDERT, GILLES GESQUIÈRE, MARCEL VOGT, JULIA BONACCORSI, MARYLINE CERQUEIRA DA CUNHA</i>	490
<i>11. Génération d'itinéraires pour le Vélo en Libre Service : Vers la prise en compte de la multi-modalité</i> <i>INES JGUIRIM, DAVID BROSSET, ANNY VIVAS ABRIL, THIERRY DUBOI, CÉLINE MAYOUSSE</i>	498
<i>12. Identifier des espaces préservés de la pollution de l'air sur le chemin de l'école</i> <i>JULIE PRUD'HOMME, MALIKA MADELIN, SOPHIE BAUDET-MICHEL</i>	503
<i>13. Mesures participatives de l'exposition individuelle à la pollution de l'air</i> <i>MALIKA MADELIN, SARAH DUCHÉ, VINCENT DUPUIS</i>	508
<i>14. Modélisation 3D de l'expansion urbaine : application au modèle SLEUTH</i> <i>GUILLAUME DA SILVA, OMAR DOUKARI, RAHIM AGUEJDAD</i>	513
<i>15. Projet PLU++ : vers un Plan Local d'Urbanisme plus intelligible et plus participatif</i> <i>MICKAËL BRASEBIN, SIDONIE CHRISTOPHE, FLORENCE JACQUINOD, ANOUK VINESSE, HORTENSE MAHON</i>	517

16. Simulation par Automate Cellulaire et Système Multi-Agents des conséquences du changement climatique sur l'aire de distribution de Pinus Cembra dans le Mercantour	
MATTHIEU VIGNAL, JULIEN ANDRIEU.....	524
17. Stratégies ouvertes d'intégration de règles métier pour la géomatique dans un système d'information de gestion	
MOUNIR TOUZANI, CHRISTOPHE PONSARD, THÉRÈSE LIBOUREL, JOËL QUINQUETON, ANNE LAURENT	528
18. Towards Automated Image Geolocalization Refining	
BERNARD SEMAAN, GUILLAUME MOREAU, MYRIAM SERVIÈRES, BILAL CHEBARO	542
19. Une application SIG pour l'optimisation de la collecte des déchets post-inondation	
SERGE LHOMME, MICKAËL BRASEBIN, ARNAUD LE GUILCHER, BRUNO BARROCA	546
20. Une approche semi-automatique pour l'extraction d'informations liées aux itinéraires culturels à partir des réseaux sociaux : cas de la Via Francigena	
ERIC KERGOSIEN, NATHALIE VALMOND-LEBLANC, NATALIA GRABAR, MARTA SEVERO.....	553
21. Une démarche de modélisation spatiale multiscalaire pour appréhender la croissance urbaine	
ALEXANDRE ORNON.....	557
22. Utiliser les réseaux opportunistes en cas de catastrophes naturelles : une solution pour pallier la saturation et/ou la coupure des réseaux de télécommunication ?	
JOHNNY DOUVINET, RACHID EL AZOUZI, JULES KOUADIO, OLFA HASSANI, DAMIEN SERRE	561
23. Vers une cartographie efficace des ambiances sonores urbaines	
KAMAL SERRHINI, GAËTAN PALKA, CATHERINE LAVANDIER, CATHERINE DOMINGUES, SAÚL GOMEZ, DOMINIQUE ANDRIEU ...	569
24. Vers une modélisation de la sémantique des trajectoires de navigation maritime côtière	
WISSAME LADDADA	587
25. Visualisation et analyse spatio-temporelle de trajectoires maritimes	
MATTIA BUNEL, CLAIRE LAGESSE, CÉSAR DUCRUET, ÉRIC MERMET, SÉLÉNA PETIT	591
26. Formalisation d'une méthode de filtrage graphique pour simplifier la cartographie des interactions spatiales	
SERGE LHOMME.....	598

[ISBN : 2-910545-11-3]



Session 1

Géovisualisation, cartographie, exploration des dynamiques spatio-temporelles

Geovisualisation des dynamiques cycliques d'événements spatio-temporels

Jacques GAUTIER¹, Paule-Annick DAVOINE², Claire CUNTY³

1. Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes
700 avenue centrale, 38401 Saint Martin-d'Hères, France
Jacques.Gautier@imag.fr

2. Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes
700 avenue centrale, 38401 Saint Martin-d'Hères, France
Paule-Annick.Davoine@imag.fr

3. UMR EVS/IRG Lyon II
5 avenue Pierre Mendès-France, 69500 Bron, France
Claire.Cunty@Univ-Lyon2.fr

RÉSUMÉ. L'identification des événements cycliques, dans le cadre de l'analyse des données temporelles, s'effectue majoritairement par des analyses mathématiques, et ne prend en compte que la dimension temporelle. Notre ambition est de proposer une méthode d'analyse visuelle permettant d'identifier les comportements cycliques d'événements spatio-temporels, qui croiserait les dimensions temporelles et spatiales. Le principe consiste à intégrer, dans un environnement de géovisualisation multi-vues, un diagramme temporel, basé sur les concepts de la Time Coil et de la Time Wave, représentant de façon simultanée les aspects linéaires et cycliques du temps. La synchronisation de ce diagramme avec une représentation cartographique adaptée permet de caractériser la relation entre la localisation d'un événement cyclique et son cycle d'apparition.

ABSTRACT. The identification of cyclic events, as part of temporal data analysis, is mainly done through mathematical analysis, and takes only into account the temporal dimension of events. Our objective is to propose a visual analysis method allowing to identify cyclic spatio-temporal events, and which takes into account both temporal and spatial dimension. The idea is to integrate, into a multi-view geovisualization environment, a temporal diagram, based on Time Coil and Time Wave concepts, in order to represent the linear and cyclic aspects of time. The synchronization of the temporal diagram with a cartographic representation allows to identify the relationship between the location and the return-period of a cyclic event.

MOTS-CLÉS : ÉVÉNEMENTS SPATIO-TEMPORELS, ÉVÉNEMENTS CYCLIQUES, GEOVISUALISATION, ANALYSE VISUELLE, TIME COIL, RELATION ESPACE/PÉRIODE DE RETOUR

KEYWORDS: SPATIO-TEMPORAL EVENTS, CYCLIC EVENTS, GEOVISUALIZATION, VISUAL ANALYSIS, TIME COIL, SPACE/RETURN-PERIOD RELATIONSHIP

1. Introduction

Depuis plusieurs années, la recherche dans le domaine de la géovisualisation a développé des concepts de représentation des dynamiques spatio-temporelles. Plusieurs méthodes (Davoine *et al.*, 2015), qu'elles soient cartographiques ou faisant appel aux technologies informatiques ont été proposées pour intégrer la dimension temporelle dans des outils de géovisualisation : collections de cartes, cartes animées, cartes interactives et dynamiques... Une méthode très utilisée est celle de l'approche multi-vues. Elle consiste à diviser l'interface de l'outil de géovisualisation en plusieurs fenêtres synchronisées, permettant notamment d'afficher l'organisation spatiale des événements dans une carte, et leur distribution temporelle dans un diagramme (Kraak, 2010 ; Moïsuc *et al.*, 2005).

Si cette technique permet de représenter aisément les dates d'occurrence des événements ainsi que leur durée, ces interfaces ne permettent pas de visualiser le caractère cyclique de certains types d'événements. L'étude des récurrences cycliques de ces événements est pourtant essentielle pour l'établissement de modèles de prédiction de certains phénomènes, comme par exemple dans la détermination des aléas dans le domaine des risques naturels. L'étude des comportements cycliques dans les données caractérisant des événements passés est majoritairement effectuée en utilisant des méthodes d'analyses mathématiques, comme par exemple l'analyse harmonique, issue des transformées de Fourier. Cependant, ces analyses ne prennent en compte que la dimension temporelle des événements, sans s'intéresser à leur dimension spatiale.

Notre but est ici de fournir une méthode d'analyse visuelle de ces cycles de retour d'événements spatio-temporels qui croiserait les dimensions temporelles et spatiales. Pour cela, nous cherchons à visualiser les récurrences cycliques des événements spatio-temporels dans les outils de géovisualisation. Cette analyse visuelle pourrait venir compléter les analyses mathématiques utilisées dans l'étude des cycles de retour sur des données temporelles.

Nous présenterons dans un premier temps les temporalités que nous cherchons à mettre en avant dans les outils de géovisualisation, et l'intérêt de prendre en compte la dimension spatiale dans l'analyse des phénomènes cycliques. Puis nous présenterons les principales méthodes permettant l'intégration de la dimension temporelle dans les outils de géovisualisation. Nous montrerons les limites des diagrammes temporels utilisés pour la visualisation des événements cycliques, et les moyens d'y remédier grâce à l'intégration de diagrammes temporels intégrant les composantes linéaires et cycliques du temps, à l'instar des diagrammes *Time Coil* (Edsall et Peuquet, 1997) et *Time Wave* (Li, 2010). Enfin nous exposerons notre proposition, qui se présente sous la forme d'un outil de géovisualisation multi-vues, intégrant une extension des diagrammes temporels *Time Coil* et *Time Wave*, et une carte, synchronisée au diagramme temporel, permettant de visualiser la composante cyclique du temps dans la représentation cartographique au moyen d'une sémiologie graphique adaptée.

SAGEO'2016 – Nice, 6-9 décembre 2016

2. Visualisation d'événements spatio-temporels récurrents

2.1. Définitions : phénomène, événement et temporalités

Un phénomène spatio-temporel peut être considéré comme étant l'évolution dans le temps d'objets spatiaux pérennes, où comme une entité spatio-temporelle inscrite dans le temps et l'espace (Haddad 2009, Galton 2004). En tant qu'entité spatio-temporelle distincte, il peut être représenté comme affectant des objets spatiaux pérennes (Grenon et Smith, 2004), ou bien représenté seul, sans que les objets impactés soient pris en compte dans la modélisation du phénomène (Reitsma, 2004).

Dans notre proposition, nous nous plaçons dans une représentation des phénomènes spatio-temporels en tant qu'entités spatio-temporelles distinctes, dont les instances sont nommées "événements spatio-temporels". Nous représentons uniquement le phénomène, et non les objets spatiaux qu'il impacte.

Hewitt (1983), Allen (1983), DiBiase (DiBiase *et al.*, 1992) et MacEachren (1994) ont décrit six « temporalités » permettant de décrire la dimension temporelle d'un événement (Figure 1).

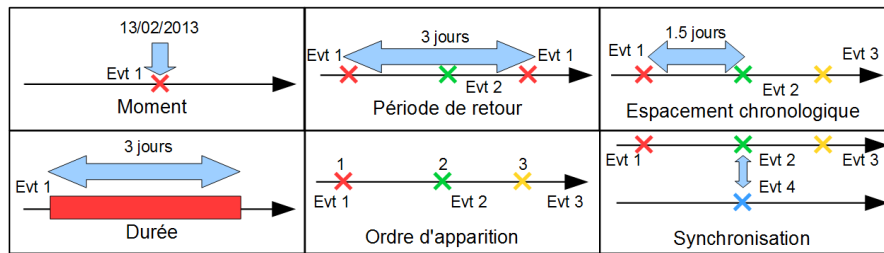


FIGURE 1. *Temporalités des événements (Hewitt, 1983 ; Allen, 1983 ; DiBiase et al., 1992 ; MacEachren, 1994)*

La notion de « période de retour » est ici ambiguë. Dans le domaine des risques, la période de retour est considérée comme l'écart de temps probable entre deux événements (Arnaud, 2009). Elle est différente de la fréquence qui est le nombre d'occurrence d'un événement durant une période donnée (Arnaud, 2009). La période de retour peut cependant être considérée comme la période temporelle entre la réapparition de deux instances d'un phénomène cyclique, soit l'inverse de la fréquence, au sens physique du terme, du phénomène. Visualiser la période de retour d'un événement revient donc à visualiser son caractère cyclique.

2.2. Analyse des événements récurrents : une analyse mathématique sur des données temporelles

L'étude des comportements cycliques dans les récurrences d'événements est très utilisée dans l'établissement de modèles de prédiction, comme par exemple dans le domaine des risques naturels (Ouillon et al., 2008 ; Jones et al., 1999). Cette analyse

Géovisualisation des dynamiques cycliques d'événements spatio-temporels

s'effectue essentiellement par des méthodes mathématiques, comme les analyses harmoniques issues des transformées de Fourier (Lafon *et al.*, 2013). Or, l'exploration visuelle des données peut permettre de mettre en lumière certaines caractéristiques qui ne seraient pas directement décelables par une simple analyse mathématique (Tufte, 1983).

Cette analyse des événements passés est essentiellement effectuée selon une dimension temporelle. Par exemple, dans le cas de l'étude des données sismologiques, il est admis que la probabilité d'occurrence d'un séisme de magnitude donnée est fonction du lieu et de l'échelle d'observation (Ouillon *et al.*, 2008). Afin d'intégrer ce facteur spatial dans l'analyse des données, les zones étudiées sont découpées en secteurs selon un maillage de 10 à 100 km de côté. Une loi probabiliste locale est calculée dans chaque zone à partir des séries temporelles correspondantes (Ouillon *et al.*, 2008). Cette analyse reste donc uniquement effectuée selon une dimension temporelle, malgré la pré-sélection spatiale des données sismiques.

La visualisation des récurrences cycliques d'événements dans un outil de géovisualisation, pourrait être une solution complémentaire aux analyses mathématiques des données concernant les événements passés. Cette méthode permettrait en outre de croiser les dimensions temporelles et spatiales lors de l'analyse. Cependant, si des environnements de géovisualisation intègrent le temps et permettent actuellement de visualiser des temporalités comme le moment, la durée ou l'ordre d'apparition, les temporalités cycliques n'y sont que peu représentées.

3. Visualisation des récurrences d'événements dans les données spatio-temporelles

3.1. La représentation du temps dans les environnements de géovisualisation

L'intégration de la dimension temporelle dans les outils de géovisualisation est basée sur quatre principales méthodes (Davoine *et al.*, 2015) :

- les cartes animées ou dynamiques, où le déroulement du temps est représenté par l'animation de la carte ;
- les collections de cartes, où chaque carte représente la distribution spatiale de l'événement à une date spécifique ;
- le cube spatio-temporel proposé par la Time-Geography (Hägerstrand, 1970), où une troisième dimension spatiale représente la dimension temporelle, utilisé essentiellement pour représenter des problématiques de trajectoires spatio-temporelles ;
- l'approche multi-vues (Figure 2.a), qui consiste à associer, au sein d'un même environnement, différentes constructions visuelles synchronisées entre elles, chacune représentant une dimension de l'information géographique (spatiale, temporelle, thématique) (Kraak, 2010 ; Moïsuc *et al.*, 2005). Dans cette approche, le temps est représenté par un graphique temporel, de

SAGEO'2016 – Nice, 6-9 décembre 2016

type *timeline* ou *timewheel* (Edsall *et al.*, 1997), selon que le temps soit appréhendé de façon linéaire ou cyclique (Figures 3 et 4) ;

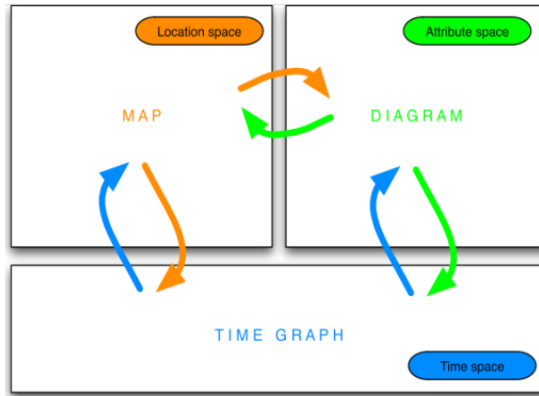


FIGURE 2. Intégration du temps dans un environnement de géovisualisation. Approche multi-vues(a) (Kraak, 2010), timeline (b) et timewheel (c)

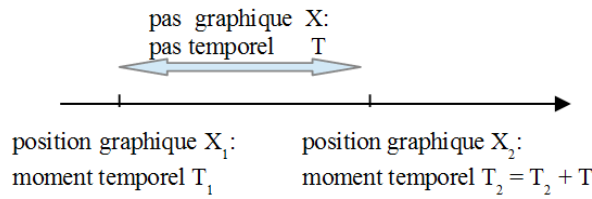


FIGURE 3. Diagramme de type timeline

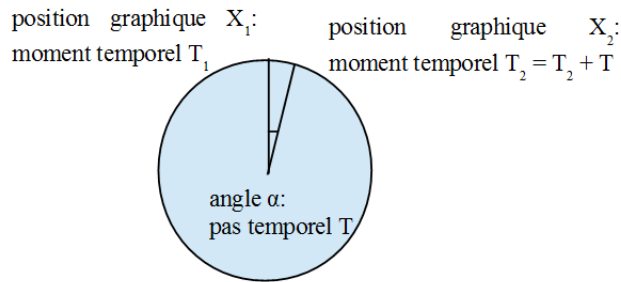


FIGURE 4. Diagramme de type timewheel

Les *timelines* (ligne de temps) et *timewheels* (roue du temps) sont généralement interactives, et permettent de modifier l'échelle temporelle d'observation, d'accéder à plusieurs granularités temporelles (jours, mois, années, ...), et de sélectionner une sous-section temporelle. Dans le cas d'une *timewheel*, le changement de l'échelle temporelle se traduit par la modification de la période cyclique représentée par le cercle de la *timewheel* (par exemple le mois, ou la semaine ou la journée).

Géovisualisation des dynamiques cycliques d'événements spatio-temporels

Des temporalités comme le moment, la durée, ou l'ordre d'apparition, sont visualisables directement sur les *timelines* et *timewheels*. Cependant, ces deux diagrammes ne permettent pas de visualiser directement le caractère cyclique d'une récurrence d'événements. Dans une *timeline* ou une *timewheel*, l'utilisateur doit identifier une distance régulière entre chaque objet graphique correspondant à une instance d'un type d'événement pour repérer une récurrence cyclique (Figure 5).

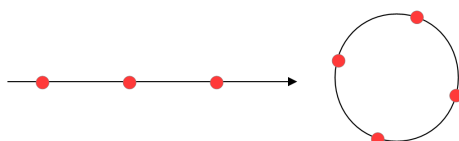


FIGURE 5. Récurrence cyclique d'événements dans une *timeline* et une *timewheel*

3.2. Visualisation des temporalités cycliques

Certains diagrammes temporels intègrent à la fois les aspects linéaires et cycliques du temps. Ces représentations mixtes permettent de visualiser le caractère cyclique d'une récurrence d'événements. Certains de ces diagrammes proviennent de la DataVisualization, ou de l'InfoVisualization, comme le *Cluster Calendar* proposé par Van Wijk (1999). D'autres ont été conçus pour être intégrés dans des environnements de géovisualisation, nous en présentons ici deux exemples.

Edsall et Peuquet (1997) ont proposé un diagramme, nommé *Time Coil*, s'appuyant sur la figure de la « bobine » (Figure 6), qui combine une *timeline* et une *timewheel*. Ce diagramme est composé d'une hélice effectuant des révolutions autour d'un axe central. L'axe de révolution de l'hélice représente le temps linéaire, tandis que chaque révolution de l'hélice représente une période cyclique. Dans la Figure 4, chaque révolution de l'hélice correspond à une période de 12 heures.

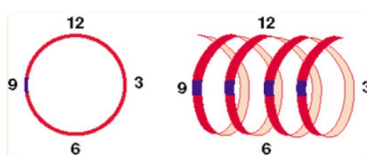


FIGURE 6. *Time Coil* (Edsall et Peuquet, 1997)

Li et Kraak (Li, 2010) ont proposé un diagramme similaire, appelé *TimeWave*, s'appuyant cette fois sur la figure de la vague (Figure 7), qui peut être vu comme une version « plane » de la *Time Coil* d'Edsall et Peuquet. Dans ce diagramme, le temps linéaire est représenté par un axe central, tandis que chaque « vague » correspond à une période cyclique.

SAGEO'2016 – Nice, 6-9 décembre 2016

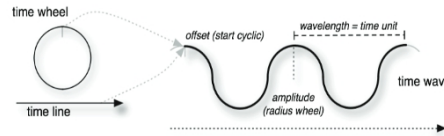


FIGURE 7. *Time Wave* (Li, 2010)

Dans ces diagrammes, une récurrence cyclique, quasi-cyclique, ou semi-cyclique d'événements peut être visualisée par un alignement, parallèle à l'axe central, d'objets graphiques représentant des événements. La période de retour de ces événements cycliques correspond à la durée représentée par une révolution de l'hélice (*Time Coil*) ou une vague (*Time Wave*). Nous appelons ici :

- 'comportement quasi-cyclique' une récurrence d'événements ayant une période de retour quasi-régulière (Figure 8) ;

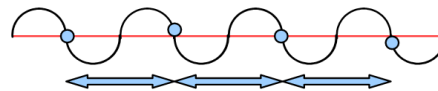


FIGURE 8. *Récurrence quasi-cyclique*

- 'comportement semi-cyclique' une récurrence d'événements ayant une période de retour régulière ou quasi-régulière, mais où l'événement n'apparaît pas forcément à chaque instance de la période de retour (Figure 9);

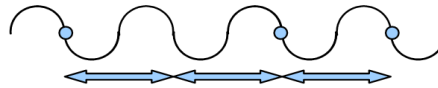


FIGURE 9. *Récurrence semi-cyclique*

Les valeurs pouvant être affectées à la période cyclique représentée par la structure de ces deux diagrammes sont pourtant limitées. Selon Hornsby et Egenhofer (2002), les périodes utilisées dans les représentations cycliques du temps sont le plus souvent basées sur les rythmes de la nature ou des activités humaines, et correspondent donc à des périodes usuelles comme les jours, les mois, les saisons, ... etc. Ainsi, les périodes cycliques représentées, par exemple dans la *Time Wave*, sont limitées à ces périodes usuelles. Par extension, les événements cycliques visualisables sont ceux dont la période de retour sera égale à ces périodes usuelles.

D'autre part, la visualisation d'une récurrence cyclique d'événements à travers ces diagrammes ne prend en compte que la dimension temporelle des événements représentés, même au sein d'un environnement de géovisualisation. Les possibles relations entre la dimension spatiale des événements et leur possible cyclicité ne sont pas visualisées.

4. Proposition

4.1. Objectif et approche méthodologique

Notre objectif est de fournir une méthode visuelle d'analyse des comportements cycliques des événements spatio-temporels, qui croiserait les dimensions temporelles et spatiales.

La méthode proposée est basée sur l'utilisation d'un environnement de géovisualisation multi-vues, et repose sur deux principes : d'une part la visualisation des récurrences cycliques, quasi-cycliques, semi-cycliques d'événements spatio-temporels au sein de la fenêtre temporelle ; d'autre part la gestion des interactions espace-temps afin de visualiser les relations entre cycles de retour et localisation spatiale des événements.

4.2. Représentation de la dimension temporelle

Dans la fenêtre temporelle, nous utilisons un diagramme hélicoïdal inspiré de la *Time Coil* d'Edsall et Peuquet (1997). Ce diagramme peut être affiché en trois dimensions à la manière de la *Time Coil*, ou dans une version "plane" en deux dimensions, semblable à la *Time Wave* (Li, 2010) comme le montre la Figure 10. L'étendue temporelle représentée par ce diagramme est modifiable de manière interactive par l'utilisateur.

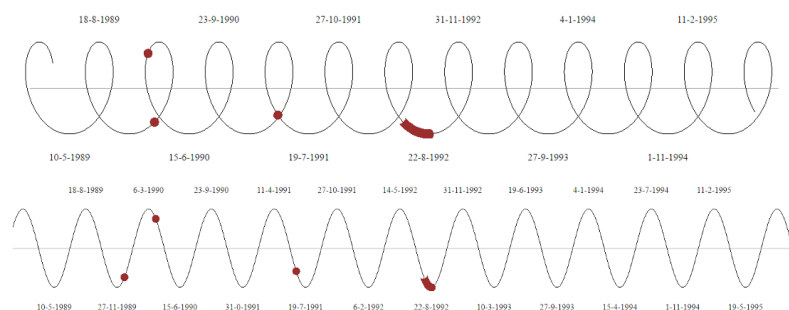


FIGURE 10. Diagramme hélicoïdal de la proposition, vu en 3D, puis dans une version plane ; les deux diagrammes représentent la même étendue temporelle (Données:coulées de lave du Piton de la Fournaise, Île de la Réunion, source OVPF)

La dimension temporelle des événements (moment d'apparition, durée) est représentée au travers d'objets graphiques. Ceux-ci peuvent être ponctuels dans le temps et représentés graphiquement par un point, ou étendus dans le temps et représentés graphiquement par un cylindre suivant la courbe de l'hélice. La couleur de l'élément graphique dépend de la classification appliquée aux événements

SAGEO'2016 – Nice, 6-9 décembre 2016

représentés. Des phénomènes naturels peuvent par exemple être classés selon leur intensité, ou leur type. Pour nos figures, nous avons représenté des événements d'une seule et même classe.

A l'instar des diagrammes *Time Coil* et *Time Wave*, ce diagramme est interactif et permet de modifier la période de temps cyclique représentée par une révolution de l'hélice. Cependant, les valeurs temporelles pouvant être affectées à la période de révolution sont étendues à toutes les valeurs possibles, y compris des valeurs non-calendaires (3 jours, 5 jours et 7 heures, ...etc).

En permettant une modification dynamique et graduelle de la composante cyclique du temps représentée, notre but est double :

- d'une part, permettre la visualisation de toute récurrence cyclique d'événements, quelle que soit la durée de sa période de retour ;
- d'autre part, faciliter une recherche visuelle de cycles de retour, en jouant sur la durée de la période de révolution de l'hélice ;

Lorsque la valeur de la période de révolution de l'hélice approche celle d'une période de retour d'un événement cyclique, on observe la formation d'un alignement des objets graphiques correspondant à ces événements (Figure 11). Les objets graphiques forment d'abord une hélice (ou une courbe sinusoïdale si le diagramme est représenté sur un même plan comme pour la *Time Wave*). Puis les objets s'alignent de plus en plus. Enfin l'alignement devient parallèle à l'axe central du diagramme lorsque la période de révolution de l'hélice est égale à celle de la période de retour. Dans la Figure 11, on représente sur un diagramme temporel une série d'événements fictifs. Ces événements sont classés en deux types A (rouges) et B (bleus). Plus on fait varier la période « P » de révolution de l'hélice vers la valeur de 45 jours, plus on observe l'alignement des objets rouges, correspondant aux événements A. Lorsque P est égale à 45 jours, les objets rouges sont alignés et parallèles à l'axe central. On observe donc une récurrence cyclique des événements de type A, avec une période de retour de 45 jours.

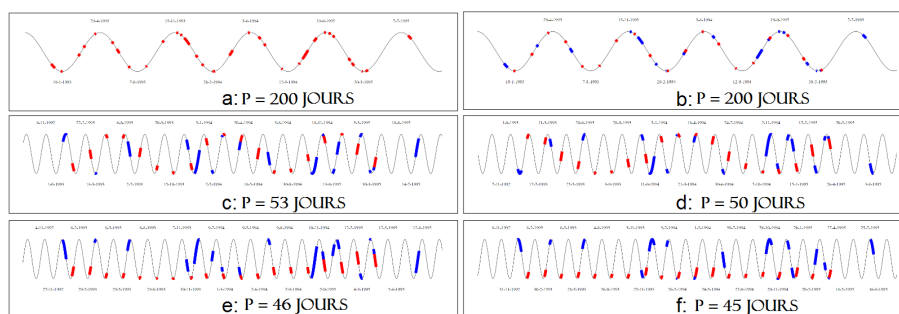


FIGURE 11. Recherche visuelle d'une récurrence cyclique par modification de la période de révolution P du diagramme hélicoïdal (Données fictives)

Géovisualisation des dynamiques cycliques d'événements spatio-temporels

4.3. Représentation de la dimension spatiale

La fenêtre spatiale se compose d'une carte représentant l'emprise spatiale des événements représentés sur le diagramme hélicoïdal. La carte ne représente pas l'espace à un instant donné, comme par exemple dans une carte animée, mais l'ensemble des événements ayant lieu durant la période de temps couverte par le diagramme hélicoïdal (Figure 12).

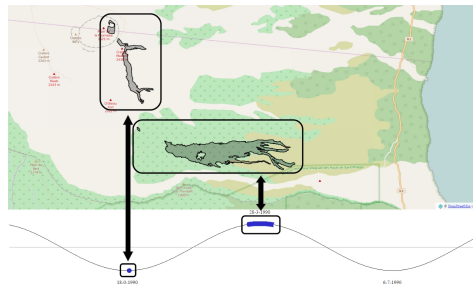


FIGURE 12. Représentation de la dimension spatiale des événements (Données: coulées de lave du Piton de la Fournaise, Île de la Réunion, source OVPF)

L'entité spatiale représentée sur la carte peut être un point, une ligne, un polygone. Dans notre exemple (Figure 12), nous représentons des coulées de lave par des objets polygonaux, correspondant à l'emprise spatiale de la coulée.

4.4. Relation Espace-temps

La fenêtre temporelle comporte, en dessous du diagramme hélicoïdal, une troisième frise temporelle, que nous appellerons « frise de couleur » (Figure 13).

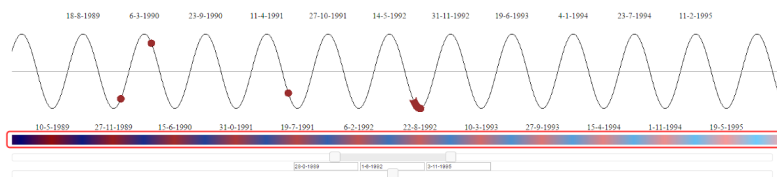


FIGURE 13. Frise de couleur (Données: coulées de lave du Piton de la Fournaise, Île de la Réunion, source OVPF)

Dans cette frise sont représentés les deux aspects du temps :

- le temps cyclique par l'alternance entre deux couleurs, ici le rouge et le bleu (Figure 14). La couleur bleue correspond à chaque sommet de l'hélice, la couleur rouge à leurs opposés par rapport à l'axe central. La couleur bleue correspond au début d'une période temporelle représentée par une

SAGEO'2016 – Nice, 6-9 décembre 2016

révolution du diagramme hélicoïdal, et la couleur rouge correspond au milieu de cette période ;

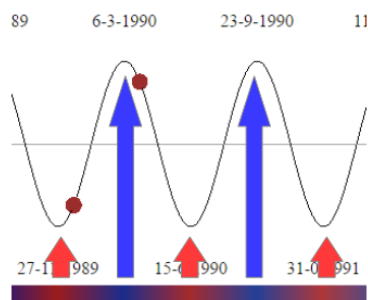


FIGURE 14. Relation frise de couleur/timeline hélicoïdale (Données: coulées de lave du Piton de la Fournaise, Île de la Réunion, source OVPF)

- le temps linéaire par un dégradé de la valeur des deux précédentes couleurs. Ici les dates les plus anciennes correspondent au rouge sombre et au bleu sombre, tandis que les dates les plus récentes correspondent au rouge clair et au bleu clair ;

La couleur de l'entité cartographiée, correspondant à un événement, dépend de la position de l'entité graphique correspondante sur le diagramme hélicoïdal. Par exemple, un événement apparaissant au début d'une période de révolution du diagramme hélicoïdal sera représenté par une entité graphique située en haut d'une boucle du diagramme temporel ; l'entité cartographique correspondante sera alors de couleur bleue (Figure 15);

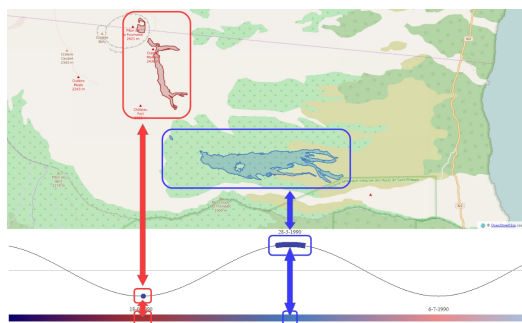


FIGURE 15. Relation espace/temps (Données: coulées de lave du Piton de la Fournaise, Île de la Réunion, source OVPF)

Une récurrence cyclique d'événements peut alors être visualisée directement dans la carte, un alignement d'événements sur le diagramme temporel correspondant à

Géovisualisation des dynamiques cycliques d'événements spatio-temporels

l'affichage sur la carte d'objets de même couleur. Dans la Figure 16, trois entités cartographiques sont de couleur bleue, et correspondent à trois entités graphiques alignées sur le diagramme hélicoïdal. Ces entités sont séparées par quatre puis par une période de révolution du diagramme hélicoïdal, celle-ci étant égale à 98 jours. L'événement pourrait donc présenter à une récurrence semi-cyclique ayant une période de retour de 98 jours. Dans l'exemple de la Figure 16, une apparition semi-régulière de trois objets ne suffit pas pour déterminer le caractère cyclique des événements représentés. Cette figure a cependant pour but de présenter la méthode effectuée pour identifier une récurrence cyclique d'événements sur la carte.

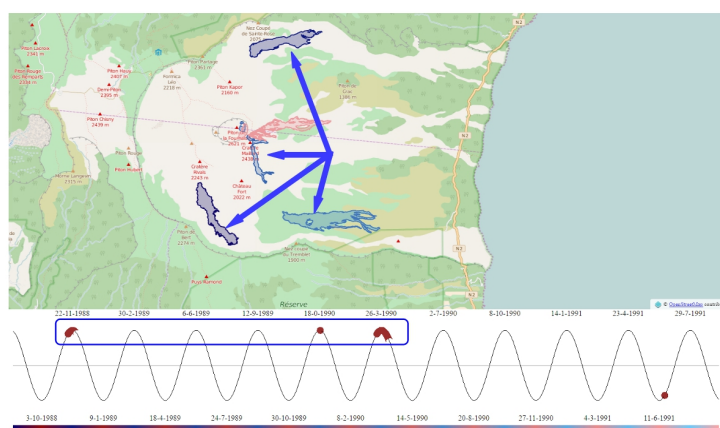


FIGURE 16. Identification de récurrences cycliques sur la carte
(Données:coulées de lave du Piton de la Fournaise, Île de la Réunion, source OVPF)

La modification de la période de révolution du diagramme hélicoïdal engendre la modification de l'aspect de la frise de couleur, et donc de la symbologie affectée aux entités de la carte. L'identification de cycles de retour par une recherche visuelle peut donc également être effectuée dans la fenêtre cartographique. Cette analyse visuelle sur la carte présente un intérêt supplémentaire dans la recherche de comportements cycliques localisés. En effet, un événement peut, dans le même temps, ne pas apparaître de façon cyclique à l'échelle d'une région, et présenter une régularité cyclique dans une zone plus restreinte.

De la même façon, un phénomène cyclique peut présenter des périodes de retour différentes selon sa région d'apparition. La modification de la période de révolution de l'hélice permet l'identification de ces récurrences cycliques localisées, lorsque des entités cartographiques, proches les unes des autres, prennent la même couleur sur la carte, tandis que les entités des autres zones de la carte prennent une autre couleur. Dans la Figure 17, les entités présentes dans le cadre bleu possèdent la même couleur lorsque la période de révolution du diagramme hélicoïdal est égale à 205 jours. Les événements correspondants (en bleu) dans le diagramme hélicoïdal sont

SAGEO'2016 – Nice, 6-9 décembre 2016

quasi alignés. Cette zone pourrait présenter une récurrence quasi-cyclique localisée de 205 jours. Comme dans la figure précédente, l'exemple de la Figure 17 ne permet pas de certifier l'existence d'une récurrence cyclique dans cette zone. Cette figure a pour but d'illustrer la méthode proposée.

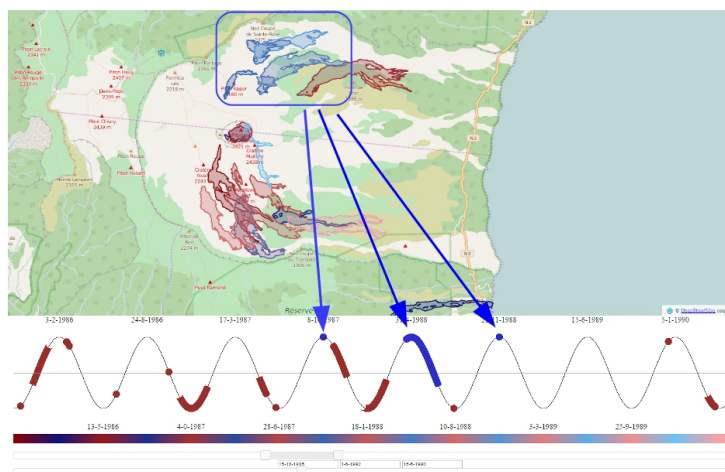


FIGURE 17. Identification de récurrences cycliques localisées sur la carte (Données:coulées de lave du Piton de la Fournaise, Île de la Réunion, source OVPF)

5. Conclusion

Nous avons proposé une méthode visuelle d'analyse des temporalités cycliques des événements spatio-temporels. Celle-ci est basée sur le concept d'un environnement de géovisualisation multi-vues intégrant :

- un diagramme temporel hélicoïdal, associant le temps linéaire et le temps cyclique, ce qui permet de visualiser directement les événements cycliques par un alignement des objets graphiques correspondants sur le diagramme temporel. La période de révolution du diagramme, représentant la composante cyclique du temps, peut être modifiée pour prendre toutes les valeurs possibles, y compris les valeurs non-calendaires. La modification dynamique et graduelle de cette période de révolution permet une recherche visuelle des cycles de retour des événements, à travers l'identification d'alignements d'objets ;
- une représentation de la composante cyclique du temps dans la carte, synchronisée avec celle du diagramme temporel. Ceci permet de visualiser directement les cycles de retour d'événements dans la carte, et autorise une recherche visuelle de des récurrences cycliques localisées.

Géovisualisation des dynamiques cycliques d'événements spatio-temporels

Cette méthode d'analyse permet la mise en relation des dimensions temporelles et spatiales dans l'analyse des cycles de retour. Elle offre également une méthode visuelle complémentaire à l'analyse mathématique des cycles de retour dans les séries temporelles. Notre prochaine étape sera de définir un protocole expérimental afin d'appliquer notre proposition sur différents types de risques naturels tels que les inondations ou les avalanches, et de tester cette proposition sur un échantillon d'utilisateurs potentiels. Ces tests permettront de valider l'apport de notre méthode dans l'analyse des événements cycliques et d'apporter des améliorations à cette proposition.

Remerciements

Cette recherche est soutenue par la région Auvergne-Rhône-Alpes dans le cadre de l'ARC 3 Environnement.

Bibliographie

- Allen J.-F. (1983). Maintaining knowledge about temporal intervals. *Communication of the ACM 1983*,
- Arnaud A., Davoine P.-A. (2009) Cartographie des temporalités dans le domaine des risques. *Comité Français de Cartographie, N°202, December 2009*
- Claramunt C., Thériault M. (1995). Managing time in GIS: an event-oriented approach dans Clifford J., Tuzhilin A. (dir.) (1995). *Proceedings of the International Workshop on Temporal Databases : Recent Advances in Temporal Databases*, Springer-Verlag, p. 23-24, Zurich, Suisse.
- Davoine P.-A., Mathian H., Saint-Marc C., Kaddouri L., Blaise J.-Y. (2015). The Visual Representations of Territorial Dynamics: Retrospective and Input from New Computing Environments. *Proceedings of 27th International Cartographic Conference, 16th General Assembly 2015*, Rio de Janeiro, Brazil.
- DiBiase D., MacEachren A.-M., Krygier J.-B., Reeves C. (1992). Animation and the role of map design in scientific visualization. *Cartography and Geographic Information Systems, 19(4): p. 201-214*.
- Edsall R., Kraak M.-J., MacEachren A.-M., Peuquet D.J. (1997). Assessing the effectiveness of temporal legends. *Environmental Visualization*, pages 667-685 *GIS/LIS*, Cincinnati, 1997
- Edsall R., Peuquet D. (1997). *A graphical user interface for the integration of time into GIS*. Research paper of The Apoala Project. Pennsylvania.
- Galton A. (2004) *Fields and objects in space, time, and space-time*. *Spatial Cognition and Computation*, vol. 4, n°1, p. 39-68, 2004
- Grenon P., Smith B. (2004), SNAP and SPAN: towards dynamic spatial ontology, *Spatial Cognition and Computation*, vol. 4, N°1, p. 69-103.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Haddad H. (2009) *Une approche pour supporter l'analyse qualitative des suites d'actions dans un environnement géographique virtuel et dynamique. L'analyse « What-if » comme exemple.* PhD Thesis. Université Laval, Québec, Canada.
- Hägerstrand T. (1970), What about people in regional science ? *Papers of the Regional Science Association*, N°24, p. 7-21.
- Hornsby L., Egenhofer M.-J. (2002). Modeling Moving Objects over Multiple Granularities. *Annals of Mathematics and Artificial Intelligence*, 36 : p. 177-194.
- Hewitt K. (1983). The Idea of Calamity in a Technocratic Age. In K. Hewitt, *The risks and hazards series*, Allen & Unwin Inc., Boston, p 3-32.
- Jones G., Chester D.K., Shooshtarian F. (1999). Statistical analysis of the frequency of eruptions at Furnas Volcano, Sao Miguel, Azores. *Journal of Volcanology and Geothermal Research*, 92 p. 31–38.
- Kraak M.-J., Ormeling F. (2010). *Cartography, Visualization of spatial data*. Routledge, 2010.
- Lafon F., Benoit M., Chevier C., C.-M. Dulic, Pineau L. (2013). *L'aléa inondation. État de l'art préalable à l'élaboration du guide inondation pour les installations nucléaires* http://www.irsn.fr/FR/expertise/rapports_expertise/Documents/surete/IRSN_Rapport-Alea-Inondation_2013.pdf p.8-24.
- Li X. (2010).. *The Time wave in time space. A visual exploration environment for spatio-temporal data.* PhD Thesis, University of Twente.
- MacEachren A.-M. (1994). Time as a cartographic variable. Dans Unwin D.J., Hearnshaw H.M., Wiley, *Visualization in Geographic Information Systems*, p 115--130.
- Moisuc B., Davoine P.-A., Gensel J., Martin H. (2005). Design of Spatio-Temporal Information Systems for Natural Risk Management with an Object-Based Knowledge Representation Approach, *Geomatica*, Vol. 59, No. 4.
- Ouillon G., Corsini M., Tabeaud M., Cossart E. (2008). *Evaluation et Gestion des Risques Naturels*, <https://e-cours.univ-paris1.fr/modules/uved/risques-naturels/html/2/21/212/index.html>
- Reitsma, F. (2004) *A New Geographic Process Data Model*. PhD thesis, Faculty of the Graduate School of the University of Maryland, College Park, 2004
- Tufte, E. (1983) *The Visual Display of Quantitative Information*. Cheshire, Connecticut, 1983

Géovisualisation de récits de catastrophes naturelles et de leurs impacts

Application aux inondations ayant impacté le système ferroviaire

Cécile Saint-Marc^{1,2}, Paule-Annick Davoine¹, Marlène Villanova-Oliver¹, Cicely Pams-Capoccioni², Dorine Chenier²

1. Université Grenoble-Alpes, CNRS, LIG
700 avenue Centrale, 38000 Grenoble, France
{prenom.nom}@imag.fr

2. SNCF Ingénierie & Projet
6 avenue François Mitterrand, 93574 La Plaine Saint-Denis, France
{prenom.nom}@reseau.sncf.fr

RESUME. Pour se préparer face aux risques à venir, il est utile d'étudier les événements passés issus de risques. Un support cartographique replace les événements historiques dans leur contexte spatial. Le doter d'animation et/ou d'interactivité peut permettre d'analyser les dynamiques des événements. Cet article présente des propositions de géovisualisation pour le récit des événements historiques issus de risques, en se basant sur le cas d'étude des inondations et de leurs impacts sur le système ferroviaire. Développée avec des experts du domaine ferroviaire, les propositions répondent à deux problématiques cartographiques : co-visualiser des événements nombreux et superposés au cours du temps et afficher les relations de causalités entre événements pour témoigner des effets dominos.

ABSTRACT. Studying past at-risk phenomena is useful to prepare against future risks. A map allow to locate historical events in their spatial context. Animation and interactivity may allow to analyse the dynamics of events. This paper deals with a visual method to represent narratives of past natural risk events in a map. It is based on the case study of past floods, which impacted the railway network. Map visualizations were developed with experts of the railway system, to overcome two main challenges: to visualize punctual events which have occurred at the same place over time, and to display causality relationships between events in order to show domino-effects.

MOTS-CLES : géovisualisation, information historique, carte spatio-temporelle, récit, risques naturels, inondations.

KEYWORDS: geovisualization, historical information, spatio-temporal map, narrative, natural risks, floods.

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

La prévention et la protection contre les risques naturels futurs gagnent à s'appuyer sur la connaissance des événements historiques (Benito et al. 2004; Coeur and Lang 2008). Au-delà du récit historique littéraire, visualiser les événements historiques issus de risques sous la forme de cartes géographiques, peut permettre aux experts d'appréhender la dynamique et les processus des risques et de les analyser dans leur contexte territorial. Ces événements naturels majeurs sont caractérisés par des effets dominos, c'est-à-dire « un aléa entraînant par effet en chaîne l'événement catastrophique » (Provitolo 2005) qui font donc partie intégrante de la catastrophe. Il est intéressant de les visualiser dans la carte, de façon à pouvoir les analyser. Ces effets dominos portent sur une diversité d'objets géographiques caractérisant le territoire, et sont d'autant plus nombreux que le phénomène naturel est important et se déroule sur une longue période. Par conséquent, les événements conséquences du phénomène naturel initial peuvent se superposer, se chevaucher ou être très proches les uns des autres ce qui pose des problèmes de visibilité et de lisibilité de la carte. Un autre problème concerne la prise en compte des temporalités dans la géovisualisation. La diversité des événements considérés (phénomène naturel initial et ceux qui s'en suivent, dommages sur les infrastructures, mesures de réaction de la part des acteurs) conduit à devoir gérer des échelles de temps emboîtées, avec des rythmes temporels différents.

Le projet DHI-Avenir, initié entre SNCF Ingénierie & Projets, le Laboratoire d'Informatique de Grenoble et Acthys-Diffusion, a abordé ces questions sur le cas d'étude des inondations historiques et de leurs impacts sur le système ferroviaire. Des propositions de représentations cartographiques ont été émises pour les résoudre, dans un contexte d'usage expert.

La première partie de cet article présente les caractéristiques des données de récit d'événements issus de risques et les problématiques de représentation qui leur sont associées. Un état de l'art des méthodes pour cartographier les récits est ensuite présenté et les avantages et limites de différentes propositions sont analysés. Une méthode de cartographie adaptée au récit des inondations et des processus issus de risques est proposée. Les perspectives d'utilisation pratique de ces propositions dans un contexte expert sont finalement discutées.

2. Problématiques liées aux données de récits historiques

2.1. Les notions de récit historique et d'événements issus de risques

En Histoire, le récit historique est une exposition neutre, sous une forme textuelle, des faits survenus dans le passé (Delacroix 2016). Le récit s'inscrit dans un territoire en suivant une chronologie linéaire d'objets géographiques. Cartographier ce récit consiste à représenter différentes dynamiques : localisation d'objets géographiques (apparition / disparition), changement d'état, de forme ou déplacement de ces objets (Kaddouri et al. 2014).

Géovisualisation de récits de catastrophes naturelles et de leurs impacts

En adaptant cette définition au contexte du récit des événements historiques issus de risques, relater les faits consistera à retranscrire sur une carte une succession d'événements qui découlent de la survenue d'une catastrophe naturelle. Précisons que le terme événement est ici employé dans un sens générique qui correspond au « fait historique » des historiens, et pas dans le sens phénoménologique souvent employé en hydrologie. Selon la définition de Zacks et Tversky (Zacks and Tversky 2001), un événement est « un segment de temps à une localisation donnée, qui est discrétisé par un observateur pour avoir un début et une fin ». Selon cette logique, une catastrophe naturelle, telle qu'une inondation, peut être segmentée en plusieurs événements élémentaires : 1) le phénomène hydrométéorologique (les précipitations gonflent le cours d'eau) ; 2) le phénomène de submersion (le cours d'eau sort de son lit et inonde la vallée) ; 3) les conséquences (l'eau envahit telle gare, détruit un pont, etc.) et 4) les mesures de réaction (arrêt des circulations, lancement de travaux, etc.). Ce sont ces événements élémentaires (phénomènes, conséquences, mesures de réaction), et leurs relations de causes à effets, qui seront intégrés à la carte.

Dans notre cas d'étude, pour les gestionnaires et les utilisateurs d'infrastructures, tels que le propriétaire du réseau ferré et les opérateurs ferroviaires, l'objectif est de mieux comprendre comment les inondations passées ont impacté le système ferroviaire. Il s'agit donc d'identifier quels événements sont survenus pendant l'inondation, à quel moment et à quels endroits, quels segments du réseau ferroviaire sont sensibles aux inondations et quel a été l'enchaînement des phénomènes naturels, des dommages et des mesures de réactions (quelles actions ont été entreprises pour remettre en état le système jusqu'au retour à la normale des circulations). Il est donc important pour les experts de distinguer dans la carte les différents types d'événements survenus pendant l'inondation, leur enchaînement chronologique, et d'identifier les relations de cause à effet entre événements.

2.2. Les données de notre cas d'étude

Les données à cartographier proviennent de récits textuels de cinq inondations survenues en France (Tarn en mars 1930, Pyrénées-Orientales en 1940, Vosges en 1947, Aude en 1999 et Somme en 2001), qui sont des synthèses de documents d'archives réalisées par la SNCF (la méthodologie de reconstitution des récits d'événements est décrite dans (Saint-Marc et al. 2014)). Ces récits décrivent les phénomènes d'inondation, leurs impacts et leur prise en charge par les acteurs du ferroviaire. Les récits, subdivisés en événements élémentaires, sont décrits dans une ontologie (Saint-Marc et al. 2016). Les relations de causalités relatives à chaque événement sont référencées. Un exemple de la structuration des données est présenté dans la FIGURE 1.

Les informations retrouvées en archives sont par nature fragmentaires et comportent des imperfections (imprécisions, incertitudes sur les lieux, les dates). Les dates et localisations de certains événements historiques ont dû être reconstituées à partir d'éléments du contexte connus, car elles sont nécessaires pour afficher ces événements dans la carte au cours du temps. Ces événements étant associés à une localisation, plus ou moins précise, leur cartographie est

SAGEO'2016 – Nice, 6-9 décembre 2016

envisageable. Toutefois, celle-ci soulève un certain nombre de problèmes discutés dans la section suivante.

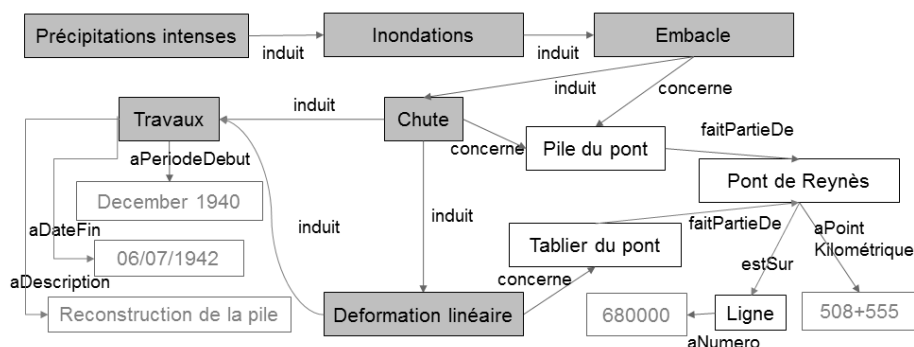


FIGURE 1. Extrait de l'instanciation de l'ontologie de récit sur l'inondation de 1940 dans les Pyrénées ; les événements, indiqués en gris, seront représentés sur la carte.

2.3. Verrous pour la visualisation de récits historiques

Les événements issus d'une inondation majeure sont nombreux et plusieurs événements peuvent s'être produits au même endroit au cours du temps. Un premier verrou est celui de la visualisation des temporalités dans les cartes. Les événements survenus pendant une inondation sont répartis sur de longues périodes de temps, avec une majorité d'événements survenus pendant les premiers jours de l'inondation. La visualisation de récit devrait refléter ces rythmes pour permettre au lecteur de comprendre la dynamique des événements.

Deuxièmement, les événements d'une inondation sont liés entre eux par des relations de cause à effet, connues dans le champ des risques majeurs sous le nom d'*effets dominos* (Provitolo 2005). Il est intéressant de représenter ces relations dans la carte pour décrire les processus à l'œuvre pendant l'inondation, par exemple pour identifier quels types de dommages ont engendré quels autres dommages. Le fait que les conséquences entre événements soient parfois à long-terme (par exemple, pendant l'inondation de 1999, l'endommagement d'une partie de la voie ferrée le 18 novembre 1999 a pour conséquence des travaux qui ne sont survenus qu'en décembre 2000), contraint aussi la façon dont ils pourront être affichés.

Troisièmement, plusieurs événements peuvent être survenus au même emplacement au fil du temps ce qui cause des problèmes de densité et de superposition visuelle entre les symboles de la carte. A titre d'exemple, 332 événements sont survenus pendant l'inondation de 1999 sur la région Languedoc-Roussillon (France), avec de 1 à 35 événements successifs à chaque localisation impactée sur la voie ferrée.

Enfin, cartographier les événements survenus pendant une inondation suppose de s'intéresser à des objets géographiques de géométries différentes. Par exemple, le

Géovisualisation de récits de catastrophes naturelles et de leurs impacts

champ d'inondation est un objet surfacique, le segment de voie ferrée endommagé est un objet linéaire mais le dispositif de signalisation mis hors service est un objet ponctuel.

Plusieurs approches cartographiques existent pour représenter des récits d'événements et permettent de répondre à certains des verrous identifiés. Un état de l'art des cartes de récits historiques est présenté dans la section suivante.

3. Approches de cartographie de récits historiques

3.1. Carte interactive animée et non animée

Historiquement, les récits racontés sous forme de carte étaient constitués soit d'une carte unique montrant des événements ou un itinéraire, comme par exemple la célèbre carte de Minard (Minard 1869), soit d'une collection de cartes chacune représentant l'état du phénomène à une date ou une période de temps (Davoine et al. 2015). La date était représentée par la teinte ou la valeur de la couleur (Bertin 1967), ou par des labels intégrés dans la carte (Vasiliev 1997). De nos jours, la majorité des géovisualisations présentant des récits sont interactives et animées (Kaddouri et al. 2014). Les techniques d'interaction offrent la possibilité à l'utilisateur d'interroger la carte et les variables dynamiques (Dibiase et al. 1992; MacEachren 1994) viennent enrichir la visualisation. Si l'animation paraît plus intuitive pour donner à voir les dynamiques (Tversky et al. 2002), il n'est pas évident qu'elle soit plus efficace pour extraire de l'information de la carte. En effet, un écueil souvent reproché aux cartes animées est qu'elles surchargent la mémoire de travail du lecteur qui doit mémoriser ce qui s'est produit aux étapes précédentes de l'animation pour comprendre les événements suivants (Fabrikant et al. 2008a). A l'opposé, les cartes non animées présentent l'inconvénient de ne pouvoir montrer qu'un nombre limité d'objets ou d'événements, sous peine que les données trop nombreuses se superposent et se masquent.

Les travaux existants avancent que cartes animées et non-animées ne sont pas concurrentes mais visent des objectifs différents (Fabrikant et al. 2008b; Boyandin et al. 2012) : une carte animée permet de détecter plus facilement des ruptures ou des changements ponctuels dans le temps ; les cartes non animées favorisent les observations globales, sur l'ensemble de la période représentée. De nombreuses géovisualisations de récit, disponibles sur internet, proposent à la fois aux lecteurs une carte animée et une vue globale non animée, cette dernière étant placée en début ou en fin d'animation (ex : Visualizing Emancipation¹, Trackin the oil spill², How music travels³, Maps of war⁴).

¹ <http://dsl.richmond.edu/emancipation/>

² <http://www.nytimes.com/interactive/2010/05/01/us/20100501-oil-spill-tracker.html>

³ <http://www.thomson.co.uk/blog/wp-content/uploads/infographic/interactive-music-map/index.html>

⁴ <http://www.mapsofwar.com/ind/history-of-religion.html>

SAGEO'2016 – Nice, 6-9 décembre 2016

3.2. Visualiser les relations en chaîne entre événements

Parmi les cartes interactives disponibles sur internet, au moins deux visualisent explicitement les relations en chaîne entre événements observables au cours du récit. La première décrit le récit de l'inondation de la Nouvelle-Orléans (Etats-Unis), lors du passage de l'ouragan Katrina en 2005⁵. A chaque étape, des flèches animées indiquent la direction d'expansion des eaux d'inondation, tandis que des symboles clignotant signalent les événements de rupture de digue. Ce procédé met l'emphase sur le fait que ces événements sont la conséquence des étapes précédentes de l'inondation. La deuxième géovisualisation est l'application Atlasciné⁶ (Caquard et al. 2012) conçue pour visualiser le déroulement spatial de l'intrigue de films canadiens. La carte animée montre les lieux successifs où se déroule le récit, ainsi que le temps passé dans chaque lieu sous la forme de cercles proportionnels qui s'élargissent au fur et à mesure que le nombre d'événements superposés augmente. Deux localisations successives dans le récit sont reliées par une ligne pleine et les lieux évoqués par les acteurs (i.e. qui sont en relation avec le lieu de l'action en cours mais non visités) lui sont reliés par des lignes pointillées.

Ces deux géovisualisations présentent aussi l'avantage de garder une trace des étapes antérieures de l'animation. Sur la carte de récit de l'ouragan Katrina, les zones inondées s'accumulent au cours du temps et ne s'effacent jamais. Dans AtlasCiné, les cercles proportionnels concentriques témoignent de la durée des localisations d'événements et restent affichés jusqu'à la fin de l'animation. Les lignes de relations entre lieux restent également affichées mais, sous forme synthétique non animée, elles occasionnent un enchevêtrement peu lisible.

3.3. Co-visualisation d'événements ponctuels superposés

Il existe deux grandes catégories de méthodes pour gérer la superposition cartographique d'entités ponctuelles. La première catégorie consiste à agréger les points qui sont trop proches dans l'espace selon l'échelle de la carte. Par exemple, la librairie « Marker Clusterer », pour Google Maps⁷, agrège les points proches automatiquement selon le niveau de zoom et affiche le nombre de points qui ont été fusionnés à chaque localisation (FIGURE 2a). La seconde catégorie de méthodes consiste à déplacer les points de leur localisation initiale, de façon à remplir l'espace graphique plus efficacement. Par exemple, la librairie « Spiderfier » pour Leaflet⁸ montre d'abord l'ensemble des points superposés et, au clic de l'utilisateur, les disperse autour de leur localisation initiale en forme de spirale (FIGURE 2b).

Une autre méthode qui présente l'avantage d'intégrer la dimension temporelle, consiste à décaler les symboles par rapport à leur localisation initiale selon un axe

⁵ http://www.nola.com/katrina/index.ssf/2015/08/katrina_flooding_map.html

⁶ <https://atlascinoproject.wordpress.com/>

⁷ Demo: [http://google-maps-utility-library-](http://google-maps-utility-library-v3.googlecode.com/svn/tags/markerclusterer/1.0/examples/simple_example.html)

[v3.googlecode.com/svn/tags/markerclusterer/1.0/examples/simple_example.html](http://google-maps-utility-library-v3.googlecode.com/svn/tags/markerclusterer/1.0/examples/simple_example.html)

⁸ Demo: <https://github.com/jawj/OverlappingMarkerSpiderfier-Leaflet>

Géovisualisation de récits de catastrophes naturelles et de leurs impacts

vertical et de les ordonner dans l'ordre chronologique. C'est le cas par exemple de deux cartes anciennes (Tufte 1990) présentées sur la FIGURE 3.

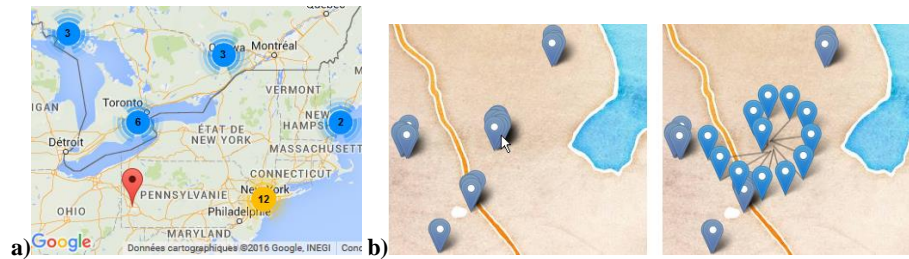
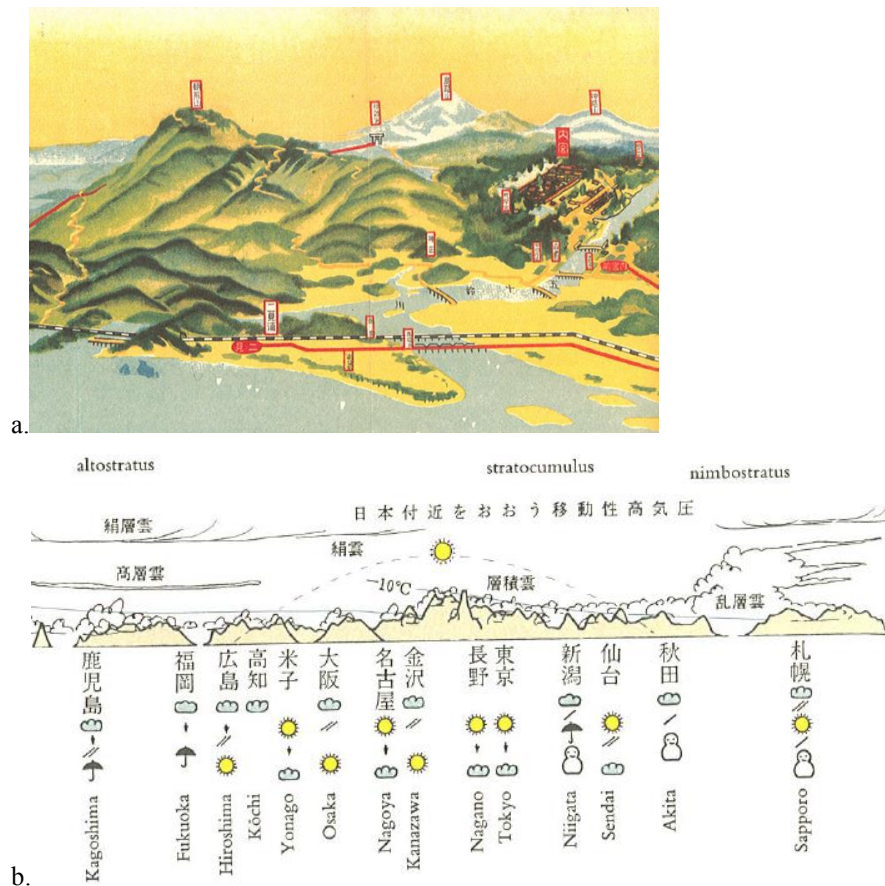


FIGURE 2 a. Exemple de méthode agrégeant les points proches (Marker Clusterer). – b. Exemple de méthode de déplacement relatif des points proches (Spiderfier).



SAGEO'2016 – Nice, 6-9 décembre 2016

FIGURE 3. *a. Extrait de la carte des routes vers le sanctuaire d'Ise. L'orientation des étiquettes de lieux, perpendiculaire au plan horizontal du reste de la carte, permet une précision des localisations et un sentiment d'équilibre visuel. – b. Extrait d'une carte météo sur le Japon. Le changement de conditions météo au cours de la journée est indiqué par des symboles iconiques, qui sont ordonnés chronologiquement de haut en bas (Tufté 1990).*

4. Propositions de géovisualisation du récit d'une inondation historique

4.1. Une simplification des données pour assurer leur co-visualisation

Pour résoudre le problème de co-visualisation d'entités géographiques ayant des géométries différentes et imparfaites, nous avons considérés le barycentre des objets qu'ils soient linéaires ou surfaciques. Les dimensions temporelles étant hétérogènes, nous considérons les événements comme des instants (c'est-à-dire des événements ponctuels dans le temps). Les événements ont également été réduits à une granularité unique, celle de la journée, en distinguant, pour les périodes, un événement début (ex : « début des travaux de réparation ») et un événement fin (ex : « fin des travaux de réparation »).

4.2. Représentation de la date des événements

La date de chaque événement est représentée par les variables couleur et valeur (Bertin 1967). L'échelle de valeur progressive permet au lecteur de distinguer les événements anciens des plus récents et de repérer l'ordre chronologique. En raison du grand nombre de dates, des teintes différentes segmentent les différentes périodes du phénomène (Davoine et al. 2012) : pendant la période hydro-climatique, la valeur de la couleur en double progression est utilisée pour différencier la progression temporelle de crue et de décrue. Les couleurs ont été choisies en fonction de leur symbolique dans la culture occidentale (rouge symbolisant une période de danger et vert un retour au calme). La période de réaction à l'inondation, qui s'inscrit dans le prolongement du phénomène naturel, est représentée par une échelle de couleur proche (sur le cercle chromatique) de celle utilisée pour représenter l'évènement, afin d'être perçue comme une continuité. Le nombre de dates unitaires étant là-aussi trop grand pour différencier l'ensemble des nuances de saturation et d'intensité, les couleurs voisines et successives dans le cercle chromatique ont été utilisées.

4.3. Une visualisation incrémentale des événements

La représentation du récit d'une inondation historique repose sur une cartographie animée de la catastrophe dans laquelle les événements apparaissent les uns après les autres suivant leur chronologie, mais sans restitution de leur durée.

Géovisualisation de récits de catastrophes naturelles et de leurs impacts

Une fois leur date d'apparition révolue, les symboles⁹ correspondant aux événements ne disparaissent pas et restent affichés sur la carte, offrant ainsi une vision complète de la catastrophe (phénomènes et conséquences). Comme plusieurs événements peuvent être successivement survenus à une même localisation, les problèmes de superposition d'information sont évités en déplaçant les événements antérieurs par rapport à leur localisation initiale. En s'inspirant de la méthode illustrée dans la FIGURE 3, les événements survenus en un même lieu sont ici empilés dans l'ordre chronologique, selon un axe (horizontal ou vertical) perpendiculaire à ligne ferroviaire sur laquelle ils sont survenus (FIGURE 4).

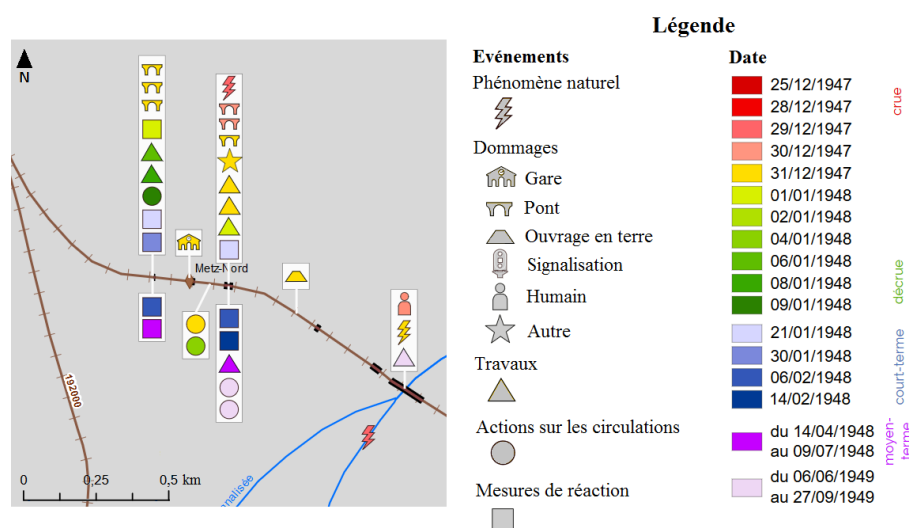


FIGURE 4. Les événements survenus à un même endroit au cours du temps sont ordonnés chronologiquement dans un cadre (le plus ancien de la série est placé en haut).

Ce mode de représentation a pour avantages i) de moins surcharger la mémoire de travail des utilisateurs car il n'est pas nécessaire de retenir les lieux antérieurs où des informations sont apparues ; ii) de faciliter l'affichage des relations de causalités entre événements lorsqu'une cause et sa conséquence ont des dates très éloignées.

4.4. Visualiser les relations de causalités entre événements

Lorsque les événements issus de risques sont liés par une relation de cause à effet, cette relation va être représentée dans la carte par une ou plusieurs lignes qui relient l'événement cause à sa ou ses conséquences. Au cours de l'animation, illustrée sur la FIGURE 5 : 1) la date du jour apparaît, 2) l'événement apparaît par

⁹ La symbologie utilisée ici est fortement inspirée de celle en vigueur dans l'entreprise.

SAGEO'2016 – Nice, 6-9 décembre 2016

un clignotement, qui permet de capter l'attention du lecteur, 3) une ligne de conséquence apparaît reliant l'événement cause à sa conséquence. Elle est parcourue par un figuré jaune qui montre la direction de la relation, 4) l'événement conséquence apparaît par clignotement et 5) et 6) les événements superposés à d'autres vont se mettre en mouvement et s'incrémenter dans des cadres blancs, afin de rester affichés en périphérie de leur localisation réelle. Cela laisse la place d'apparaître à de nouveaux symboles.

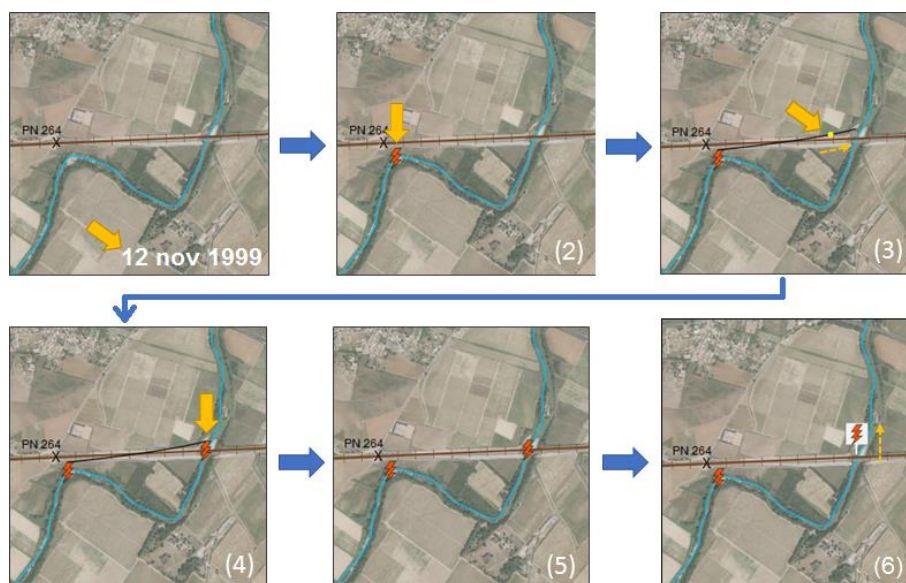


FIGURE 5. Déroulement d'une animation (cas d'un événement ayant une conséquence).

A tout moment, le lecteur peut mettre l'animation en pause et ré-explorer sur demande les relations de cause à effet. Le clic sur un événement affiche les relations de causalité dans lesquelles intervient cet événement grâce à des lignes qui lie les événements entre eux, à la manière d'une carte en oursin. Dans le cas non-animé, la couleur des lignes distingue les causes, reliées par des lignes rouges à l'événement cliqué, et les conséquences, qui lui sont liées par des lignes noires (FIGURE 6). La variable couleur a été utilisée ici mais d'autres méthodes visuelles pourraient être employées, telles que des variations de texture (trait plein, pointillé). L'usage de flèches a également été envisagé mais produit un effet de fouillis lorsque les relations de causalité sont nombreuses.

Géovisualisation de récits de catastrophes naturelles et de leurs impacts

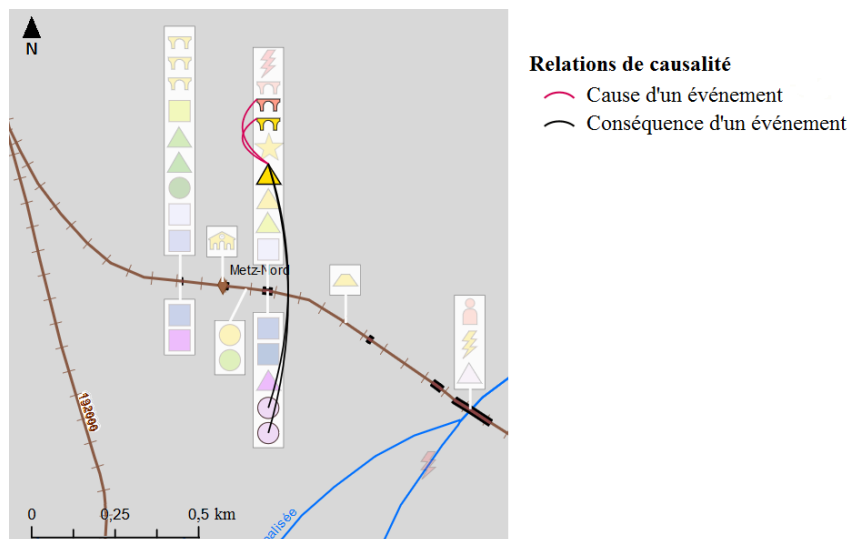


FIGURE 6. Dans la vue non animée, affichage des relations de causes et conséquences d'un événement sur demande grâce au clic.

4.5. Visualisation synthétique du temps

A la fin de l'animation, tous les événements survenus sur la période d'étude sont affichés dans la carte, fournissant une vue synthétique de leur répartition spatiale. La couleur attribuée aux événements renseigne sur leurs dates, mais avoir une vue chronologique globale de l'ensemble de la période est plus difficile, surtout car les événements sont spatialement dispersés. Un graphique en ligne de temps a donc été associé à la carte : il fournit une vue globale de la période étudiée. La ligne de temps a été dessinée au pas de temps du jour, ce qui correspond à la granularité des dates explicitées dans les archives historiques. Néanmoins, la représentation des événements issus d'une inondation dans une ligne de temps soulève deux problèmes.

Premièrement, les événements sont survenus sur de longues périodes de temps et sont concentrés en majorité dans les premières semaines de l'inondation, laissant de longues périodes vides dans le graphique. Pour composer avec ces espacements temporels, la solution adoptée a été de condenser graphiquement les périodes vides les plus longues, tout en leur attribuant une sémiologie différente qui contraste avec le reste de la ligne de temps : dans notre cas, un arrière-plan de couleur gris montre les sauts temporels. Cela permet de les distinguer des périodes « ordinaires ». Les événements à moyen et long-terme survenus entre ces périodes vides, qui sont peu nombreux et eux-mêmes épars dans le temps, sont groupés dans un même intervalle temporel.

Deuxièmement, les événements qui sont survenus le même jour en début d'inondation apparaissent graphiquement superposés dans la ligne de temps. Pour

SAGEO'2016 – Nice, 6-9 décembre 2016

pallier ce problème de superposition deux lignes de temps sont proposées : une régulière, où chaque jour occupe le même espace graphique, et une proportionnelle aux événements, dans laquelle la largeur d'une journée est proportionnelle au nombre d'événements survenus ce jour (FIGURE 7). Elles permettent de mettre en valeur des temporalités différentes, parmi les types de temporalités identifiés dans la littérature (Hewitt 1983; Dibiase et al. 1992; MacEachren and Taylor 1994) :

- La ligne de temps régulière permet de visualiser la date, la durée, la fréquence des événements et l'espacement temporel entre événements.

- La ligne de temps proportionnelle aux événements permet de visualiser la date, le taux de changement et l'ordre des événements. En revanche, contrairement à ce que le lecteur pourrait penser intuitivement, elle ne met pas en évidence la synchronisation entre événements. En effet, les événements survenus un même jour sont classés dans un ordre logique d'apparition mais rien ne traduit la durée de ces événements, c'est-à-dire si plusieurs événements sont arrivés quasi-simultanément (dans un intervalle de quelques secondes) ou sont espacés de dizaines d'heures. Cela doit être bien explicité aux utilisateurs avant tout usage de l'interface pour éviter les fausses interprétations.

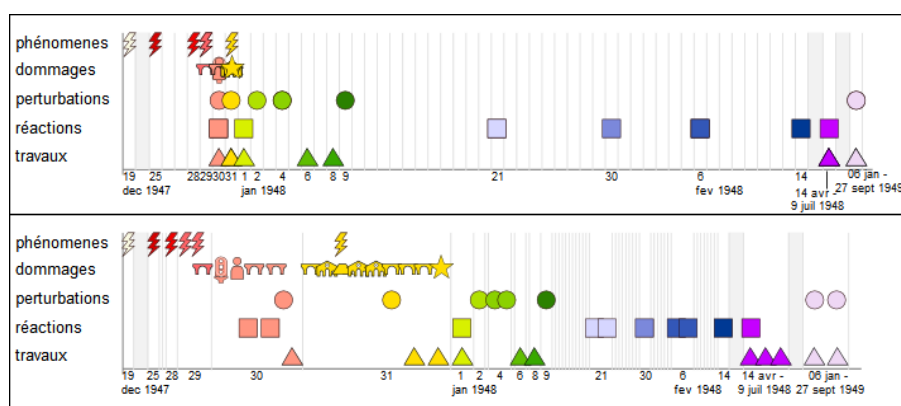


FIGURE 7. Lignes de temps montrant les événements survenus suite à l'inondation de 1947 à Metz. (Haut) Ligne de temps régulière – (Bas) Ligne de temps proportionnelle au nombre d'événements.

5. Conclusion et perspectives

Une méthode a donc été proposée pour naviguer dans le récit d'une inondation majeure, pour visualiser les événements naturels et leurs impacts en chaîne sur le système ferroviaire. Trois contributions principales ont été formulées pour faciliter la visualisation d'informations spatio-temporelles nombreuses. Premièrement, au cours de l'animation, les événements superposés sont déplacés par rapport à leur localisation initiale et sont ordonnés chronologiquement dans des conteneurs

Géovisualisation de récits de catastrophes naturelles et de leurs impacts

surimposés à la carte. L'interactivité de la carte est essentielle pour que l'utilisateur puisse accéder à des informations complémentaires sur les événements. Deuxièmement, la carte donne à voir explicitement les relations de cause à effet entre événements, sous la forme de lignes qui lient les événements entre eux, ce qui peut permettre aux utilisateurs d'explorer et d'analyser ces relations. Troisièmement, une ligne de temps interactive, synchronisée à la carte, fournit une vue synthétique des temporalités associées aux événements. L'application proposée a été implémentée dans un prototype web, développé en utilisant les langages HTML/CSS pour la mise en page, SVG/SMIL pour les vues graphiques et l'animation et Javascript pour l'interactivité. Les fichiers de données sont stockés dans des tableaux JSON.

L'application a été testée auprès de 22 experts du domaine ferroviaire au cours de sessions d'entretiens qualitatifs. L'analyse des tests préliminaires confirme l'intérêt de ces propositions et la possibilité d'extraire à la fois de l'information sur le déroulement de l'inondation, ses caractéristiques phénoménologiques et ses effets dominos. Ils ont également confirmé l'intérêt de combiner carte animée et vue non animée globale sur l'événement : selon eux, la première permet de mieux prendre connaissance de la chronologie des événements tandis que la deuxième permet de retrouver plus rapidement une information cible.

Les propositions émises se basent sur les données relatives aux inondations historiques mais elles pourraient être généralisée à l'étude d'autres types de risques incluant des effets dominos (ex : séismes, risques technologiques). Dans le domaine des inondations historiques, une autre voie d'analyse pour les experts du ferroviaire serait de pouvoir visualiser plusieurs inondations successives simultanément dans la carte afin de pouvoir étudier la répartition spatiale de leurs impacts respectifs. Cela fera l'objet de la suite du projet DHI-Avenir.

Bibliographie

- Benito G, Lang M, Barriendos M, et al (2004) Use of Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation. Review of Scientific Methods. Nat Hazards 31:623–643. doi: 10.1023/B:NHAZ.0000024895.48463.eb
- Bertin J (1967) Sémiologie graphique: Les diagrammes-Les réseaux-Les cartes. EHESS, Paris (France)
- Boyandin I, Bertini E, Lalanne D (2012) A Qualitative Study on the Exploration of Temporal Changes in Flow Maps with Animation and Small-Multiples. Eurographics Conf Vis 2012 31:1005–1014. doi: 10.1111/j.1467-8659.2012.03093.x
- Caquard S, Naud D, Gonzalès R (2012) Esquisses géographiques des récits cinématographiques Canadiens contemporains. Can Geogr 56:508–530. doi: 10.1111/j.1541-0064.2012.00443.x
- Coeur D, Lang M (2008) Use of documentary sources on past flood events for flood risk management and land planning. Comptes Rendus - Geosci 340:644–650. doi: 10.1016/j.crte.2008.03.001
- Davoine P-A, Mathian H, Saint-Marc C, et al (2015) The Visual Representations of

SAGEO'2016 – Nice, 6-9 décembre 2016

- Territorial Dynamics : Retrospective and Input from New Computing Environments. In: 27th International Cartographic Conference (ICC 2015). Rio de Janeiro (Brazil), p 16
- Davoine P, Saint-Marc C, Dimuro A, et al (2012) Constructions cartographiques pour la caractérisation de l'aléa volcanologique du Piton de la Fournaise. In: SAGEO 2012. Liège (Belgique), pp 1–17
- Delacroix C (2016) HISTOIRE (Histoire et historiens) - L'écriture de l'histoire. In: Encycl. Universalis. <http://www.universalis.fr/encyclopedie/histoire-histoire-et-historiens-l-ecriture-de-l-histoire/>. Accessed 24 May 2016
- Dibiase D, MacEachren AM, Krygier JB, Reeves C (1992) Animation and the role of map design in scientific visualization. *Cartogr Geogr Inf Syst* 19:201–214.
- Fabrikant SI, Rebich Hespanha S, Montello DR, et al (2008a) A visual analytics approach to evaluate inference affordance from animated map displays. In: *GeoVA Workshop - GIScience 2008*. Park City (Utah, USA), pp 1–7
- Fabrikant SI, Rebich-Hespanha S, Andrienko N, et al (2008b) Novel Method to Measure Inference Affordance in Static Small-Multiple Map Displays Representing Dynamic Processes. *Cartogr J* 45:201–215. doi: 10.1179/000870408X311396
- Hewitt K (1983) The Idea of Calamity in a Technocratic Age. In: Hewitt K (ed) *Interpretations of Calamity from the Viewpoint of Human Ecology*. Allen & Unwin, Boston (USA), p 304
- Kaddouri L, Blaise J-Y, Davoine P-A, et al (2014) État des lieux des Représentations Dynamiques des Temporalités des Territoires.
- MacEachren AM (1994) *How maps work. Representation, Visualization and Design*. The Guilford Press, New York (NY, USA)
- MacEachren AM, Taylor DRF (1994) *Visualization in Modern Cartography*. Pergamon Press, Inc.
- Minard CJ (1869) *Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie, 1812-1813*.
- Provitolo D (2005) Un exemple d'effets de dominos : la panique dans les catastrophes urbaines. *Cybergeo Eur J Geogr* 19. doi: 10.4000/cybergeo.2991
- Saint-Marc C, Chenier D, Coeur D, et al (2014) Apprendre du Passé pour Optimiser la Prévention et la Gestion des Inondations sur le Ferroviaire. In: *Symposium International Géo-rail 2014*, 6-7 novembre. Marne-La-Vallée (France), p 10 p.
- Tufte ER (1990) *Envisioning Information*. Graphic Press, Cheshire (CO, USA)
- Tversky B, Morrison JB, Betrancourt M (2002) Animation : can it facilitate ? *Int J Hum Comput Stud* 247–262. doi: 10.1006/ijhc.1017
- Vasiliev I (1997) Mapping Time. *Cartographica* 34:1–51. doi: 10.3138/D357-234G-2M62-4373
- Zacks JM, Tversky B (2001) Event structure in perception and conception. *Psychol Bull* 127:3–21. doi: 10.1037/0033-2909.127.1.3

L'approche cartographique de la représentation du mouvement spatial. L'exemple des flux commerciaux maritimes euro-méditerranéens.

Françoise Bahoken¹², Claire Lagesse²³, César Ducruet²³

¹ Institut français des sciences et technologies des transports, des réseaux et de l'aménagement (IFSTTAR-AME / SPLOTT)

² Géographie-Cités / Equipe PARIS (UMR 8504)

³ Institut des Systèmes Complexes - Paris Ile-de-France (ISC-PIF)

RÉSUMÉ. *La cartographie du mouvement spatial implique une réflexion autour de la modélisation cartographique, complémentairement aux données numériques représentées. Elle ouvre ainsi une recherche sur la représentation des interactions et interrelations spatio-temporelles. Les approches existantes agissent soit sur le « fond de la carte » (aspects statistiques du traitement des données, en lien avec leur thématique) soit sur sa « forme », c'est-à-dire sur les aspects graphiques liés au tracé (« edge bundling ») et à la symbolisation (sémiologie graphique) du processus de représentation.*

Cette communication a pour objectif d'examiner le rôle joué par le fond de carte dans la cartographie de flux s'exprimant par voie maritime. Purement géographique, l'approche diffère des précédentes en ce qu'elle approxime la représentation du mouvement en mobilisant un maillage de l'espace des flux. Deux possibilités sont présentées selon la manière dont l'espace – discret ou continu – formé des positions des couples de lieux d'origine et de destination, est mobilisé pour représenter des interactions spatiales (flux et mouvements).

ABSTRACT. *Mapping spatial movement implies to rethink cartographic modeling in addition to digital data issues. This opens research pathways about the representation of spatiotemporal interactions and interrelations. Existing approaches focus either on the map background itself (data analysis issues, to answer thematic questions) or on its "shape", i.e. on the pure graphical aspects of visualization (edge bundling) and on the semiology rules behind it. This paper aims to further examine the role of the map background in maritime flow mapping. This essentially geographic approach differs from existing ones in that it approximates the representation of movement by modeling the mesh of the space of flows. Two possibilities are presented according to how space (discrete or continuous; based on O-D pairs of places) is mobilized to represent spatial interactions (flows and movements).*

MOTS-CLÉS : *matrice origine-destination, flux spatialisé, mouvement spatial, cartographie de flux, maillage territorial, maillage spatial.*

KEYWORDS: *origin-destination matrix, spatialized flow, spatial movement, flow mapping, territorial mesh, spatial mesh.*

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

La cartographie de données de flux par voie maritime est généralement réalisée à l'échelle mondiale, dans le cadre d'approches qui sont soit réticulaires, soit ponctuelles.

Les premières, dans le champ de la théorie des graphes, assimilent la trajectoire des navires à une matrice de liens décrivant un ensemble de relations commerciales (Ducruet, 2015). La cartographie du commerce maritime étatsunien d'Edward L. Ullman (1949) est un des premiers exemples connus de ce genre. Les secondes mobilisent une approche individu-centrée sur les navires, à partir de leurs positions dans l'espace-temps. Depuis la carte pionnière d'André Siegfried (1940) sur les trajectoires des navires de commerce britanniques en 1937, il faut attendre les années 2000 pour que d'autres cartographies de ce type soient produites, à partir de sources diverses comme les carnets de bord (voir Garcia-Herrera et al., 2005) ou les données AIS (*Automatic identification System*) ou VMS (*Vessel monitoring system*) (Le Guyader et al., 2011), à l'instar des recherches sur les trajectoires d'objets mobiles (voir Buard et al., 2015). Enfin, les données satellites comme les AVHRR (*Advanced Very High Resolution Radiometer*) ont également permis de cartographier la densité de navires sur le globe (Tournadre, 2014).

Les secondes considèrent les relations commerciales ainsi que les quantités transférées par couples de ports d'Origine et de Destination (OD). Leur cartographie est toujours réalisée selon une *logique de flux*, à l'exception notable des travaux pionniers de C. J. Minard (1845-1877) qui mobilise un procédé distributif, correspondant davantage à une *logique de mouvement* (Bahoken, 2016). Certaines études ont, par ailleurs, proposé d'utiliser un carroyage pour mesurer la fréquence des flux et produire des cartes de densité (McKenzie, 1975), ou encore les données ponctuelles (AIS) pour proposer une mesure composite de l'accessibilité mondiale des villes (Nelson, 2008).

Dans la *logique de flux* classique, les flux maritimes sont, en effet, généralement représentés par une ligne droite tracée entre les OD, en raison du recours au formalisme euclidien (vol d'oiseau). Le processus de représentation ne prend pas en compte l'espace géographique. Cette pratique est exacte et logique : la définition même du flux ne contient pas cette information. Le flux se définit, en effet, comme le transfert d'une quantité (d'individus ou de déplacements) nécessairement agrégée entre deux OD, son cheminement n'est pas connu avec précision, puisqu'il a été perdu lors des procédures de d'agrégation spatio-temporelles. L'espace sert le plus souvent de support à la représentation ou d'habillage, éludant les nombreuses contraintes topographiques liées aux déplacements. Il en résulte que la cartographie de flux, selon une *logique de flux*, est préjudiciable à la représentation de certains d'entre eux : par exemple, de ceux qui s'expriment à l'échelle mondiale et/ou par voie maritime.

L'approche cartographique de la représentation d'un mouvement

Les flux maritimes, parce qu'ils sont localisés en mer, c'est-à-dire sur une surface continue bornée par un trait de côtes, lui-même ponctué de ports qui sont localisés à terre (en façade maritime), souffrent particulièrement de cette *logique de flux*. Leur représentation conduit à des cartes impossibles où les navires sont supposés franchir des espaces continentaux (Grataloup, 2011). Au-delà de cette question maritime, la spécificité des flux observés selon un mode de transport oblige à considérer l'infrastructure concernée. Il est dès lors intéressant de s'interroger sur la manière dont l'espace des positions (x, y) des OD peut être intégré dans le processus cartographique et cela, quelle que soit l'échelle d'observation. Ce questionnement interroge alors la mise en œuvre d'une *logique de mouvement spatial*, explorée notamment dans le cadre de flux aériens (Bahoken, 2016). Cette logique considère en effet la friction de l'espace (géographique) dans le processus de modélisation cartographique, en complément de sa modélisation numérique.

L'objectif de cette communication est d'explorer différentes possibilités de géovisualisation des mouvements maritimes mondiaux, réduits à l'espace euro-méditerranéen. Nous présentons dans un premier temps un état théorique de la question de la cartographie du mouvement spatial. Celui-ci nous conduit à revenir sur la proposition de J. Bertin (1973) de représenter de manière complémentaire des interactions et des relations spatiales. Dans un second temps, nous recourons à des matrices de flux commerciaux mondiaux élaborées dans le cadre du projet *World Seastems*. La reproduction de cette proposition et son enrichissement ouvrent la voie à de nombreuses variations cartographiques. Dans un troisième temps, nous présentons les possibilités de la cartographie des mouvements spatiaux, en focalisant l'attention sur les aspects carto-graphiques, ceux qui agissent sur le fond de carte, en complément éventuels des aspects graphiques (forme) et carto-statistiques (fond) non traités ici.

2. Etat de la question

La question de la cartographie du mouvement spatial s'inscrit pleinement dans un cadre plus général de réflexion sur l'analyse du déplacement entre deux OD, situées dans un espace donné, pendant une période.

2.1. La cartographie du mouvement spatial

Le mouvement spatial, tout comme le flux, correspond à un changement de position, à un déplacement entre deux lieux symbolisé sur la carte par un lien. Lorsqu'il est valué (enrichi d'un poids), ce lien caractérise un flot figuré par une bande, éventuellement orientée. Ce lien peut se rapporter à un réseau planaire, lorsqu'il existe une infrastructure de transport qui supporte le déplacement, sinon il est non planaire. Dans le premier cas, la mesure est celle d'un trafic effectué selon un mode de transport, routier par exemple, elle ne traduit pas un flux *stricto sensu*.

SAGEO'2016 – Nice, 6-9 décembre 2016

Qu'il soit planaire ou non, la construction cartographique de ce flux consiste à agir sur la seule dimension de *largeur* de la bande, en application du principe de sémiologie cartographique spécifique à la variable visuelle taille. La construction cartographique du mouvement est d'ailleurs similaire. Il est, en effet, possible de démontrer que la cartographie de flux est en réalité une approximation de la représentation du mouvement spatial, la différence entre les deux notions étant imputable à la manière dont l'espace des interactions est pris en compte dans l'analyse et, surtout, dans l'interprétation de la carte (Bahoken, 2016).

Ce questionnement intéresse des aspects sémiologiques, voire numériques de la construction du flux. Il peut être étendu aux aspects cartographiques relevant du fond de carte. En effet, le déplacement (flux, trafic ou mouvement), en tant que caractéristique intrinsèque du couple de lieux (Goodchild et al., 2007), peut également être interprété sur le plan graphique. Partant de là, il est possible de développer un raisonnement théorique et conceptuel sur la signification de la *longueur* du lien (exprimée en mm) perçu sur la carte, et ses conséquences en termes d'interprétation de la figure.

Si l'on considère les modalités de prise en compte de l'espace dans la réalisation d'une carte, la *longueur cartographique* du lien intéresse l'espace des positions des OD, autrement dit l'espacement relatif des lieux d'origine et de destination qui traduit une acception de la *distance parcourue* (Bahoken, 2016) par la quantité en déplacement. Cette posture introduit donc une réflexion sur la prise en compte de la composante spatiale dans la cartographie d'un transfert, du *mouvement spatial* par une action éventuellement menée sur la *longueur* du lien, conjointement à sa *largeur*, déjà mobilisée pour symboliser l'éventuelle valeur de la quantité dont il est question.

Il est important de noter que cette réflexion menée dans le cadre d'analyse des flux s'inscrit dans un champ plus large d'investigations qui a émergé au début des années 1970. Il vise à la prise en compte des concepts de distances-temps dans l'analyse des réseaux de transport. Les réflexions menées par W. Tobler (1979) dans sa thèse, puis par A. L'Hostis (1997), pour la construction de la carte en relief ou chronocarte, ont en effet participé à la prise en compte de métriques spatiales et/ou temporelles dans la construction cartographique. Leurs travaux cartographiques s'inscrivent dans l'objectif d'illustration de recherches théoriques qui ont accompagné le développement des transports et la *révolution de la vitesse*. A. Bretagnolle (1995, 1996) décrit bien l'évolution de la carte qui s'en suit pour illustrer ce changement de paradigme. On observe, en effet, le développement de cartes où l'approche cartographique – c'est-à-dire l'action menée sur les positions des lieux, sur le fond de carte – est privilégiée par rapport à des actions numériques (sur les données statistiques). Cela se traduit par le développement de procédés d'anamorphoses vectorielles (unipolaires et multipolaires) qui permettent de substituer à la métrique euclidienne traditionnelle des métriques réseaux ou de distance-temps ; aussi par des cartes en isolignes où la « [...] *représentation*

L'approche cartographique de la représentation d'un mouvement

symbolique exprime le rapport au temps, par la superposition de lignes d'égal temps de transport, d'égale vitesse ou d'égale accessibilité » (Bretagnolle, 2005 : 56).

Dans ce contexte des années 1970, on peut ainsi facilement imaginer que les réflexions menées par ses contemporains ont influencé J. Bertin (1973) dans le choix de cartes devant illustrer des flux commerciaux, en l'occurrence « *Les relations non orientées* » (Bertin, 1973 : 345), elles sont présentées sur la fig. 1.

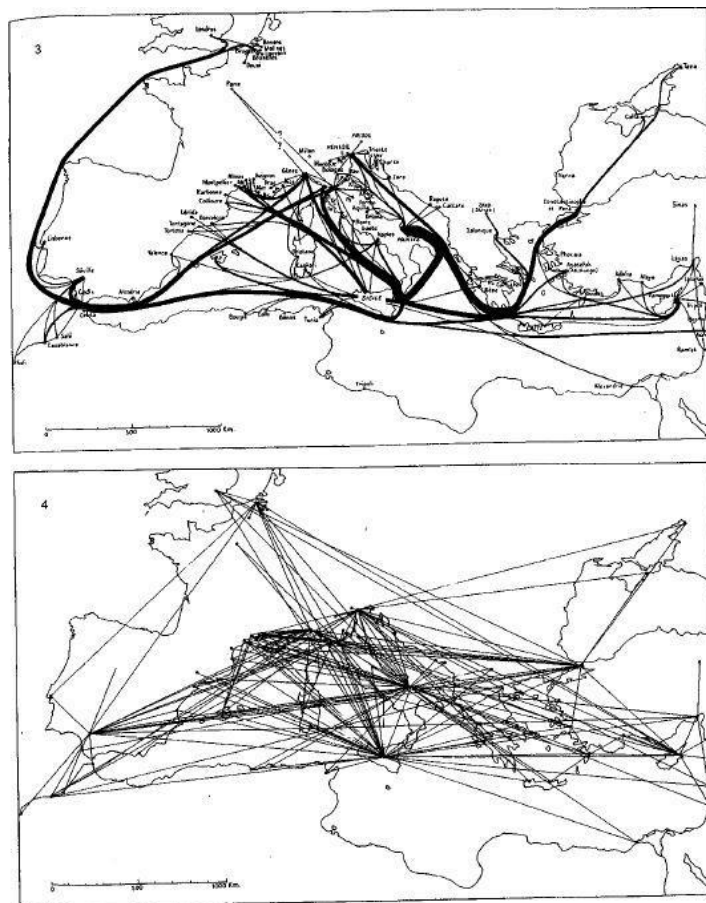


Figure 1. Cartes illustrant des « relations non orientées », d'après (Bertin, 1973)

Les cartes de fig 1. représentent des flux commerciaux sur l'espace méditerranéen.

SAGEO'2016 – Nice, 6-9 décembre 2016

2.2. La proposition cartographique

La présentation de ces deux cartes, en apparence anodine, ne l'est pas en réalité. De quoi est-il question sur ces cartes ? Quelle est la raison pour laquelle Jacques Bertin met ces deux cartes en face à face ? Quelles(s) amélioration(s) ces cartes apportent-elles dans la compréhension des phénomènes cartographiés. Pour y répondre, il convient de revenir à ce qu'indique l'auteur à leur propos.

« (3) *Le commerce du blé en Méditerranée, au XV^e siècle (Mélanges L. FEBVRE, T. II, Paris, A. Colin, 2953). Il ne suffit pas de tracer les itinéraires réellement parcourus pour représenter un système de relation.*

Une carte des routes maritimes, même pondérée, ne montre pas l'orientation commerciale des centres d'activité. Elle montre la densité des bateaux en mer.

Les relations maritimes marchandes entre les villes de l'Europe et de la Méditerranée n'apparaissent seulement dans leur diversité, leur poids et leur orientation géographique que lorsque chaque relation, bien que maritime, est représentée par une droite (4) [sur la carte du bas] » (Bertin, 1973 : 344).

En dépit de l'impossibilité de répondre directement à ces questions, nous pouvons toutefois avancer deux hypothèses (Bahoken, 2016 : 172). La présentation de deux cartes, à visée complémentaire, pour cartographier une information similaire justifie l'idée de décomposer la matrice OD sous-jacente en sous matrices complémentaires, en même temps qu'elle la valide sur des flux affectés sur réseaux (ici des routes maritimes). La seconde hypothèse introduit cette possibilité cartographique de raisonner selon une *logique de mouvement* – en complément éventuel d'une *logique de flux* classique, ou d'une *logique de liens* – pour représenter des interactions spatiales (ici, la seule route de navires marchands en Méditerranée sur la carte du bas de la Fig. 1 : l'information sur les flux ayant été dégradée à la seule existence d'une relation entre les ports).

Quoiqu'il en soit, cette proposition d'association de cartes élaborées selon deux logiques de raisonnement différentes, pour représenter des liens (approche Théorie des graphes) et des mouvements spatiaux (approche gravitaire), apparaît fondamentale. En effet, par rapport aux travaux antérieurs, notre analyse diffère en ce sens qu'elle ne porte pas tant sur la prise en compte du temps de transport (ou de la vitesse) pour générer un espace relatif, cognitif... que sur l'expression raisonné des déplacements dans un territoire, compte tenu des contraintes liées à la représentation. En effet, nous raisonnons sur un espace géométrique assorti d'une métrique géographique et nous mobilisons une matrice des positions relatives dans un système de projection cartographique donné. L'objectif n'est pas tant de montrer la proximité des lieux liés à l'utilisation de tel ou tel mode de transport que de figurer le déplacement entre deux lieux – lié ou non à un mode de transport donné,

L'approche cartographique de la représentation d'un mouvement

compte tenu des contraintes physiques ou topographiques de déplacement. C'est pourquoi notre approche considère le rôle joué par la friction de l'espace (i.e. le frein de la distance) dans l'expression des mouvements spatiaux. Elle met en évidence d'une part l'importance théorique du détour et sa figuration sur une carte (Lhostis, 2015) et, d'autre part, valide par la même l'hypothèse de son existence. C'est pourquoi il ne s'agira pas tant ici de trouver le chemin le plus court sur le réseau formé des déplacements des navires que d'explorer les différentes manières de représenter sur une carte le déplacement entre deux OD, compte tenu des possibilités techniques et technologiques actuelles.

Le cas d'application est illustré par une extraction de la base de données issue des fichiers de la *Lloyd's List* (2008). L'objectif principal de ce travail est ainsi d'enrichir la compréhension d'un des plus importants vecteurs d'échanges de l'économie-monde contemporaine, tout en apportant une lumière nouvelle sur les dynamiques locales, régionales et mondiales qui l'animent. Mieux visualiser la répartition des volumes et densités des flux maritimes se doit cependant de répondre à un certain nombre de questions méthodologiques portant à la fois sur les données et les outils permettant de telles analyses.

3. Présentation des données

D'après Francis Galton cité par A. Bretagnolle (2005), « *la ligne directe entre deux points de l'océan est rarement [jamais ?] la route la plus rapide pour les bateaux de commerce* ».

3.1. La base de données *Lloyd's List* (2008)

Les données sur lesquelles repose cette étude proviennent de la *Lloyd's*, compagnie d'assurance anglaise consignante les déplacements de l'ensemble des navires recensés à l'échelle mondiale, depuis le XVIII^{ième} siècle. Notre extraction sur la base de quatre mois continus en 2008 détaille, pour chaque déplacement d'un navire : le port précédent (lieu d'origine), le port intermédiaire (lieu de transit) et le port suivant (lieu de destination) ; pour chaque navire : son nom, son type de cargaison, l'opérateur, le pavillon, etc. ; et pour chaque port : son nom, son pays, coordonnées géographiques.

Nous avons construit, à partir du graphe des trajectoires, une matrice globale correspondant à cette échelle spatio-temporelle. Les nœuds sont les ports et il existe un arc entre deux nœuds si au moins un navire a réalisé la liaison entre les deux ports durant la plage de temps observée (l'année 2008). Cette matrice globale peut être découpée en catégories, les liens peuvent en effet être valués selon différents attributs, à savoir :

- le nombre de navires ayant effectué le trajet entre chaque couple de ports ;
- le tonnage total de marchandises échangées entre chaque couple de ports ;
- la distance (orthodromique) du trajet ;

SAGEO'2016 – Nice, 6-9 décembre 2016

La matrice globale est donc formée de cinq champs : les identifiants des lieux d'origine (i) ; de destination (j) ; la distance kilométrique parcourue (Dij), une fréquence de navires circulant sur le trajet (Xij) et une quantité totale de marchandises transférée (Fij), entre (i) et (j), exprimée en tonnes. Ces deux derniers champs expriment un flux échangé entre deux ports.

3.2 Décomposition des flux maritimes

La matrice globale étant orientée, nous l'avons décomposée de manière à ne considérer que sa composante symétrique, qui correspond à la représentation d'un volume bilatéral d'échanges (Bahoken, 2016a). Le graphe correspondant n'est donc plus orienté : seule une bande représentant les quantités transférées entre deux ports sera donc représentée.

Les deux cartes suivantes représentent les flux de la *Lloyd* à la manière de la Fig.1. Le graphe étant assez dense (plus de 24 000 liaisons OD sont disponibles et représentées), un traitement sémiologique modifiant la valeur d'opacité de la teinte noire des figurés a été appliqué afin que les motifs de ces flux soient lisibles (Bahoken, 2014). La Fig. 2 représente ainsi les relations interportuaires euro-méditerranéennes existantes et la Fig. 3 décrit des interactions spatiales, par des bandes de largeur proportionnelle au volume bilatéral transféré entre les ports.

Figure 2. Cartographie des liaisons commerciales

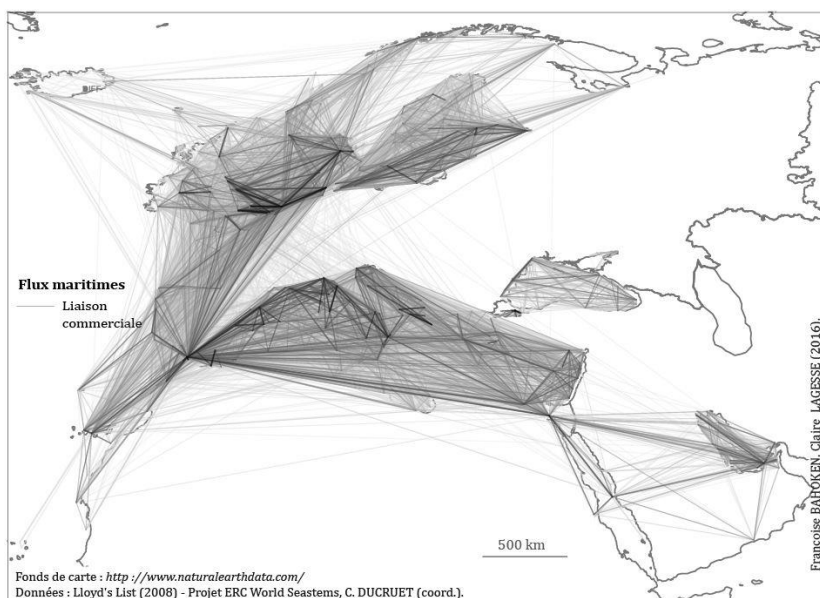
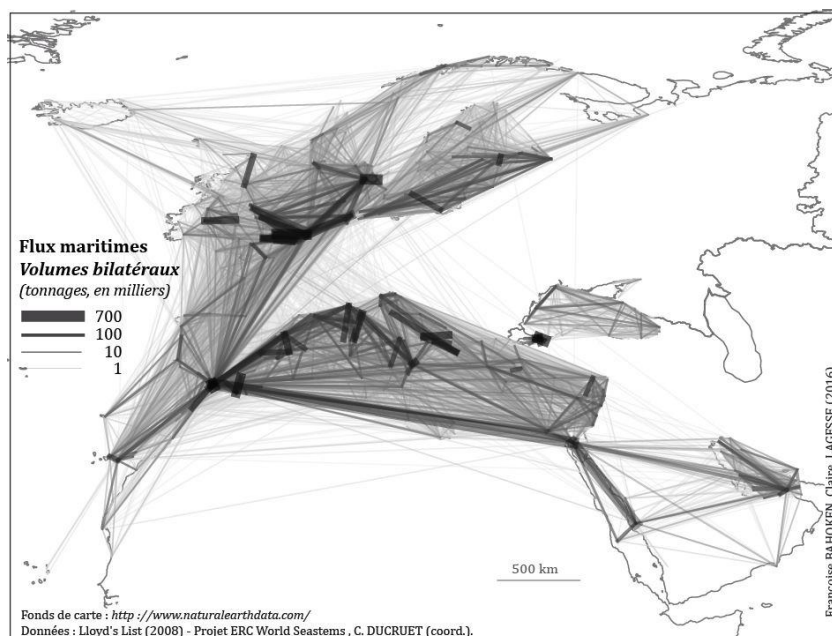


Figure 3. Cartographie des flux maritimes

Ces deux cartes sont générées dans le cadre de *Logique de liens* (Fig. 2) et de *flux* (Fig. 3), selon le formalisme euclidien. C'est d'ailleurs la raison pour laquelle les flux sont représentés par un trait droit : les voies maritimes n'étant pas connues avec précision. Elles ne tiennent donc pas compte de l'espace d'expression de ces flux, en particulier d'éventuelles contraintes physiques (topographiques) au transport maritime qui les caractérise. C'est pourquoi elles sont assez difficiles à analyser. Tout au plus peut-on identifier certains *hubs* portuaires, comme en Italie du Sud (Gioia Tauro) ou au Moyen Orient (Dubai) et certains passages stratégiques, comme Gibraltar ; enfin des zones de concentration (Benelux).

Il est important de noter que les deux cartes de la Fig. 1 décrivent bien la même information : des échanges commerciaux de blé. Elles diffèrent toutefois sur le plan de la sémantique : la première représente un volume total de mouvements spatiaux, tandis que la seconde décrit des relations (L_{ij}) commerciales entre les ports méditerranéens, tout comme la Fig. 2. De ce point de vue, on comprend que pour décrire correctement les mouvements maritimes commerciaux de la Lloyd's, il convient de changer de logique de raisonnement ou de formalisme spatial.

Nous proposons d'analyser les différentes possibilités de géovisualisation du déplacement de ces navires dans la section suivante.

SAGEO'2016 – Nice, 6-9 décembre 2016

4. Les possibilités cartographiques de la représentation des mouvements maritimes de navires

La représentation cartographique du mouvement observé dans un espace géographique peut être réalisée selon trois grandes familles de possibilités qui sont numériques, graphiques et cartographiques. Elles caractérisent une action qui porte respectivement sur le fond de la carte (l'information thématique), la forme de la carte et le fond de carte (l'information géographique).

La première possibilité est à manier avec précaution, car elle mobilise les données numériques de manière à privilégier l'espace des positions pour renforcer la perception du rôle joué par les contraintes au déplacement. Etant donné qu'elle conduit à la représentation de flux présentant des valeurs faibles mais de fréquences élevées, elle entraîne une survalorisation des mouvements importants qui s'expriment plutôt à longue distance, ce qui n'est pas le cas ici puisque l'espace est restreint. C'est pourquoi elle ne sera pas examinée.

La seconde possibilité caractérise une action réalisée sur les seuls aspects graphiques et/ou sémiologiques, pour donner une illusion du cheminement dans l'espace. Elle appartient à la famille des fusions graphiques de type *edge Bundling* (Holten et al, 2005) qui, si elles sont intéressantes, souffrent de deux problèmes : elles ne sont que visuelles et, de ce fait, le résultat obtenu n'est pas adapté à la représentation des mouvements maritimes à l'échelle mondiale, l'approche conduisant à la réalisation de cartes impossibles (typiquement : des flux par voie maritime qui franchissent des continents).

Pour qu'elle soit réellement cartographique, et dotée d'une signification thématique en rapport avec ces flux maritimes mondiaux, cette approche graphique nécessite de prendre en compte l'espace : le fond de carte qui symbolise le territoire d'expression de ces flux. L'une des raisons majeures tient à un *effet de bords* inverse qui conduit à la manifestation *d'effets d'alignements* et *d'effets d'itinéraires* caractéristiques des cartes de flux et mouvements mondiaux (Bahoken, 2016). Cet *effet de bords* est lié au fait que les ports sont situés en bordure d'une surface fermée mais non nécessairement continue que franchissent des navires. Les positions des ports, délimitant la zone d'étude, y sont d'autant plus importantes qu'elles ne définissent pas son centre, mais ses marges, à la différence des analyses portant sur des lieux. Lorsque l'espace est pseudo fermé, comme dans le cas de la méditerranée, il devient donc impératif de tenir compte des marges de la zone d'étude

La troisième possibilité consiste donc à considérer la géographie des lieux d'origine et de destination de ces flux dans le cadre d'une approche spécifiquement carto-graphique.

L'approche cartographique de la représentation d'un mouvement

4.2. L'approche carto-graphique de représentation du mouvement

L'approche carto-graphique consiste à agir sur la position des lieux, par conséquent sur le fond de carte, en lien avec la thématique des données. Sur le plan théorique, cela revient à considérer indirectement le mode de transport, voire l'infrastructure, à considérer le cout généralisé de déplacement. Sur le plan cartographique, la zone d'expression (de représentation) des flux sera ainsi réduite aux portions terrestres pour les modes ferroviaires ou routiers, ou aux portions maritimes ou fluviales.

Il existe des flux marchands variés soumis à différents types de contraintes de déplacement. Lorsque les flux maritimes sont limités à l'espace navigable, les flux routiers sont restreints à l'espace terrestre. Quant à eux, les flux aériens sont émancipés de ces contraintes physiques (ce qui ne signifie pas l'absence de voies aériennes). Tous sont cependant soumis, dans leur représentation, à un effet lié au choix du système de projection.

Dans le cas des flux maritimes, l'effet de proximité entre les ports induit une contrainte forte. Il n'est pas possible de transformer l'implantation des lieux, autrement dit d'assimiler les points d'OD à des zones qui en seraient représentatives (de la même manière que les barycentres correspondent au centres de mailles politico-administratives, ...) puisque l'espace maritime de circulation des navires leur est exclu (les ports ne flottent pas). Les flux maritimes s'expriment en effet sur un espace par essence continu (les mers et océans), non discrétisé *a priori*, il ne peut l'être facilement *a posteriori*. Cette contrainte revêt d'autres avantages : l'intérêt de n'être pas assujéti à un maillage prédéfini, ce qui soulève la possibilité de mobiliser un maillage théorique pour partitionner l'espace.

4.2.1. Les possibilités de partitionnement des mers et des océans

Pour agir sur le fond de carte, c'est-à-dire sur les mers et les océans, il convient de modifier la surface continue qu'ils représentent de manière à ce que l'espace des positions (x,y) des ports, situés sur ses bords soit affecté.

En, règle générale, cette méthode passe par un changement de système de projection cartographique ou par la mise en œuvre de déformations morphologiques ou de modifications des positions des lieux.

En faisant le choix de conserver la géographie des ports, il est nécessaire de changer de paradigme, en discrétisant l'espace maritime. Nous réalisons pour cela une partition des mers et océans qui servira de maillage pour la reconstitution d'itinéraires, des principaux points des passages devant permettre de reconstruire puis d'incurver le trait figurant le mouvement de navires.

Plusieurs possibilités sont envisageables :

– l'utilisation d'itinéraires empiriques (par exemple des « rails maritimes ») pour contraindre la circulation des navires dans des couloirs bien délimités. C'est l'une

SAGEO'2016 – Nice, 6-9 décembre 2016

des approches mises en œuvre par D. Le Guyader et al. (2011) pour identifier les principales « zones de navigation » des navires de charge dans la rade de Brest.

- l'utilisation de découpages existants, de zones de circulation règlementée, pour segmenter l'espace en mailles, généralement hétérogènes. Dans ce cas, l'objectif serait de relier les barycentres pour reconstituer l'itinéraire le plus court entre deux OD. Le site internet *Marine Regions* propose par exemple une liste intéressante de régions des mers/océans (<http://www.marineregions.org/sources.php>), de même que le découpage en eaux territoriales ou celui de la FAO¹. O. Joly (1999), par exemple, avait utilisé le découpage du monde en zones REEDS pour mettre en œuvre l'approche des réseaux maritimes *via* la théorie des graphes qu'il propose.
- L'utilisation de maillages théoriques, générant des partitions hétérogènes de l'espace maritime sur lesquels il est possible d'agir.

4.2.2. L'utilisation de maillages théoriques

Plutôt qu'utiliser un découpage existant, orienté selon la problématique sous-jacente à sa création, il est possible de recourir à des maillages théoriques.

Différents types de maillages en fonction de la forme : régulières ou irrégulières, carrées, rectangulaires, nids d'abeille, existent dans la littérature. D. Dorling (1991) a beaucoup discuté dans sa thèse des avantages comparés de différents maillages terrestres. A notre connaissance, peu d'auteurs semblent avoir segmenté l'espace maritime dans l'objectif de reconstituer des trajectoires.

La méthode développée par Lagesse et al. (2016) consiste à construire un maillage théorique de la surface continue des océans. À partir d'un tirage de points aléatoire sur les mers et océans, plus ou moins dense selon la distance à la côte, une triangulation de B. Delaunay (1934) est construite. Un graphe est ainsi généré dont les sommets sont constitués par le tirage de points initial, et les arcs sont issus du calcul de triangulation. Le graphe est ensuite complété, en s'assurant que les connections sont faites aux différents canaux ou détroits ainsi qu'aux limites de la zone d'étude, différents développements² complémentaires dans ce sens sont actuellement en cours.

Chaque port est ensuite associé à un nœud de ce graphe, selon sa position géographique. Les chemins les plus courts sur le graphe sont enfin calculés entre chaque couple de ports, en utilisant l'algorithme de Dijkstra (Dijkstra, E. W., 1959). Le graphe prédéfini permet d'agrèger les trajectoires sur les routes maritimes ainsi construites. Elles se révèlent proches de celles empruntées par les navires aujourd'hui, selon les observations des données AIS.

¹ Organisation des nations unies pour l'alimentation et l'agriculture.

² A noter que nous n'avons pas retenu de solution liée à l'agrégation car elle aurait abouti à la génération de ports "fictifs" dont il faudrait définir arbitrairement la position géographique. Nous perdriions alors en précision de l'information géographique représentée.

L'approche cartographique de la représentation d'un mouvement

Nous présentons trois cartes complémentaires décrivant les mouvements maritimes enregistrés par la Lloyd's, respectivement la fréquence de lignes (Fig. 4), la quantité transportée (tonnage) sous une forme discrète (Fig. 5) puis continue (Fig. 6). On observe que la Fig. 4 (à la différence de la Fig. 3) met en évidence les grandes artères et corridors du trafic maritime, notamment le fait que ces deux entités n'ont pas de connexion directe à ce niveau de trafic (75%). Il est intéressant de noter que ce résultat valide des travaux antérieurs selon lesquels 80% du trafic maritime mondial réalisé au niveau des arcs interportuaires se concentre dans des arcs dont la distance (orthodromique) est inférieure ou égale à 500 km, résultats par ailleurs logiques puisque les gros flux sont généralement localisés, ils s'expriment davantage à courte distance.

Figure 4. Cartographie de la fréquence des mouvements maritimes

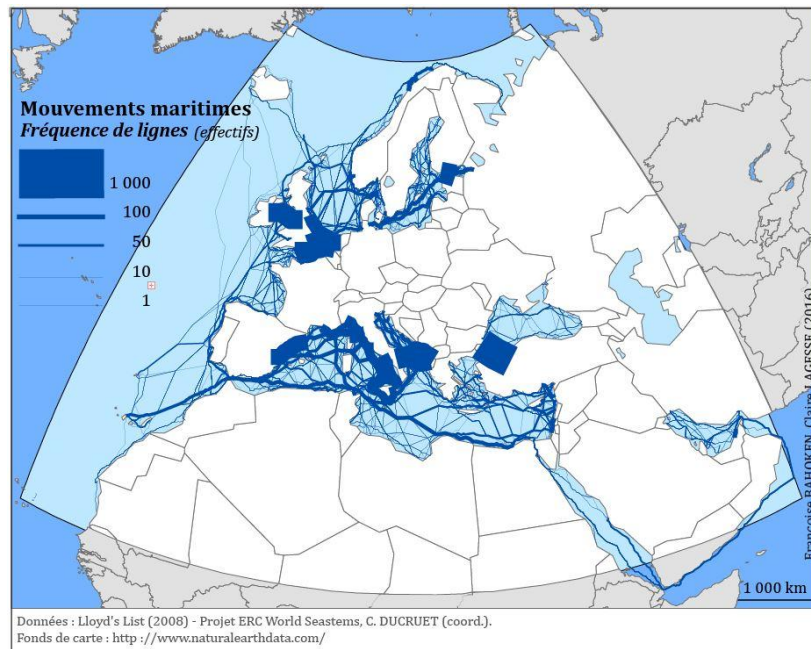
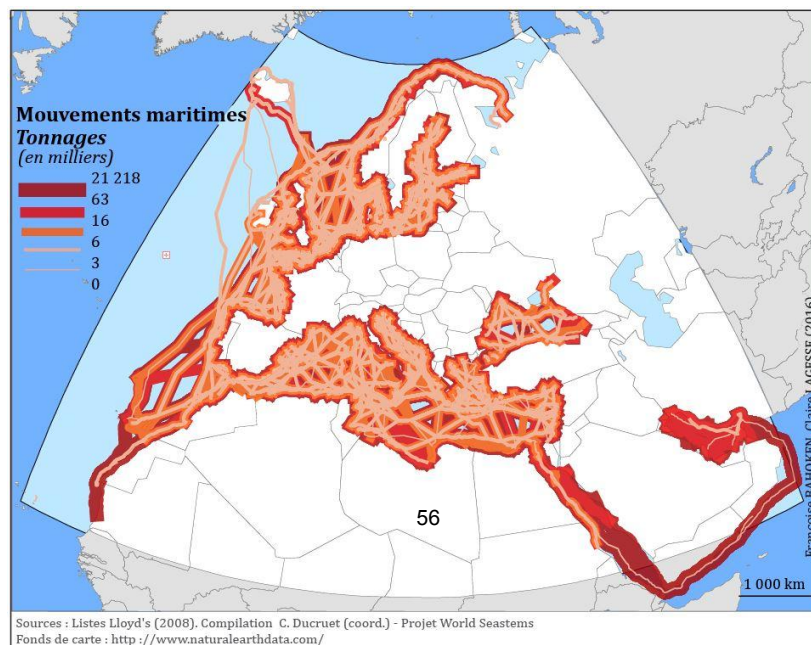


Figure 5. Cartographie des volumes de mouvements maritimes (tonnages, en milliers)



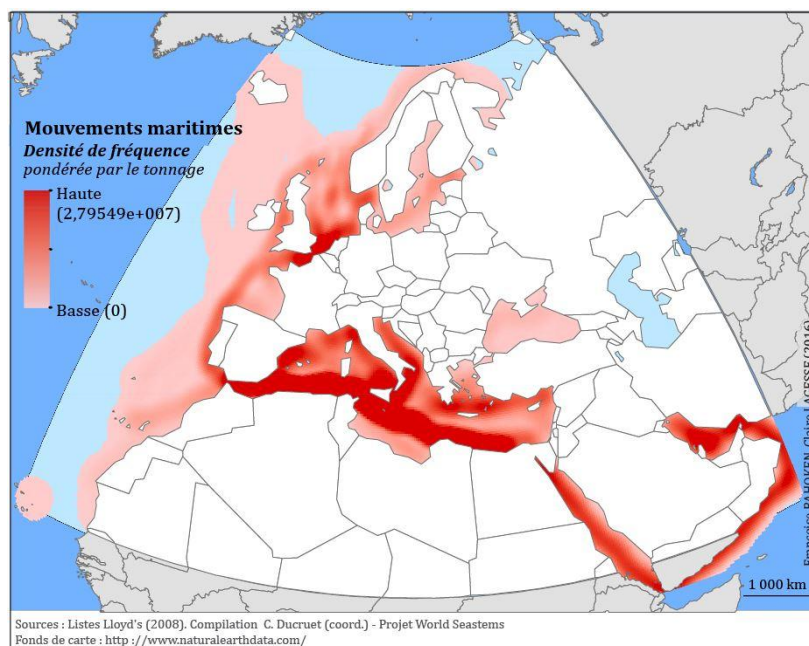
SAGEO'2016 – Nice, 6-9 décembre 2016

Sur la Fig. 4, hormis le fait que la carte puisse être faussée par certains points de passage comme Suez ou Istanbul, elle revêt tout de même un intérêt thématique. Elle décrit effectivement des zones très locales de concentration des mouvements autour de grands pôles comme Barcelone, le Benelux, où le trafic est massif et fréquent ; des zones de forte fréquence, mais au trafic moins massif, comme la Finlande-Estonie, le Royaume-Uni et l'Irlande (transport maritime à courte distance à forte fréquence).

La Fig. 5 est plus difficile à lire en l'état, en raison de la masse de figurés qu'il serait intéressant de filtrer (mais ce n'était pas l'objet ici). C'est pourquoi elle peut être considérée comme une carte intermédiaire entre les versions a) et c). Elle souligne toutefois l'importance du couloir massifié de Suez / mer Rouge. Cela étant, le fait de représenter l'ensemble des classes de valeurs de mouvements procure un effet de lissage et d'homogénéité du trafic en dehors de ces zones.

La Fig. 6 correspond à l'aboutissement de cette analyse cartographique.

Figure 6. Cartographie de la densité de mouvements maritimes (*densité de fréquence pondérée par le tonnage, en milliers*)



Cette Fig. 6 montre - et bien mieux qu'ailleurs - les lignes de force de l'espace maritime euro-méditerranéen avec, cette fois, le lien conséquent nord-sud atlantique reliant le couloir méditerranéen de transit au nord scandinave. Elle illustre comment l'Europe est irriguée par les flux maritimes et le caractère fortement inégal de cette

L'approche cartographique de la représentation d'un mouvement

desserte, de même que la forte dépendance Europe-Asie avec le cordon ombilical de Suez.

Conclusion

A l'échelle mondiale, différents *effets de position des lieux* (ici des ports) et *d'itinéraires*, lié à l'alternance terres-mers entre les couples de lieux empêchent généralement la perception claire des motifs des flux. S'il a notamment été résolu par un choix adéquat du système de projection approprié, dès les années 1920, le procédé distributif de C. Minard – qui consiste à représenter la trajectoire des navires – apportait déjà une solution efficace.

On montre ici l'importance des choix cartographiques, leur caractère essentiel car ils se répercutent au niveau thématique. Il est important de souligner que le projet *World Seastems* est né, en grande partie, du constat selon lequel aucune cartographie de l'évolution des grandes routes maritimes mondiale n'existait à ce jour, malgré le fait que le transport maritime concentre non moins de 90% du volume des échanges mondiaux. La disponibilité de données historiques, comme celles de la Lloyd's, majoritairement des matrices OD, repose la question du choix et de l'adéquation des méthodes et procédés de visualisations de ces mouvements maritimes qui seraient les plus à même de retracer, de la façon la plus pédagogique possible, l'évolution du commerce maritime mondial au cours des siècles passés.

Les analyses présentées ici ont été focalisées sur la représentation du mouvement des navires dans l'espace maritime, elles n'ont volontairement pas examiné les questions liées à la représentation des ports. Une *logique de lieux* focalisant l'attention sur les nœuds aurait pu être introduite, par exemple pour mettre en évidence ceux qui centralisent le réseau. En effet, les graphes (Fig. 2 et Fig. 3), inexploitable, en l'état, de ce point de vue, révèlent toutefois l'importance de certains points de passage. A l'inverse, sur les cartes des Fig. 5 et 6, les dits nœuds sont dilués dans le mouvement d'ensemble. Se pose alors la question de savoir s'il ne faudrait pas au préalable extraire le réseau de certains nœuds, avant d'en analyser les relations.

6. Bibliographie

Bahoken, F. (2016) *Contribution à la cartographie d'une matrice de flux*, Thèse de doctorat en Géographie - Sciences des Territoires, Université Denis-Diderot, Paris 7.
URL : <https://halshs.archives-ouvertes.fr/tel-01273776>

Bahoken, F. (2016a) L'approche cartographique de la décomposition des matrices de flux, *M@ppemonde*, n°116.
URL : <http://mappemonde.mgm.fr/num44/articles/art14404.html>

Bahoken, F. (2014) Pour le renouvellement de la sémiologie de la carte de flux, *Revue Cartes et géomatique du comité français de cartographie*, n°, pp. 109-121

SAGEO'2016 – Nice, 6-9 décembre 2016

Bertin, J. (1973) *Sémiologie graphique. Les diagrammes – Les réseaux Les cartes*, Mouton-Gauthier-Villars, 2ème édition, Les réimpressions de l'EHESS (1998), pp. 344-345.

Bretagnolle, A. (2005) De la théorie à la carte: histoire des représentations géographiques de l'espace-temps, in : Volvey A. (dir.), *Echelles et temporalités*, Atlande, Clefs Concours : Géographie thématique, pp.55-60.

URL : https://halshs.archives-ouvertes.fr/file/index/docid/156769/filename/Theorie_a_carteHAL.pdf

Buard, E., Devogele, T., Ducruet, C. (2015) Trajectoires des objets mobiles dans un espace support fixe. *Revue Internationale de Géomatique*, 25(3), pp. 331-354.

Delaunay, B. (1934) Sur la sphère vide. *Bulletin de l'Académie des Sciences de l'URSS, Classe des sciences mathématiques et naturelles*, 6, pp. 793–800.

Dijkstra, E. W. (1959) A note on two problems in connexion with graphs, *Numerische Mathematik*, 1, pp. 269–271. doi:10.1007/BF01386390.

Ducruet, C. (2013) Histoire maritime et cartes en ligne, *M@ppemonde*, n°109, Rubrique Internet. URL : <http://mappemonde.mgm.fr/num37/internet/int13101.html>

Ducruet, C. (dir.) (2015) Maritime networks. Spatial Structures and Time Dynamics. *Routledge Studies in Transport Analysis*.

Garcia-Herrera, R., Können, G., Wheeler, D., Prieto, M.R., Jones, P., Koek, F. (2005) A climatological database for the world's oceans 1750-1854. *Climatic Change*, 73, pp. 1-12.

Goodchild M. F., Yuan M., Covas J. T. (2007) Towards a general theory of geographic representation in GIS, *International Journal of Geographic Information Science*, 21(3), pp. 239-260.

Grataloup, C., (2011) *Représenter le monde*, La Documentation Française, La Documentation photographique – Les dossiers, n° 8084, 64 p.

Guyader (Le), Brosset, D., Gourlemon, F. (2011) Exploitation de données AIS pour la cartographie du transport maritime, *M@ppemonde*, n 104, 15 p. URL : <http://mappemonde.mgm.fr/num32/articles/art11405.pdf>

Joly, O. (1999) *La structuration des réseaux de circulation maritime*. Thèse de doctorat en Aménagement du Territoire, Université du Havre, Le Havre.

Lagesse, C., Mermet, E., Petit, S., Ducruet, C. (2016) Geovisualisation of the worldwide maritime network over 30 years. Challenges and preliminary results,

L'approche cartographique de la représentation d'un mouvement

Proceedings of the 2nd International Workshop on Maritime Flows and Networks (WIMAKS'16), Session 1: Géovisualization of Maritime Networks, Paris 25-26 april 2016.

McKenzie, F.D. (1975) *Maritime Dynamic Traffic Generator. Volume III: Density Data on World Maps*. Working Paper No. AD-A012 498, *Transportation Systems Center*, Cambridge, Massachusetts.

Nelson, A. (2008) *Travel time to major cities: A global map of accessibility*. Global Environment Monitoring Unit, Joint Research Centre of the European Commission, Ispra, Italy.

Siegfried, A. (1940) *Suez, Panama et les routes maritimes mondiales*. Paris, Armand Colin.

Tournadre, J. (2014) Anthropogenic pressure on the open ocean: The growth of ship traffic revealed by altimeter data analysis, *Geophysical Research Letters*, 41(22), pp. 7924-7932.

Ullman, E.L. (1949) Mapping the world's ocean trade: A research proposal, *The Professional Geographer*, 1(2), pp. 19-22.

Zimmermann, M. (1942) Compte-rendu, *Etudes Rhodaniennes*, 17(3), pp. 183-185.

Proposition méthodologique pour la visualisation et l'exploration des mobilités géographiques

Emilie Lerond¹, Jean-Philippe Antoni¹, Olivier Klein²

1. *ThéMA UMR 6049 CNRS - Université Bourgogne Franche-Comté*
32 rue Mégevand, 25000 Besançon ; 4 boulevard Gabriel, 21000 Dijon
(prenom).(nom)@univ-fcomte.fr

2. *Luxembourg Institute of Socio-Economic Research, 11 Porte des Sciences,*
4366 Esch-sur-Alzette
olivier.klein@liser.lu

RÉSUMÉ. Les données de mobilité quotidienne sont multidimensionnelles et désagrégées, ce qui les rend difficiles à manipuler et à analyser. Bien que des méthodes de représentation et d'analyse aient déjà été proposées, aucune d'elle ne semble parvenir à faire émerger efficacement de la connaissance depuis ces données. Nous proposons de combiner ces solutions et une méthodologie de ré-agrégation innovante des données pour visualiser et explorer les données de mobilité et en faire émerger de l'information.

ABSTRACT. Daily mobility data are multidimensional and disaggregated, what make them hard to manipulate and to analyze. Although methods of representation and analyze have been proposed, none of them seem to be able to efficiently bring out knowledge from data. We propose to combinate this solutions and a methodology of innovative reaggregation to visualize and explore mobility data.

MOTS-CLÉS : mobilités quotidiennes, agrégation, visualisation, espace-temps, données désagrégées

KEYWORDS: daily mobility, aggregation, visualization, space-time, disaggregated data

Proposition pour la visualisation des mobilités

1. Introduction**1.1. Contexte**

Les sources de données de mobilité individuelle se sont diversifiées, aussi bien du côté des outils de collecte (smartphones, objets connectés, capteurs) que de celui d'applications (applications téléphoniques, simulation spatiale). Ces données partagent en général le fait d'être désagrégées et multidimensionnelles : elles se caractérisent par des attributs à la fois spatiaux (localisation, géométrie), thématiques (motif, mode de déplacement) et temporels (heure de début et de fin, durée) qui décrivent les déplacements des individus. Leur volume et la structure de leurs attributs les rendent difficiles à manipuler. On peut toutefois penser que malgré leur complexité, elles offrent, du fait de leur nature, une possibilité nouvelle pour comprendre les mobilités et identifier les règles comportementales qui en sont à l'origine. Nous cherchons justement à traiter et à structurer ces données pour en faire émerger de l'information par diverses méthodes d'analyse, d'exploration et de visualisation des données.

Plusieurs solutions ont déjà été proposées pour représenter des données géographiques spatio-temporelles, en agrégeant tout ou partie de leurs attributs. D'un côté, en conservant l'ensemble de leurs attributs, la *Time Geography* se heurte à des difficultés lorsqu'un jeu de données comporte un grand nombre de déplacements : l'aquarium spatio-temporel de Hägerstrand perd son efficacité au-delà d'une cinquantaine de déplacements (Hägerstrand, 1982; Chardonnel, 2007; Klein, 2007; Kaddouri *et al.*, 2014). De l'autre côté, en conservant uniquement leurs attributs spatiaux, la représentation des déplacements par champs vectoriels, qui réduit le nombre de flux grâce à une transformation mathématique, permet de dégager des tendances générales depuis un grand volume de déplacements en les synthétisant (Tobler, 1978; Gerber *et al.*, 2012); cependant, cette représentation conduit à un lissage des données qui ne permet plus de distinguer les caractéristiques thématiques individuelles. Dans un cas comme dans l'autre, ces méthodes présentent un intérêt, mais elles ne restituent pas l'ensemble des données de départ.

1.2. Objectif de recherche

Ainsi, malgré leur pertinence, les méthodes existantes semblent limitées pour la représentation de données de mobilité. Selon (Klein, Cauvin, 2011), il existe trois grandes familles de représentation du mouvement : (1) la sémiotique, qui offre de nombreuses possibilités de représentations « classiques » (Bertin, 1967), (2) les filtres répondant partiellement à la surabondance d'information contenue dans les matrices de mouvements et (3) la modélisation qui propose des représentations dérivées explicatives. Le choix d'une solution sémiotique est associé à plusieurs critères liés principalement à ce que l'on veut représenter : la nature du mouvement (niveau de mesure) ; son importance, ses trajets,

SAGEO'2016 – Nice, 6-9 décembre 2016

sa direction, son sens ; la signification de ses extrémités (point de départ et d'arrivée). Les représentations spatio-temporelles proposées par Hägerstrand appartiennent à cette première famille de méthodes, de même que la multitude de travaux qui, dans des domaines et sur des espaces variés, les prolongent et continuent de mobiliser la recherche en cartographie pour produire de nouvelles images synthétiques des mobilités, notamment urbaines (Lenntorp, 1999 ; Kwan, 2000 ; Thériault *et al.*, 2002 ; N. Andrienko *et al.*, 2003 ; Kraak, 2003 ; Neutens *et al.*, 2011 ; Bach *et al.*, 2014). Toutefois, l'approche sémiotique présente des limites majeures dès que le nombre de lieux concernés augmente. Pour résoudre ce problème, des réductions sont indispensables. Elles prennent la forme d'opérations de filtrages thématiques (seuils quantitatifs d'apparition des flux) ou spatiaux (regroupements d'unités spatiales). Enfin, la modélisation des mouvements avec expression visuelle, et plus particulièrement celle des flux, a été principalement développée par (Tobler, 1978 ; 1981 ; Lavin, Cervený, 1987 ; Turk, Banks, 1996 ; Phan *et al.*, 2005). Trois types de représentations prédominent : les vecteurs isolés, les champs vectoriels et les « courants ». On peut y ajouter les potentiels d'attractivité dérivés des champs vectoriels (Tobler) et le vecteur unitaire (Lavin et Cervený).

Dans ce contexte, l'objectif est de renouveler les méthodes de représentations des données de mobilité pour les adapter à leur caractère multidimensionnel et individuel. Nous considérons que c'est en combinant les méthodes existantes à d'autres approches de visualisation qu'il est possible de gérer, afficher, explorer et analyser les données. Ce couplage de méthodes favorise l'émergence d'information, depuis des données « brutes », au moyen de transformations successives des données. De cette manière, la visualisation des données est censée permettre de saisir d'un coup d'oeil les tendances générales qui ressortent d'un jeu de données, en les rendant visibles. Pour cela, des représentations graphiques sont employées (nuage de points, diagramme, etc.), ainsi que des représentations cartographiques ; ces représentations visent à structurer les données et à mieux les voir afin de mieux comprendre les phénomènes géographiques complexes (Jiang *et al.*, 2003). Pour dégager des informations depuis un grand volume de données, il semble utile de combiner la visualisation avec d'autres méthodes qui peuvent renforcer ses capacités d'exploration des données.

D'un point de vue méthodologique, ces méthodes se fondent sur des principes de l'analyse exploratoire des données (AED) et de l'analyse exploratoire des données spatiales (AEDS), dont les fondements visent à résoudre l'inadaptation de la statistique classique aux données spatio-temporelles (Tukey, 1977 ; Waniez, 1991 ; Chen *et al.*, 2011). L'AEDS multiplie le nombre et le type de représentations visuelles et les représentations statistiques pour voir sous différents angles les données et pour amener l'observateur à s'interroger sur les données et à émettre des hypothèses à tester (Banos, 2001 ; Cauvin *et al.*, 2008). Les hypothèses testées, confirmées ou infirmées, sont ensuite reconsidérées pour être améliorées et « augmentées » : ainsi, après une phase inductive qui conduit l'utilisateur à des hypothèses, une phase déductive amène à affiner la réflexion

Proposition pour la visualisation des mobilités

grâce à des allers-retours entre hypothèses et déductions, jusqu'à l'affleurement d'informations nouvelles et donc, de connaissance sur les données.

Dans ce contexte, nous positionnons notre travail dans le cadre de la cartographie thématique, entendue comme une série de transformations (Cauvin *et al.*, 2007a ; 2007b) et employée comme une « cartographie de recherche ». Ce sont les transformations effectuées sur les données (en tant que données brutes, comme matière de départ de la réflexion) qui conduisent ainsi à l'émergence d'information (en tant que données transformées) par un raisonnement inductif sur le jeu de données conduisant à l'émission d'hypothèse(s) (Figure 1). L'objectif de cette cartographie de recherche est l'exploration des données en faisant varier les types de transformations, les points de vue ainsi que les méthodes de représentations, et les niveaux d'observation des dimensions des données.

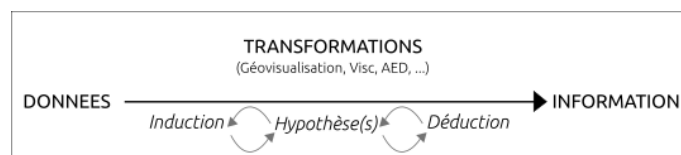


FIGURE 1. Des données à l'information

Pour mettre en œuvre cette cartographie de recherche, nous proposons ainsi de combiner les solutions visuelles existantes avec l'AED et d'opérer, préalablement aux analyses, des agrégations sur les données. Cette combinaison d'agrégations constitue la principale innovation de notre proposition méthodologique à la visualisation et à l'exploration des mobilités géographiques.

2. Méthodologie

Après une brève présentation du jeu de données de mobilité utilisé pour notre travail (2.1), nous exposons les techniques d'agrégation (2.2.) sur les dimensions spatiales (2.3), temporelles (2.4) et spatio-temporelles (2.5) des données.

2.1. Terrain d'étude et données

Dans le cadre de ce travail, nous disposons d'un jeu de données issu du modèle de simulation spatiale MobiSim, plateforme de simulation LUTI individualisée développée au sein du laboratoire ThéMA (Tannier *et al.*, 2015 ; Antoni *et al.*, 2016). Elle fournit des données de mobilités individuelles, désagrégées et structurées, qui correspondent aux positions successives des individus en dépla-

SAGEO'2016 – Nice, 6-9 décembre 2016

gement dans l'espace urbain. Le jeu de données compte environ 180.000 individus, dont les déplacements sont connus à la seconde sur une journée type de simulation, à l'échelle de la Communauté d'Agglomération du Grand Besançon (CAGB). Ces déplacements sont constitués par un point d'origine et un point de destination connus à une heure de départ et à une heure d'arrivée (au format heure:minute:seconde) et dont le motif (travail, étude, domicile, loisirs, réseau social, accompagnement, commerce ou service) et le mode (voiture, transports en commun, vélo, marche à pied) sont également connus. Ces informations sont stockées dans une table attributaire correspondant à ces points.

2.2. Agrégation en trois étapes

Afin de guider l'exploration de différentes agrégations, nous testons une méthode en trois étapes qui s'applique, pour l'instant, à chaque dimension des données prise individuellement. Cette méthode consiste en la sélection (1) de l'opérateur d'agrégation, (2) du type de chaînage et (3) du niveau d'agrégation.

1. Un opérateur (somme, moyenne, valeur centrale) est sélectionné pour agréger les données selon l'attribut sur lequel porte l'agrégation : nombre d'individus, âge médian, part de déplacement en transport en commun, etc. La structure de l'unité d'agrégation est également choisie à cette étape : pour un attribut spatial, la forme de l'unité (carrés, hexagones, etc.) et celle de la grille (régulière ou non).

2. Les agrégations à plusieurs niveaux (sur le même attribut) peuvent être calculées de manière directe ou itérative (Figure 2). Le chaînage direct permet d'agréger les données à un niveau n_i depuis la donnée source, correspondant au niveau n_0 . Le chaînage itératif permet d'agréger les données à un niveau n_i depuis le niveau précédent n_{i-1} ; le chaînage itératif est en partie conditionné par le choix de l'unité d'agrégation à l'étape 1.

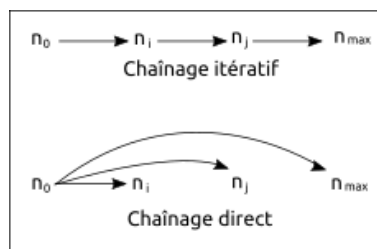


FIGURE 2. Type de chaînage d'agrégation

3. L'étendue des niveaux d'agrégation sur un attribut est comprise entre n_0 et n_{max} (l'attribut est entièrement agrégé et ne compte alors plus qu'une seule unité), ce qui définit l'étendue de n_i . Entre ces deux limites, chaque niveau

Proposition pour la visualisation des mobilités

n_i peut être testé et chacun d'eux produit une représentation différente du jeu de données, plus ou moins facile à gérer, analyser et visualiser. Chaque représentation au niveau n_i doit être évaluée et comparée aux autres niveaux afin de mesurer l'information obtenue et de, éventuellement, déterminer un intervalle de niveaux dans lequel un grand nombre d'informations affleure tout en étant aisé à manipuler.

2.3. Agrégation spatiale

Les déplacements sont réalisés sur un territoire, à partir d'un réseau de transport (réseau routier ou de transport en commun) entre des lieux. Il semble peu judicieux de chercher à les analyser sans prendre en compte l'occupation du sol, dont ils sont dépendants. Nous disposons ainsi de deux attributs spatiaux à agréger : l'occupation du sol et le réseau routier.

Nous appliquons l'agrégation en trois étapes aux attributs retenus (Figure 3). Pour représenter l'occupation du sol, nous retenons l'attribut « bâtiment », dont le pourcentage de surface bâtie par unité d'agrégation est calculé. La structure retenue pour l'instant est une grille régulière de cellules hexagonales. De ce fait, seule le chaînage direct d'agrégation est applicable dans cet exemple. L'étendue des niveaux est compris entre l'échelle du bâtiment (n_0) et celui de la zone d'étude (n_{max} , soit environ 16km de côté); n'importe quel n_i peut être testé entre ces deux extrêmes.

Caractéristiques	Attribut Opérateur Unité	Forme urbaine Pourcentage Hexagone	Caractéristiques	Attribut Opérateur Unité	Réseau transport Somme Hexagone
Chaînage	Type Nombre d'itération	Direct Non applicable	Chaînage	Type Nombre d'itération	Direct Non applicable
Niveaux	n_0 n_i n_{max}	Bâtiment 125m Zone d'étude	Niveaux	n_0 n_i n_{max}	Tronçon routier 500m Zone d'étude

FIGURE 3. Exemple d'application de l'agrégation en trois étapes à l'occupation du sol et au réseau

Pour représenter le réseau routier, nous effectuons une agrégation sur la géométrie du réseau lui-même. Afin de maximiser le nombre de connexions entre cellules (et donc le nombre de tronçon), une grille hexagonale est retenue et seule une agrégation directe est possible. Cette agrégation permet de représenter linéairement le réseau sous forme de « carte de métro ». Pour cela, chaque tronçon du réseau routier est rattaché au tronçon de centroïde à cen-

SAGEO'2016 – Nice, 6-9 décembre 2016

troïde dont il est le plus proche (Figure 4¹) : de la sorte, les tronçons routiers - sans considération pour leur attributs thématiques - proches sont agrégés en un seul tronçon. Cette représentation du réseau routier permet de mieux en voir la structure sans la grever, la connectivité du réseau entre chaque centroïde de cellule devenant ainsi plus visible que sa topologie (Stott, Rodgers, 2005).

L'agrégation spatiale ainsi mise en place permet de naviguer entre différents niveaux d'agrégation, depuis les données désagrégées du jeu de données (n_0) à un niveau très élevé d'agrégation (n_{max}), en passant par plusieurs niveaux intermédiaires. C'est dans la multiplication des vues à différents niveaux d'agrégation spatiale que l'on entend voir émerger des structures, qui sont susceptibles d'être visibles à divers niveaux (par exemple, une organisation selon un axe sud-ouest nord-est du réseau routier) ou de ne l'être qu'à un niveau particulier (toutes les cellules valent 0 ou 100% de surface bâtie).

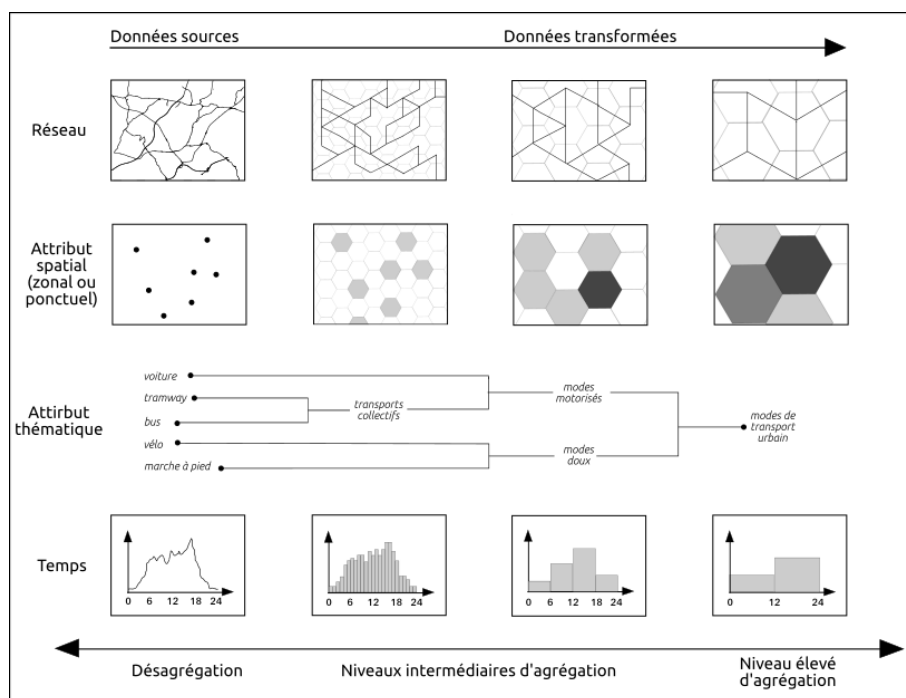


FIGURE 4. Techniques différenciées d'agrégation

1. L'agrégation sur les dimensions thématiques n'a pas été présentée, mais est envisagée pour un développement futur.

Proposition pour la visualisation des mobilités

2.4. Agrégation temporelle

De même que pour les données spatiales, nous faisons le choix d'agréger les données temporelles. L'agrégation temporelle peut s'effectuer de manière « classique » en regroupant les secondes en minutes, les minutes en heures, les heures en périodes de journée (matinée, après-midi), etc. Cela revient finalement à discrétiser le temps de la journée en classes d'étendues égales et conduit à construire une carte ou une vue pour chaque classe temporelle : 2 cartes pour représenter la journée découpée en 12 heures, 24 cartes pour chaque heure de la journée, ou 86.400 pour chaque seconde de la journée. Dans le premier « lot » de deux cartes, les périodes de temps sont trop grandes (43.200 secondes, où, pour chacune d'elles, chaque individu peut effectuer un déplacement) pour que la carte soit interprétable. Inversement, le lot de 86.400 cartes est trop volumineux pour permettre au lecteur d'en déduire des tendances ou organisations dans les déplacements individuels, en raison du trop grand nombre de cartes à comparer les unes aux autres. C'est donc à des niveaux intermédiaires d'agrégation temporelle qu'il faut visualiser les données de mobilité quotidienne pour les analyser et les explorer commodément.

Cependant, même à des niveaux intermédiaires d'agrégation, le recours à des classes d'étendues temporelles égales paraît en un sens contraignant, de même que l'usage de l'agrégation spatiale à l'IRIS peut l'être pour les données spatiales (Antoni *et al.*, 2012). En effet, si l'on considère le lot de 24 cartes pour chaque heure de la journée, certaines de ces cartes comptabilisent peu de déplacements, par exemple pendant la nuit, alors que d'autres connaissent des pics de déplacements (par exemple la tranche 8h-9h), en raison des rythmes urbains. Face à ce constat, chercher une discrétisation temporelle, à chaque niveau d'agrégation (qui correspond alors à un nombre de classes), fondée sur les méthodes de discrétisation des variables thématiques (méthode de Jenks, classes d'écart-type, etc.) paraît permettre l'homogénéisation des classes (et donc des « lots de carte »). Cela permet de représenter dans des classes temporelles plus petites les moments comptant le plus de déplacements, ce qui réduit donc le nombre de déplacements visualisés simultanément afin de faciliter leur analyse.

2.5. Agrégation spatio-temporelle

D'un point de vue spatio-temporel, lorsque l'on ne s'intéresse qu'à un petit nombre d'individus, les concepts et représentations proposés par la *Time geography* peuvent suffire ; lorsque l'on s'intéresse à un grand nombre d'individus, celle-ci est malheureusement insuffisante. Nous proposons donc d'agréger spatialement et temporellement les données de mobilité pour pouvoir les saisir dans l'espace et dans le temps. Pour cela, plusieurs méthodes de *clustering* spatio-temporels des données a été proposées (G. Andrienko, Andrienko, 2008 ; 2010a ; 2010b), comme les algorithmes Netscan (Kharrat *et al.*, 2009) ou Traclus (Lee

SAGEO'2016 – Nice, 6-9 décembre 2016

et al., 2007), qui permettent de réduire la taille des données en cherchant les ressemblances spatio-temporelles entre les déplacements des individus (trajectoires similaires dans l'espace et dans le temps). De plus, le résultat de tels *clustering* peut être combiné avec un cube spatio-temporel, qui permet d'intégrer le temps sur un axe vertical à une cartographie en abscisse et en ordonnée de l'espace. On obtient alors une représentation des mobilités quotidiennes à l'intérieur de *voxels*, dont la position à l'intérieur du cube spatio-temporel indique la localisation dans l'espace et dans le temps d'individus en mouvement et dont la couleur ou le grain peut donner une information complémentaire.

Ces propositions d'agrégation des données spatiales, temporelles et spatio-temporelles, ont finalement pour objectif de permettre à l'utilisateur d'analyser les données de mobilités quotidiennes en naviguant parmi les niveaux d'agrégation pour choisir à la fois le volume et le type de données visualisées, ainsi que la forme que prend cette visualisation. De cette manière, en augmentant le nombre de visualisation et en le faisant varier, des hypothèses différenciées peuvent être émises et faire émerger de l'information sur les mobilités quotidiennes.

3. Résultats

Des premiers résultats ont été obtenus sur l'agrégation spatiale de l'occupation du sol et du réseau routier de la CAGB. L'occupation du sol est prise en compte uniquement à partir des bâtiments résidentiels, dont le pourcentage de surface au sol est calculé pour chaque cellule de 125 mètres de côté (Figure 5). Cette agrégation permet de visualiser les zones résidentielles de la CAGB et leur concentration. La ville de Besançon se distingue au centre de la zone, et un grand nombre de villes périphériques, proches et concentrées sont visibles au nord, ouest et sud-ouest. Au contraire, la partie est compte un petit nombre de cellules « pleines », qui sont également plus espacées. Ainsi, la représentation agrégée de la surface bâtie dans des cellules permet de conserver l'organisation urbaine de la zone d'étude (la forme, la localisation et la concentration des villes sont maintenues) tout en facilitant la manipulation des données (32.590 bâtiments individuels pour la donnée de base, contre 11.071 cellules ici). Cette visualisation conduit à poser l'hypothèse que l'urbanisation à l'ouest et au nord est plus dense et prend la forme de corridors.

Le réseau routier représenté sur la Figure 5 est réalisé à partir des cellules de 1000 mètres de côté. Cette représentation agrégée du réseau réduit considérablement le nombre de tronçons routiers (de 26.871 tronçons pour le réseau routier de départ à 9.384 tronçons dans celui-ci) tout en maintenant la connectivité du réseau : le recours à des cellules hexagonales permet, depuis chaque cellule, d'obtenir jusqu'à six connexions vers les cellules attenantes. On distingue ainsi quelques zones moins bien connectées au reste du réseau : si les cellules au centre de la zone d'étude sont connectées à chacune de leurs cellules

Proposition pour la visualisation des mobilités

voisines, ce n'est pas le cas des cellules dans la partie est de la zone, dont certaines sont connectées à seulement deux ou trois de leurs cellules contiguës. On peut alors formuler l'hypothèse, depuis cette agrégation, que la connectivité n'est pas homogène dans la zone d'étude, et que des inégalités d'accessibilité peuvent en découler.

En le comparant au réseau routier réalisé à partir des cellules de 1000 mètres de côté, celui obtenu à partir des cellules hexagonales de 500 mètres de côté rend encore plus visible la structure du réseau (Figure 6). Les cellules du centre de la zone sont toujours très bien connectées, mais celles des périphéries sont moins connectées (deux ou trois connexions), voire complètement isolées du réseau, en particulier dans la périphérie est et nord-est. Ce niveau d'agrégation du réseau routier permet d'émettre une hypothèse sur l'existence d'effet-barrières dans la zone d'étude.

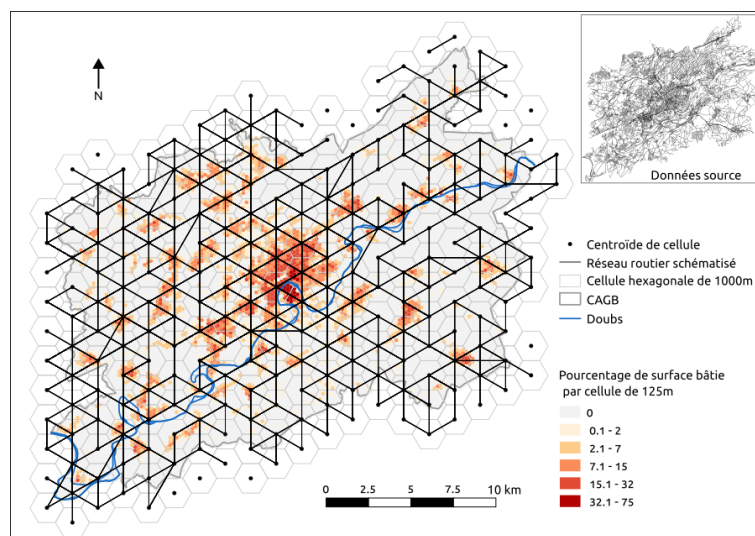


FIGURE 5. Agrégation des bâtiments de la CAGB dans des cellules hexagonales de 125 mètres de côté et du réseau routier dans des cellules de 1000 mètres de côté

SAGEO'2016 – Nice, 6-9 décembre 2016

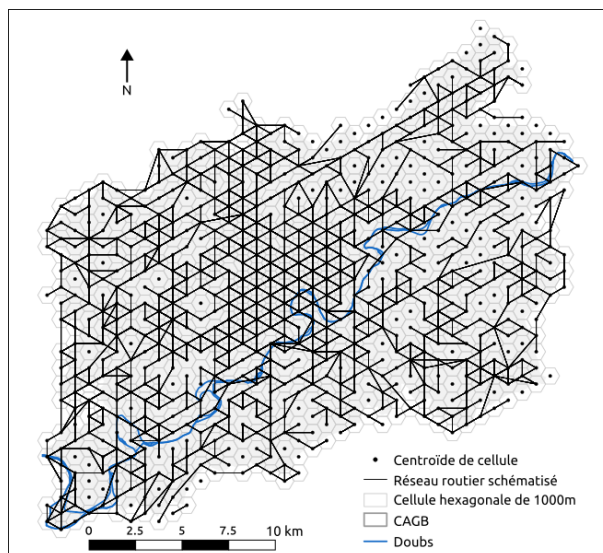


FIGURE 6. Agrégation du réseau routier de la CAGB dans des cellules hexagonales de 500 mètres de côté

4. Conclusion

Les données de mobilité étant de nature spatio-temporelle, trois dimensions élémentaires constituent leur organisation (Peuquet, 1994) : espace (de où vers où ?), thème (quoi ?), temps (quand ? en combien de temps ?). Appliquée à notre cas d'étude, cette structure prend la forme d'un graphique « radar » (Figure 7) organisant les données mobiles selon 5 dimensions constituantes : réseau, occupation du sol, temps, modes de transport et motifs de déplacement. Chacune de ces dimensions peut alors être manipulée individuellement, notamment sous la forme d'agrégation/désagrégation, afin de parcourir le jeu de données et d'en faire émerger des structures et des processus sous-jacents.

Chaque axe du graphique est gradué en fonction du niveau d'agrégation effectué sur l'attribut considéré. La combinaison de divers niveaux d'agrégation sur chaque dimension crée alors une « silhouette » sur le graphique, qui permet d'obtenir une vue particulière des données. Ce graphique doit permettre de tester « toutes » les combinaisons possibles : il constitue une sorte de protocole d'exploration des données.

Pour l'instant, nous nous sommes focalisés sur le réseau, avec deux niveaux d'agrégation. Le réseau créé à partir des cellules de 1000 mètres lisse les données, alors que celui créé à partir des cellules de 500 mètres commence à faire

Proposition pour la visualisation des mobilités

émerger une structure, et semble rendre la mise en relation entre les lieux plus perceptible. Ainsi, l'exploration des niveaux d'agrégation doit être poursuivie, en intégrant l'ensemble des dimensions des données. Chaque vue créée doit ensuite être associée à des indicateurs globaux (Goodchild, 1986) ou locaux (Hagen-Zanker, 2006) pour améliorer son interprétabilité; par exemple, un indicateur de densité de connexion de chaque point du réseau agrégé pour analyser les unités agrégées. La qualité et l'efficacité de ces indicateurs doit conduire à l'évaluation des vues obtenues, afin de déterminer les plus efficaces (en d'autres termes, quelles silhouettes du graphique sont les plus favorables à l'exploration des données). Cette évaluation est à compléter par passation d'enquête sur un panel d'utilisateurs pour valider le choix des vues, en fonction, non seulement, de la maniabilité des données dans un SIG, mais aussi de la quantité et la qualité des informations obtenues (obtient-on plus d'informations? obtient-on une information plus facilement interprétable et compréhensible?). Cela doit, finalement, permettre de mettre en place une méthode pour la découverte d'informations depuis des données géographiques. Leur efficacité, en tant que gains temporel et informationnel, doit être vérifiée en s'appuyant sur la cognition, pour guider l'utilisateur dans la découverte d'informations depuis des données géographiques.

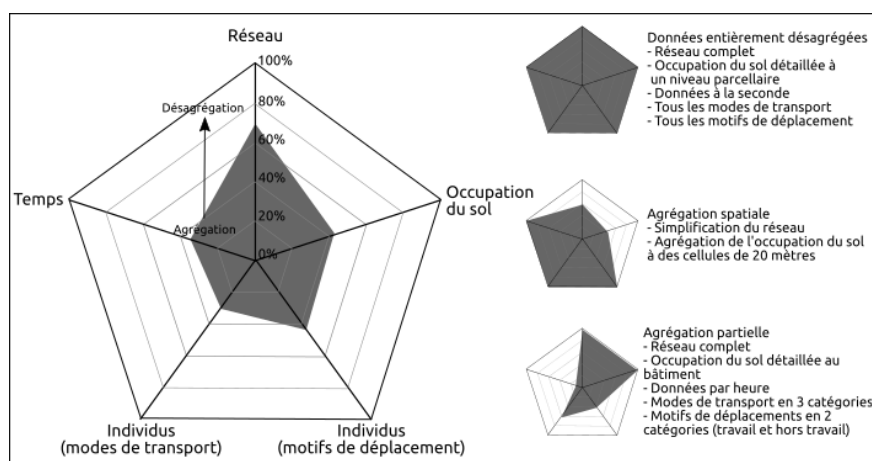


FIGURE 7. Combinaison de niveaux d'agrégation

Bibliographie

- Andrienko G., Andrienko N. (2008). Spatio-temporal aggregation for visual analysis of movements.
- Andrienko G., Andrienko N. (2010a). A general framework for using aggregation in visual exploration of movement data. *The Cartographic Journal*, vol. 47.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Andrienko G., Andrienko N. (2010b). Interactive Cluster Analysis of Diverse Types of Spatiotemporal Data. *ACM SIGKDD Explorations Newsletter*, vol. 11, n° 2, p. 19–28.
- Andrienko N., Andrienko G., Gatalaky P. (2003). Exploratory spatio-temporal visualization: an analytical review. *Journal of visual languages and computing*, vol. 14, p. 503–541.
- Antoni J.-P., Klein O., Moisy S. (2012). La discrétisation temporelle. Une méthode de structuration des données pour la cartographie dynamique. *Cartes & géomatique*, p. 27–31.
- Antoni J.-P., Lunardi N., Vuidel G. (2016). Simuler les mobilités individuelles. Les enjeux de l'information géographique. *Revue internationale de géomatique*, vol. 2, p. 237–262.
- Bach B., Dragicevic P., Archambault D., Hurter C., Carpendale S. (2014). A review of temporal data visualizations based on space-time cube operations. In *Eurographics Conference on Visualization*.
- Banos A. (2001). A propos de l'analyse exploratoire de données . *Cybergeog. European journal of geography*.
- Bertin J. (1967). *Sémiologie graphique* (Mouton/Gauthier-Villars éd.).
- Cauvin C., Escobar F., Serradj A. (2007a). *Cartographie thématique. Tome 1 : une nouvelle démarche*. Hermès Science.
- Cauvin C., Escobar F., Serradj A. (2007b). *Cartographie thématique. Tome 2 : des transformations incontournables*. Hermès Science.
- Cauvin C., Escobar F., Serradj A. (2008). *Cartographie thématique. Tome 3 : Méthodes quantitatives et transformations attributaires*. Hermès Science.
- Chardonnel S. (2007). Time-geography: Individuals in Time and Space. In L. Sanders (Ed.), *Models in spatial analysis*.
- Chen J., Shaw S.-L., Yu H., Lu F., Chai Y., Jia. Q. (2011). Exploratory Data Analysis of activity diary data: a space-time GIS approach. *Journal of Transport Geography*, vol. 19, n° 3, p. 394–404.
- Gerber P., Klein O., Carpentier S. (2012). Local mobility and cross-border periurbanisation. In C. Sohn (Ed.), *An emerging cross-border metropolitan region*, p. 141–159. Peter Lang.
- Goodchild M. (1986). Spatial Autocorrelation Manuel de logiciel.
- Hagen-Zanker A. (2006). Map comparison methods that simultaneously address overlap and structure. *Journal of geographical systems*, vol. 8, p. 165–185.
- Hägerstrand T. (1982). Diorama, Path and Project. *Tijdschrift voor Economische en Sociale Geografie*, vol. 73, n° 6, p. 323–339.
- Jiang B., Huang B., Vasek V. (2003). Geovisualisation for Planning Support Systems. In S. Geertman, J. Stillwell (Eds.), *Planning Support Systems in Practice*, p. 177–191. Springer.

Proposition pour la visualisation des mobilités

- Kaddouri L., Blaise J.-Y., Davoine P.-A., Mathian H., Saint-Marc C. (2014). *Etat des lieux des représentations dynamiques des temporalités des territoires*. Rapport technique.
- Kharrat A., Sandu Popa I., Zeitouni K., Faiz S. (2009). Caractérisation de la densité de trafic et de son évolution à partir de trajectoires d'objets mobiles. In *UbiMob'09 Proceedings of the 5th French-Speaking Conference on Mobility and Ubiquity Computing*, p. 33–40.
- Klein O. (2007). *Modélisation et représentations spatio-temporelles des déplacements quotidiens urbains. Application à l'Aire Urbaine Belfort-Montbéliard*. Thèse de doctorat non publiée, Université de Strasbourg.
- Klein O., Cauvin C. (2011). How to grasp the geographic movement? A heuristic guide for its cartographic representation. In *Proceedings of the 25th International Cartographic Conference*.
- Kraak M. J. (2003). The space-time cube revisited from a geovisualization perspective. In *Proceeding of 21st International Cartographic Conference*.
- Kwan M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation research part C*, vol. 8, p. 185–203.
- Lavin S., Cervený R. (1987). Unit-vector density mapping. *The Cartographic Journal*, vol. 24, p. 131–140.
- Lee J.-G., Han J., Whang K.-Y. (2007). Trajectory Clustering: A Partition-and-Group Framework. In *SIGMOD'07 Proceedings of the 2007 ACM SIGMOD International conference on management of data*, p. 593–604.
- Lenntorp B. (1999). Time-geography - at the end of its beginning. *GeoJournal*, vol. 48, p. 155–158.
- Neutens T., Schwanen T., Witlox F. (2011). The prism of everyday life: towards a new research agenda for time geography. *Transport Reviews*, vol. 31, n° 1, p. 25–47.
- Peuquet D. (1994). It's about Time: a conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the association of american geographers*, vol. 84, n° 3, p. 441–461.
- Phan D., Xiao L., Winograd T. (2005). Flow map Layout, Information Visualization (InfoVis). In *Proceeding IEE Symposium on Graphics*, p. 6.
- Stott J. M., Rodgers P. (2005). Automatic metro map design techniques. In *Proceeding 22nd International Cartographic Conference (ICC'05)*.
- Tannier C., Hirtzel J., Stephenson R., Couillet A., Vuidel G., Youssoufi S. (2015). Conception and use of an individual-based model of residential choice in a planning decision process. Feedback from an experimental trial in the city of Besançon, France. *Progress in planning*.
- Thériault M., Claramunt C., Séguin A.-M., Villeneuve P. (2002). Temporal GIS and statistical modelling of personal lifelines. In D. Richardson, P. Van Oosterom (Eds.), *Advances in Spatial Data Handling*, p. 433–449.
- Tobler W. (1978). Migration fields. In W. Clark, E. Moore (Eds.), *Population mobility and residential change*, vol. 25, p. 215–232. Evanston.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Tobler W. (1981). A model of geographical movement. *The Geographical Analysis*, vol. 13, n° 1, p. 1-20.
- Tukey J. W. (1977). *Exploratory Data Analysis*. Pearson.
- Turk G., Banks D. (1996). Image-guided streamline placement. In *SIGGRAPH 96 Conference Proceedings*, p. 453-460.
- Waniez P. (1991). De la "cartomatique" aux systèmes d'information géographique : méthodes et techniques nouvelles pour l'analyse spatiale. *Espace, populations, sociétés*, vol. 3, p. 505-516.

Pour une cartographie romanesque de Paris au XIXe siècle. Proposition méthodologique

Noémie Boeglin¹, Michel Depeyre¹, Thierry Joliveau¹, Yves-François Le Lay²

1. Laboratoire Environnement-Ville-Société, Université Jean Monnet de Saint-Etienne

6 rue Basse-des-Rives, F42000 Saint-Etienne, France

noemie.boeglin@univ-st-etienne.fr, michel.depeyre@univ-st-etienne.fr,

thierry.joliveau@univ-st-etienne.fr

2. Laboratoire Environnement-Ville-Société, Ecole Normale Supérieure de Lyon

15 parvis René Descartes, F69007 Lyon, France

yves-francois.le-lay@ens-lyon.fr

RESUME. L'article propose une méthode pour localiser, cartographier et analyser les occurrences de lieux citées dans un corpus de 31 romans du XIXe siècle dont l'action se situe pour tout ou partie à Paris. L'objectif à terme plus lointain est de construire une plateforme facilitant l'analyse cartographique d'autres textes évoquant Paris à la même époque. La méthode expérimentée associe l'usage de la plateforme d'analyse textuelle open-source TXM et l'élaboration d'un Système d'Information Géographique adapté, supportant un atlas toponymique des lieux parisiens entre 1830 et 1914 et fournissant une interface de géovisualisation. Les quatre étapes successives sont décrites : rassembler et corriger les versions numériques des textes des romans, extraire les Entités Spatiales Nommées par l'écriture de requêtes adaptées dans la plateforme textométrique TXM, localiser ces entités par la mobilisation de sources historiques en constituant un gazetteer adapté et proposer des formes de cartographie adaptées aux spécificités des espaces littéraires. Une discussion et des perspectives concluent le papier.

ABSTRACT. We propose a method to locate, map and analyze the occurrences of places cited in a corpus of 31 novels of the nineteenth century whose action takes place wholly or partly in Paris. The more distant term goal is to build a platform facilitating the cartographic analysis of other texts evoking Paris at the same time. The experienced method combines the use of a textual analysis platform (TXM) and the development of a convenient Geographic Information System supporting a gazetteer of Parisian places between 1830 and 1914 and an interface of geovisualization. The four stages are described: (a) collect and correct digital versions of novels, (b) extract spatial named entities by writing suitable queries in the TXM lexicometric platform, (c) locate these entities by using historical sources and constituting an appropriate gazetteer, and (d) offer mapping forms adapted to the specificities of literary spaces. Discussion and perspectives conclude the article.

SAGEO'2016 – Nice, 6-9 décembre 2016

MOTS-CLES : RIG – GEOPARSING – CARTOGRAPHIE – ANALYSE TEXTUELLE – TEXTOMETRIE – HISTOIRE URBAINE – PARIS – LITTÉRATURE – SIG – XIXe SIECLE

KEYWORDS: GEOPARSING – CARTOGRAPHY – GIS – TEXTOMETRY – LITERATURE – PARIS – URBAN HISTORY – 19TH CENTURY

1. La question posée

Dans le cadre de sa thèse d'histoire sur les paysages parisiens évoqués dans 31 romans du XIXe siècle, Noémie Boeglin effectue un inventaire cartographique des lieux nommés dans son corpus. Cet inventaire manuel s'avère fastidieux et approximatif. La réalisation de cartes pose aussi des problèmes complexes tant que l'on reste dans une approche manuelle, même informatisée, en particulier si l'on veut procéder à des visualisations nombreuses par roman, par auteur ou selon une logique temporelle. L'équipe de recherche s'est constituée afin de : 1) proposer une méthodologie rigoureuse permettant de répondre de manière opérationnelle à ce problème précis ; 2) réfléchir à la généralisation possible de cette méthode et envisager l'extraction et la cartographie automatique des noms de lieu dans les textes de fiction.

2. Etat de l'art et problématique

Le rapprochement entre la géographie et les disciplines littéraires et l'emprunt de concepts et de techniques cartographiques pour l'étude des espaces romanesques est déjà ancien (Moretti, 2000) ; (Moretti, 2008) ; (Bulson, 2007). F. Moretti en particulier a proposé une approche de « lecture distante » des textes utilisant des méthodes statistiques empruntées aux sciences humaines. Il propose ainsi une « géographie de la littérature » en replaçant par exemple dans la topographie concrète des lieux les personnages des romans du XIXe dont l'intrigue se passe à Londres ou à Paris, réactualisant sur une base systématique des propositions anciennes. Si le mode de production de F. Moretti était manuel, les méthodes informatiques de recherche d'information géographique (RIG) dans les textes, relevant plus ou moins de ce que l'on appelle en anglais *geoparsing*, se sont largement développées pour tenter d'extraire le plus automatiquement possible ces références textuelles aux lieux. Le développement du domaine des humanités numériques a contribué à accroître fortement l'intérêt pour ces approches en littérature et en histoire, comme l'illustrent par exemple les travaux de (Gregory et al., 2015) ou la publication récente du texte de référence de (Cooper, Donaldson, et Murrieta-Flores, 2016). Il n'est pas utile de développer ici les différentes méthodes de RIG, à laquelle de nombreux travaux ont été consacrés ces dernières années. On trouvera dans (Gaio et al., 2008) et (Sallaberry, 2014) une présentation de différentes techniques utilisables pour repérer et extraire les lieux de documents textuels. Les solutions diffèrent selon que ces lieux apparaissent sous forme de toponymes, appelés dans la littérature spécialisée Entités Spatiales Nommées, ou qu'ils sont évoqués de manière implicite ou allusive. L'objectif visé dans ce papier est de proposer une méthode pour trouver les Entités Spatiales Nommées dans les textes en utilisant les marqueurs explicites simples : « rue », « place », « théâtre »

etc. La recherche de toponymes aux formes plus floues nécessite de mobiliser des solutions plus complexes.

Le deuxième domaine dans lequel nous faisons des propositions est celui de la visualisation cartographique de ces lieux, qui reste sommaire et même parfois déficiente chez F. Moretti ou dans les travaux de critique littéraire. Au-delà de l'application des règles admises de la sémiologie classique, on peut faire l'hypothèse que la cartographie littéraire doit s'adapter à la nature originale du rapport à l'espace de son objet et qu'il est nécessaire de proposer des solutions nouvelles. Dans son projet d'Atlas littéraire de l'Europe¹, une équipe de l'Université de Zürich a tenté de décrire la spécificité de ces espaces fictionnels, dont le caractère imaginaire nécessite un mode de projection original. (Piatti et al., 2008) proposent une liste de solutions pour rendre compte sémiologiquement de ces espaces littéraires par nature incertains, flous, détournés... Notre contribution dans ce papier porte sur des propositions d'un ordre un peu différent, permettant de comparer la manière dont plusieurs romans décrivent le même espace, en l'occurrence l'espace parisien.

L'objectif de cette première phase de la recherche est en effet de retrouver automatiquement dans les textes de 31 romans d'auteurs variés (voir Tableau 1) les Entités Spatiales Nommées présentes et de les utiliser pour cartographier l'empreinte spatiale de ces romans ou de leurs auteurs. Compte tenu des délais de la première phase de ce projet, nous ne travaillons pour l'instant qu'avec des solutions informatiques existantes et disponibles, articulées manuellement et sans développement informatique spécifique. La méthode proposée se décompose en trois étapes : 1) le repérage des Entités spatiales dans les textes ; 2) la localisation des Entités Spatiales Nommées dans l'espace parisien ; 3) la géovisualisation et l'analyse spatiale des résultats.

<i>La maison du Chat-qui-pelote</i> , Honoré de Balzac, 1830	<i>Jack</i> , Alphonse Daudet, 1876
<i>Ferragus</i> , Honoré de Balzac, 1833	<i>L'Assommoir</i> , Emile Zola, 1877
<i>La fille aux yeux d'or</i> , Honoré de Balzac, 1835	<i>Une belle journée</i> , Henry Céard, 1881
<i>Le Père Goriot</i> , Honoré de Balzac, 1835	<i>Pot-Bouille</i> , Emile Zola, 1882
<i>César Birotteau</i> , Honoré de Balzac, 1837	<i>Le vingtième siècle</i> , Albert Robida, 1883
<i>Les Mystères de Paris</i> , Eugène Sue, 1842	<i>Sapho</i> , Alphonse Daudet, 1884
<i>L'envers de l'histoire contemporaine</i> , Honoré de Balzac, 1848	<i>Bel-ami</i> , Guy de Maupassant, 1885
<i>Les demoiselles de magasin</i> , Paul de Kock, 1863	<i>L'œuvre</i> , Emile Zola, 1886
<i>Paris au XXème siècle</i> , Jules Verne, 1863	<i>La vie électrique</i> , Albert Robida, 1892
<i>L'éducation sentimentale</i> , Gustave Flaubert, 1869	<i>Paris</i> , Emile Zola, 1897
<i>La Curée</i> , Emile Zola, 1871	<i>La Charpente</i> , J.H. Rosny jeune, 1900
<i>Le ventre de Paris</i> , Emile Zola, 1873	<i>Mr. Bergeret à Paris</i> , Anatole France, 1901
	<i>La Maternelle</i> , Léon Frapié, 1904
	<i>La Vague rouge</i> , J.H. Rosny aîné, 1910
	<i>Dans les rues</i> , J.H. Rosny aîné, 1913

Tableau 1. Les 31 romans du corpus

¹ <http://www.literaturatlas.eu/en/>

SAGEO'2016 – Nice, 6-9 décembre 2016

3. La réalisation

3.1. Le repérage des Entités Spatiales Nommées dans le texte

Il est d'abord nécessaire de rassembler et de corriger les versions numériques des textes des romans, souvent fautives car tirées de procédures OCR. On obtient un corpus de 4 778 464 mots, interrogeable dans son ensemble ou par fraction (Figure 1). Ce corpus est chargé dans la plateforme d'analyse textométrique TXM¹ (Heiden, 2010) (Heiden, Magué, et Pincemin, 2010).

TXM propose notamment un module logiciel appelé *Corpus Query Language* (CQP) qui traite des requêtes et fonctionne comme un moteur de recherche. Les requêtes sont construites en *Corpus Query Language* (CQL) pour exprimer un motif linguistique (un ou plusieurs mots) au moyen de leurs propriétés (forme graphique, lemme ou encore catégorie grammaticale). Ces requêtes permettent ainsi de trouver toutes les mentions d'Entités Spatiales Nommées des différentes catégories : voies (rue, avenue, boulevard, quai, etc.), cours d'eau (rivière, bassin, canal...), bâtiments (théâtre, hôpital, école...), parcs et autres. Le Tableau 2 en donne un exemple révélateur de la difficulté pratique de l'analyse de données textuelles. L'expression est d'abord conçue pour chercher toutes les occurrences d'une catégorie nommée, à savoir les rues. Pour en appréhender les diverses réalisations, il importe de neutraliser la casse (%c), les éventuels signes diacritiques (%d) et le pluriel : rechercher un lemme permet de saisir un mot sous ses formes variables (en les ramenant à une entrée du dictionnaire). Le reste de l'expression spécifie progressivement la requête. Il s'agit d'abord de préciser que la catégorie nommée doit être suivie d'un ou de plusieurs mots commençant par une majuscule ([word="\p{Lu}.*"]): la « rue Plumet » et la « rue du Grenier-Saint-Lazare » par exemple sont désormais considérées. Pour ne pas négliger la « rue du Temple » ou la « rue de l'Homme-Armé », la requête doit préciser qu'un ou deux mots ([]?) peuvent s'insérer entre la rue et le mot avec majuscule. Il faut encore en exclure quelques formes graphiques et plus particulièrement un éventuel signe de ponctuation qui signifierait le début d'une nouvelle proposition : des chaînes telles que « rue. Il » ou « rue... Ah » se voient ainsi exclues de la recherche.

Au terme d'un enrichissement progressif, la requête s'est complexifiée mais débouche sur un index des rues qui peut être traité dans les meilleurs délais. Certes cette requête *ad hoc* ne met en œuvre qu'une recherche ciblée et doit être adaptée lorsque de nouveaux questionnements se font jour. Mais plus généralement le langage CQL se prête bien à la recherche de tout motif linguistique (et non seulement des entités spatiales), si bien que son apprentissage peut être remobilisé ensuite selon un point de vue de sémantique interprétative (en s'intéressant par exemple aux adjectifs qui qualifient les rues parisiennes).

¹ <http://textometrie.ens-lyon.fr/>

Cartographie romanesque de Paris

```
[frlemma="rue"%cd] [word!="\.,\;\;|\:|\!|\?|\.\.\-|une|\-|où|ainsi|et|aurait|-l"%c]? [word!="\.,\;\;|\:|\!|\?|\.\.\-|une|\-|où|ainsi|et|aurait|-l"%c]? [word="p{Lu}.*" & word!="Ça|Ah|O|Venez|Et|M|L."]
```

Tableau 2. Exemple de requête TXM pour extraire les rues des 31 romans

TXM fournit une liste de concordances, c'est-à-dire l'ensemble des Entités Nommées trouvées par la requête, encadrées par les passages du texte où elles apparaissent. (Tableau 3). Il est possible de naviguer dans la liste des concordances obtenues et d'accéder en permanence au texte lui-même à partir de l'extrait correspondant. Cela donne à l'analyste un outil précieux pour explorer son corpus textuel à partir des Entités Spatiales Nommées et analyser le contenu littéraire et les descriptions associées aux différents lieux. Cette fonction a permis par exemple de rechercher rapidement les mentions des passages et galeries couvertes, pour faire une analyse approfondie de cet élément urbanistique très important du Paris de l'époque.

text_id	Contexte gauche	Pivot	Contexte droit
education	qu'elle était sur le trottoir de la	rue Tronchet	depuis longtemps. Elle y attendait quelque chose d'indéterminé, de
education	blaguait, était très gai. Par la	rue Duphot	, ils atteignirent les boulevards. Des lanternes vénitiennes, suspendues aux
education	direct impossible; et ils entraient dans la	rue Caumartin	, quand, tout à coup, éclata derrière eux un bruit
education	l'emmena jusqu'à l'hôtel de la	rue Tronchet	, dans le logement préparé pour l'autre. Les fleurs n'
education	étaient partis. A l'angle de la	rues Saint-Honoré	, des hommes en blouse le croisèrent en criant: « Non
education	. Des compagnies de la ligne occupaient les	rues Saint-Thomas	et Fromanteau. Une barricade énorme bouchait la rue de Valois.
education	et Fromanteau. Une barricade énorme bouchait la	rue de Valois	. La fumée qui se balançait à sa crête s'entrouvrit,
education	s'en retournèrent au Palais-Royal. Devant la	rue Fromanteau	, des cadavres de soldats étaient entassés sur de la paille.
education	barricades du quartier Latin, s'était battu	rue Rambuteau	, avait sauvé trois dragons, était entré aux Tuileries avec la
education	, et lui annonça qu'il existait,	rue Saint-Jacques	, un club intitulé le Club de l'Intelligence. Un nom
education	« Le citoyen Jean-Jacques Langreneux, typographe,	rue Dauphine	voudrait qu'on élevât un monument à la mémoire des martyrs
education	meubles et quitter le bel appartement de la	rue Drouot	. Elle en prit un autre, sur le boulevard Poissonnière,
education	de Frédéric. Puis il l'emmena déjeuner	rue de Chartres	, chez Parly et, comme il avait besoin de se refaire
education	; et, quand ils remontèrent ensemble la	rue Vivienne	, leurs épaulettes se touchaient fraternellement. Re entré chez lui, Frédéric
education	des Gobelins et de l'Hôpital et les	rues Godefroy	et Moufflard. Quatre barricades formaient, au bout des quatre voies
education	Fontainebleau au secours d'un camarade blessé logeant	rue Bellefond	; personne d'abord ne voulut le croire; on examina ses
education	le ranima. Ils tournèrent ensuite par la	rue du Marché-aux-Chevaux	. Le Jardin des Plantes, à droite, faisait une grande
education	accompagna jusqu'à l'École polytechnique. La	rue Saint-Victor	était toute sombre, sans un bec de gaz ni une lumière
education	caporal pour le mener au haut de la	rue Saint-Jacques	, à la mairie du XIIe arrondissement. La place du Panthéon
education	, l'annonçait. Au haut de la	rue Vivienne	, un peloton de mobiles parut. Alors, les bourgeois poussèrent
education	au haut d'une barricade, dans la	rue Lafayette	, un gamin enveloppé d'un drapeau tricolore criait aux gardes nationaux
education	retourna chez lui; car il possédait,	rue Saint-Martin	, une maison où il s'était réservé un pied-à-terre; et
education	les autres, il s'enfonça dans la	rue Hauteville	, fit le tour du Gymnase, revint sur le boulevard,
education	sa fille, à l'entrée de la	rue Saint-Denis	. Ils s'en retournèrent sans rien dire; lui, n'
education	qu'elle allait avec sa bonne dans la	rue Rumford	chercher un médecin. On les laissa passer. Au coin de
education	. Il était sur le trottoir de la	rue Paradis	, devant une maison; et il regardait au second étage,
education	. C'était une succursale intime de la	rue de Poitiers	. Il y rencontra le grand M. A., l'
education	et Dufour. Ces entretiens continuèrent dans la	rue de la Roquette	, bordée par des boutiques, où l'on ne voit que
education	fiacre et partit. Au coin de la	rue de Marbeuf	, il lut sur une planche en grosses lettres: « Maison
education	» « Arnoux ! » « Oui :	rue de Fleurus	! » « Quel numéro ? » « Est -ce que je
education	, » à sept heures du soir,	rue Jacob	. Il avait même son passeport, par précaution; et il
education	dans sa boutique. Mais il logeait toujours	rue Paradis	, car il possédait deux domiciles. Rue Paradis, le portier
education	Paradis, car il possédait deux domiciles.	Rue Paradis	, le portier jura que M. Arnoux était absent depuis la
education	fit conduire chez lui, à Montmartre,	rue de l'Empereur	. Sa maison était flanquée d'un jardinet, clos par une
education	novembre, Frédéric, en passant dans la	rue de Mme	Arnoux, leva les yeux vers ses fenêtres, et aperçut contre
education	il avait vu au premier dîner dans la	rue de Choiseul	, qui ensuite avait été chez Rosanette, était revenu chez Mme
education	le faubourg Saint-Martin. Au coin de la	rue de Provence	, Frédéric mit pied à terre pour gagner les boulevards. Il
education	Tous les soirs, régulièrement, depuis la	rue de Grammont	jusqu'à la rue Montmartre, il se traîne devant les cafés
education	depuis la rue de Grammont jusqu'à la	rue Montmartre	, il se traîne devant les cafés, affaibli, courbé en
envers	sans caractère assises entre le Terrain, la	rue d'Arcole	, la cathédrale et le quai de la Cité. Ce point
envers	en cas de convenance mutuelle. S'adresser	rue Chanoinesse	, près Notre- Dame, à monsieur Millet, épicier, qui
envers	émût à cette parole, et il revint	rue Chanoinesse	, où il ne voulait déjà plus aller. Le prêtre et
envers	un que l'autre d'entrer dans la	rue Massillon	, qui fait face au petit portail nord de la cathédrale,
envers	la cathédrale, de tourner ensemble dans la	rue Chanoinesse	, à l'endroit où, vers la rue de la Colombe

SAGEO'2016 – Nice, 6-9 décembre 2016

Tableau 3. Les concordances obtenues avec la requête du Tableau 2

La plateforme TXM permet des analyses statistiques sur les occurrences de mots ou de syntagmes en fonction d'entités définies par l'utilisateur : romans, auteurs, période d'édition ou de narration, etc. (Figures 1 et 2).

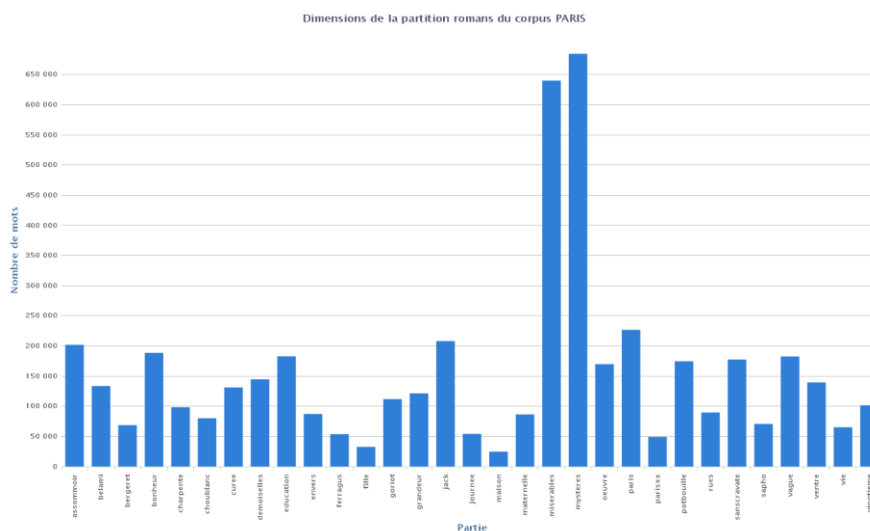


Figure 1. Le nombre de mots par roman

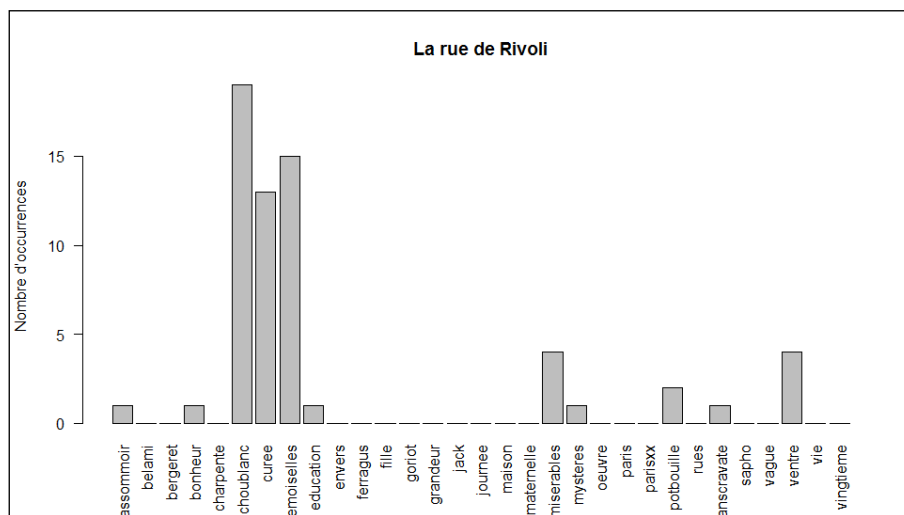


Figure 2. Le nombre de mentions de la rue de Rivoli par roman

3.2. La constitution d'un index géohistorique des lieux littéraires parisiens

La localisation des Entités Spatiales Nommées trouvées dans les textes nécessite une recherche approfondie dans les sources historiques. Il n'existe par exemple pas d'index constitué des rues de Paris portant indication des changements de nom au cours du temps. Il n'y a pas non plus d'annuaire général ou global des théâtres ou des églises existant aux différentes époques. *Gazettier* est le nom couramment donné dans la littérature aux atlas toponymiques. Les *Gazettiers* généralistes et contemporains tels que Geoname¹ s'avèrent de peu de secours dans ce contexte. Toute cartographie romanesque parisienne nécessite la constitution d'un *Gazettier* historique, intégrant le plus grand nombre possible d'Entités Spatiales Nommées précisément localisées à Paris. Les informations existent dans de nombreux documents dispersés. Certaines informations se trouvent directement sur le Web (sur Wikipedia ou DbPedia ou sur divers sites spécialisés), d'autres dans des documents conservés dans les archives, dont certains sont publiés sous forme de collections en ligne. Les localisations précises des entités se trouvent principalement dans les cartes et plans historiques de Paris. Il est donc indispensable de constituer parallèlement au corpus des textes de roman un corpus de cartes devant répondre à cet objectif de localisation. Il servira aussi à documenter l'espace de référence des romans en facilitant le repérage cartographique des occurrences de lieu trouvées dans les textes.

Pour le début de la période (1830), nous avons pris comme référence principale les données du plan Vasserot, disponibles au téléchargement en ligne sur le site du projet Alpage² (Figure 3) (Noizet, Bove, et Costa, 2013). Au plan lui-même, superbe document accessible en format raster, ont été ajoutés par le projet Alpage les limites des bâtiments et les filaires de voies au format vecteur. On constate cependant que seul un faible nombre de noms de rues sont renseignés dans la base. Pour la période de la fin du siècle et les années 1900, alors que les limites de Paris se sont fortement étendues suite à l'annexion en 1860 des communes ou parties de communes périphériques et aux fortes transformations d'Hausmann dans le centre de Paris, nous nous appuyons sur un plan de rue courant de l'époque. D'autres pourront être ajoutés. Ces plans sont projetés dans le référentiel contemporain du plan de rue disponible en *opendata* sur le site de l'APUR³ afin de permettre ultérieurement une visualisation des localisations romanesques sur les plans actuels, dans une optique de développement d'un tourisme littéraire.

Un premier travail de référencement des rues a montré que les données disponibles dans la base de données d'Alpage et dans le fichier de rues contemporains ne permettaient de retrouver qu'environ un tiers des noms de rues présents dans les 31 romans. Un travail de recherche est en cours pour compléter ce *gazettier* et l'étendre, au-delà de la voirie, aux églises et aux théâtres en priorité. La

¹ <http://www.geonames.org/>

² <http://alpage.huma-num.fr>

³ <http://www.apur.org/article/donnees-disponibles-open-data>.

SAGEO'2016 – Nice, 6-9 décembre 2016

structure de données adoptée permettra d'associer toutes les Entités Spatiales Nommées trouvées dans les romans à un nom de rue ayant existé à l'époque en fondant l'association sur la date du roman ou l'époque du récit. Il est à noter que constituer un tel *gazetier* des lieux parisiens du XIXe reste une tâche d'ampleur raisonnable même s'il concerne un espace urbain vaste et complexe. Il est d'abord finalisé vers la cartographie des espaces romanesques d'une période donnée, 1830-1914. Il peut aussi être partagé avec d'autres projets visant la connaissance de l'espace parisien à la même époque. Il peut enfin être conçu comme un projet collaboratif. On sait l'intérêt grandissant pour les initiatives de localisation des œuvres culturelles, qu'elles soient cinématographiques ou littéraires (Joliveau, 2009), dont le site comme Placing Literature¹ est une bonne illustration. Il est possible d'envisager ce *gazetier* comme une application de ce qu'on pourrait appeler suite à (Goodchild, 2007) une démarche d'information géographique littéraire volontaire.

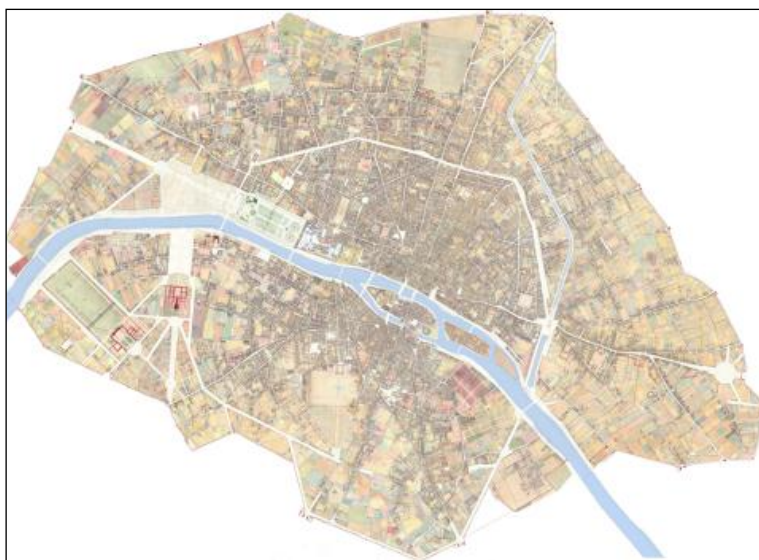


Figure 3. Le plan Vasseroz tiré du site Alpage (source : <http://alpage.huma-num.fr>)

3.3. La géovisualisation et l'analyse spatiale

Les Entités Spatiales Nommées localisées dans les romans peuvent facilement être associées à leur position géographique dans un SIG. Il devient alors aisé de construire des requêtes thématiques et spatiales, de produire des géovisualisations et des cartes concernant l'ensemble du corpus, un roman, un auteur, une période ou un quartier particuliers. Nous avons fait des tests à partir des Entités Spatiales Nommées de type « rue » déjà localisées, qui correspondent à un tiers environ des

¹ <http://www.placingliterature.com/>

Cartographie romanesque de Paris

900 rues repérées avec TXM dans les 31 romans. Plusieurs formes de géovisualisation ont été testées pour représenter les nombres de citations des rues : par ligne ou par point, par symboles textuels, qualitatifs ou quantitatifs (Figures 4 à 7). Le choix de la meilleure visualisation dépend des objectifs et du mode de rendu (carte papier, visualisation interactive à l'écran ...). Les modes d'implantations par symboles proportionnels ponctuels ou linéaires sont plus clairs mais sont moins évocateurs du phénomène que les variations de taille de label, en particulier quand ils conservent les orientations des rues (Fig. 4).

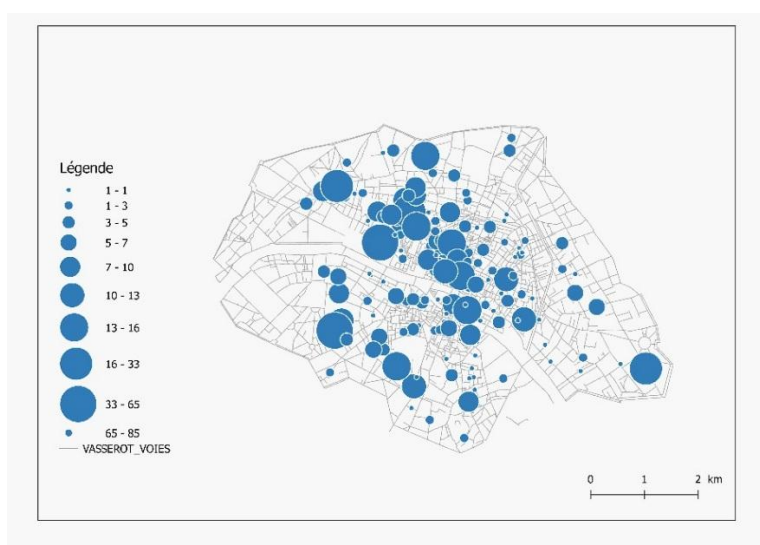


Figure 4. Nombre de citations de rue dans les romans par symboles proportionnels en implantation ponctuelle (échantillon 1/3)

SAGEO'2016 – Nice, 6-9 décembre 2016

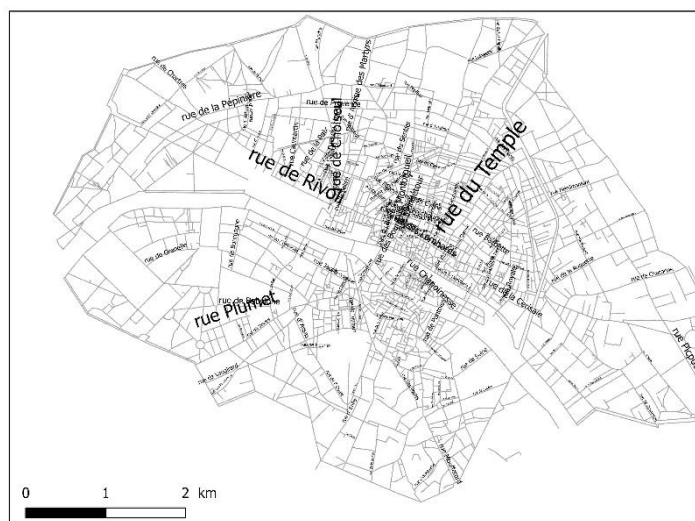


Figure 5. Nombre de citations des rues dans les romans par taille de label proportionnelle en implantation ponctuelle (échantillon 1/3)

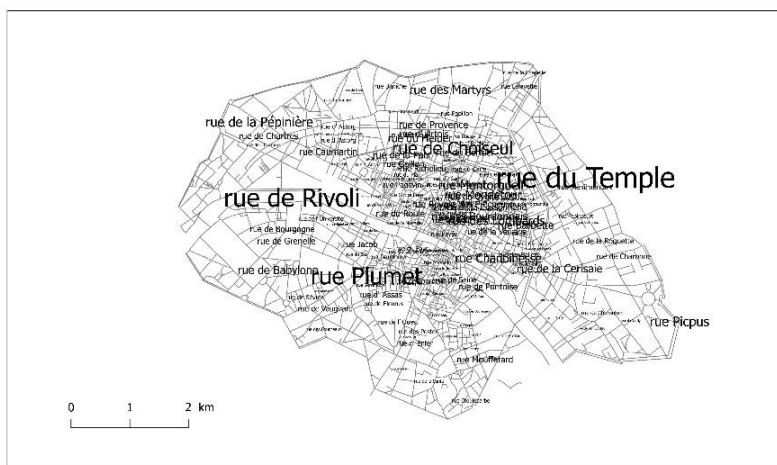


Figure 6. Nombre de citations de rues dans les romans par taille de label proportionnelle en implantation linéaire (échantillon 1/3)

Cartographie romanesque de Paris



Figure 7. Nombre de citations de rues dans les romans par symboles proportionnels en implantation linéaire (échantillon 1/3)

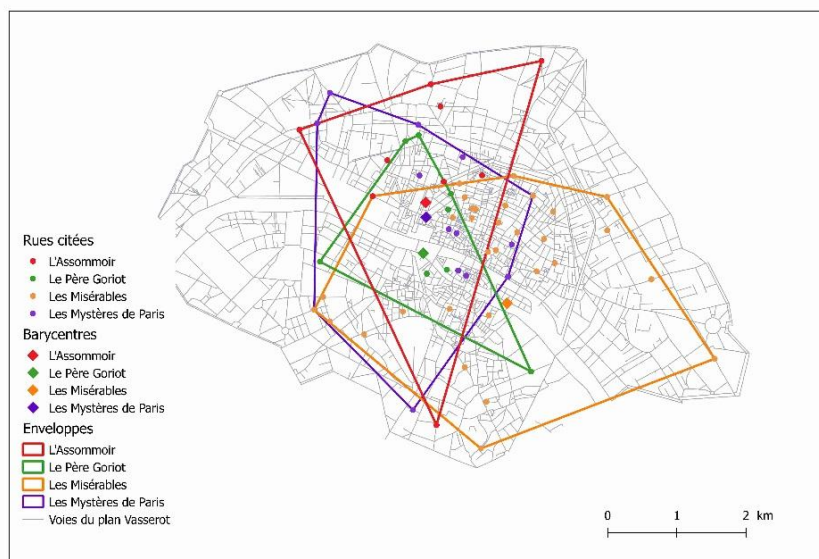


Figure 8. Les citations de rues dans quatre romans (semis de point, enveloppes convexes et barycentres) (échantillon 1/3)

SAGEO'2016 – Nice, 6-9 décembre 2016

L'usage de fonctions de mesure et d'analyse spatiale permet de synthétiser les informations de manière à comparer les empreintes spatiales de différents romans ou différents auteurs. Les Figures 8 à 10 illustrent l'emploi de différents indicateurs classiques de statistiques spatiales pour résumer des semis de points : enveloppes convexes, centres moyens ou barycentres pondérés, ellipses de dispersion standard et interpolations spatiales à partir des rues nommées dans quatre romans. Cela permet de simplifier le tracé des figures. On constate sur l'échantillon de voies concerné que les empreintes de *L'Assommoir*, du *Père Goriot*, *des Misérables* et des *Mystères de Paris* se chevauchent au centre de la ville et se différencient en allant vers la périphérie. Ces calculs nécessitent cependant tous de réduire le linéaire de rue à son centroïde et conduisent à perdre une information importante.

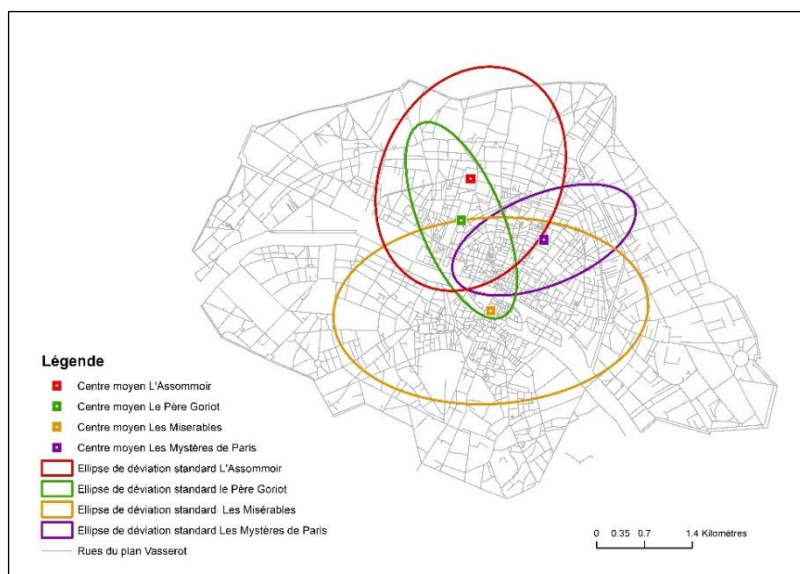


Figure 9. Les citations de rues dans quatre romans (centres moyens et ellipses spatiales de dispersion) (échantillon 1/3)

Cartographie romanesque de Paris

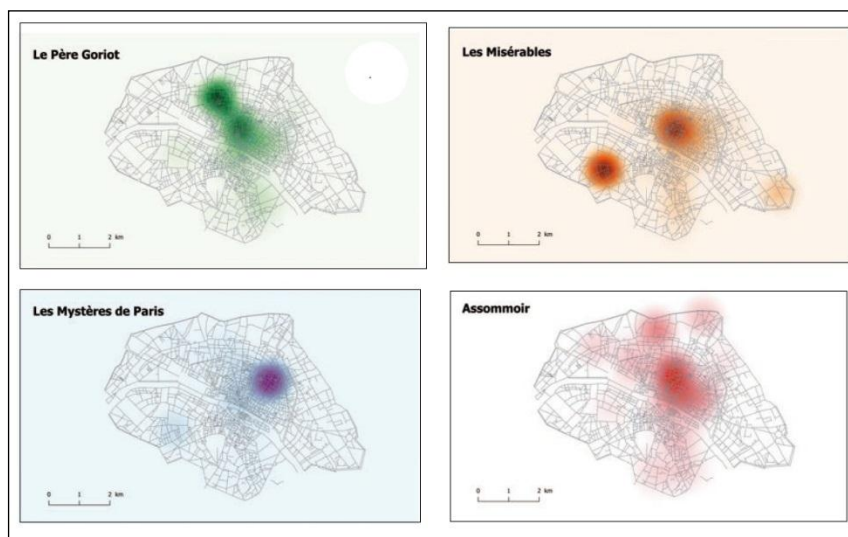


Figure 10. Les citations de rues dans quatre romans par interpolation spatiale (échantillon 1/3)

4. Discussion et perspectives

Les premiers résultats de ce projet démontrent l'intérêt de notre proposition méthodologique pour produire une cartographie romanesque des romans parisiens du XIXe siècle à partir d'une recherche automatique des Entités Spatiales Nommées. Les requêtes TXM semblent fonctionner de manière efficace et répondre aux besoins. Des problèmes liés à des caractères non pris en compte dans les requêtes ont été corrigés par les auteurs de la plateforme. D'autres problèmes inattendus sont apparus, telle que l'existence de deux signes typographiques différents pour les apostrophes qui ont contraint à homogénéiser le corpus. Des analyses complémentaires doivent maintenant être menées pour repérer des cas où les rues non trouvées ne proviendraient pas des limites actuelles de notre Gazetteer : graphies erronées, structures de phrases complexes et résoudre la question des noms de rues inventés par l'auteur. Le choix de TXM pour extraire des Entités Nommées était lié à deux motivations principales. D'abord la plateforme est relativement simple à mettre en œuvre sans compétences informatiques trop spécialisées. Ensuite, elle propose une interface interactive pour explorer les textes par l'intermédiaire des Entités Spatiales Nommées, ce qui est particulièrement utile dans notre cas d'usage. Il existe d'autres outils de reconnaissance d'entités nommées ou d'autres plateformes de traitement de corpus textuels qui intègrent ce type de fonction comme GATE par exemple (<https://gate.ac.uk/>). Un travail d'analyse comparée de leurs performances et de l'intérêt qu'ils présentent pour un projet comme le nôtre est à faire.

Le travail de constitution de l'index géographique des Entités Spatiales Nommées n'étant pas terminé, il n'est pas encore possible de valider l'intérêt des géovisualisations pour l'analyse des empreintes spatiales des différents romans ou auteurs. Mais même sans disposer de cartes ou d'indicateurs exhaustifs, il reste

SAGEO'2016 – Nice, 6-9 décembre 2016

indéniable que l'accès direct au corpus de texte par l'intermédiaire des noms de lieux transforme fortement les manières d'explorer l'espace et les paysages romanesques parisiens. Disposer d'outils d'interrogation, de visualisation et d'analyse spatiale s'avère utile pour émettre des hypothèses et synthétiser les résultats liés à la spatialisation des romans.

La prochaine étape est d'utiliser la plate forme textométrique pour caractériser automatiquement le contenu sémantique associé aux Entités Spatiales Nommées en recherchant les qualificatifs employés pour les décrire. Par exemple, il sera possible de préciser si une rue est grande, sombre, animée ... ou bien quels personnages sont convoqués dans tel ou tel lieu. Une autre piste serait de représenter la dynamique temporelle des lieux dans la chronologie du récit. Il en effet possible d'associer à chacun des citations sa place dans le texte par le chapitre ou même par son rang dans l'ordre des mots. L'objectif à terme serait de construire une plateforme facilitant l'analyse spatiale et géographique des Entités Spatiales Nommées ouvertes à tous les romans évoquant Paris à la même époque. Cette plateforme pourrait être mobilisée par les opérateurs du tourisme littéraire, en donnant un accès *in situ* aux extraits de texte associés à des lieux explicitement nommés. Comme nous l'avons déjà écrit, le *geocrowdsourcing* (ou information géographique volontaire) permettrait de documenter littérairement l'espace parisien et de spatialiser les œuvres littéraires traitant Paris au XIXe siècle.

Pour conclure, soulignons que l'équipe ne pense pas que la dimension géographique des romans ni même les questions de localisation romanesque se réduisent à la cartographie d'Entités Spatiales Nommées trouvées dans les textes littéraires. Mais celles-ci, qui sont assez aisées à localiser de manière automatique, peuvent servir de point d'ancrage pour cartographier des lieux non nommés explicitement et explorer la dimension géographique des textes romanesques, moyennant la création de nouvelles méthodes de recherche d'information géographique textuelle.

NB. Merci aux relecteurs pour leurs remarques judicieuses qui ont permis d'améliorer ce texte.

Bibliographie

Bulson E. *Novels, Maps, Modernity: The Spatial Imagination, 1850-2000*. 1^{re} éd.[s.l.] : Routledge, 2007. 188 p.ISBN : 0-415-97648-0.

Cooper D., Donaldson C., Murrieta-Flores P. *Literary Mapping in the Digital Age*. [s.l.] : Routledge, 2016. 457 p.ISBN : 978-1-317-10455-1.

Gaio M., Sallaberry C., Etcheverry P., Marquesuzaa C., Lesbegueries J. « A global process to access document's contents from a geographical point of view ». *J. Vis. Lang. Comput.* 2008. Vol. 19, n°n° 1, p. 3–23.

Goodchild M. F. « Citizens as sensors ; the world of volunteered geography ». *GeoJournal*. 2007. Vol. 69, n°4, p. 211-221.

Cartographie romanesque de Paris

Gregory I., Donaldson C., Murrieta-Flores P., Rayson P. « Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research ». *International Journal of Humanities and Arts Computing* [En ligne]. mars 2015. Vol. 9, n°1, p. 1-14. Disponible sur : < <https://doi.org/10.3366/ijhac.2015.0135> >

Heiden S. « The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme ». In : *24th Pacific Asia Conference on Language, Information and Computation*. [s.l.] : Institute for Digital Enhancement of Cognitive Development, Waseda University, 2010. p. 389-398. Disponible sur : < <https://halshs.archives-ouvertes.fr/halshs-00549764/document> > (consulté le 16 juin 2016)

Heiden S., Magué J.-P., Pincemin B. « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ». In : *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*. [s.l.] : Edizioni Universitarie di Lettere Economia Diritto, 2010. p. 1021-1032. Disponible sur : < <https://halshs.archives-ouvertes.fr/halshs-00549779/document> > (consulté le 16 juin 2016)

Joliveau T. « Connecting Real and Imaginary Places through Geospatial Technologies: Examples from Set-Jetting and Art-Oriented Tourism ». *Cartographic Journal*. février 2009. Vol. Vol. 46, n°1, Cinematic Cartography Special Issue, p. 36-45.

Moretti F. *Atlas du roman européen*. Paris : Editions du Seuil, (édition originale italienne : 1997), 2000.

Moretti F. *Graphes, cartes et arbres. Modèles abstraits pour une autre histoire de la littérature*. Paris : Les prairies ordinaires (Edition originale en anglais. 2005), 2008.

Noizet H., Bove B., Costa L. *Paris, de parcelles en pixels : Analyse géomatique de l'espace parisien médiéval et moderne*. Saint-Denis (Seine-Saint-Denis); Paris : Coédition PU Vincennes, 2013. 354 p. ISBN : 978-2-84292-364-8.

Piatti B., Bär H., Reuschel A., Hurni L., Cartwright W. « Mapping Literature: Towards a Geography of Fiction ». In : *Proceedings of the Art and Cartography Conference, Berlin Heidelberg: Springer*. Vienna : [s.n.], 2008. Disponible sur : < <http://www.literaturatlas.eu/2012/03/23/ein-literarischer-atlas-europas-poster/> > (consulté le 8 novembre 2010)

Sallaberry C. *Recherche d'information géographique dans les corpus textuels*. Collection Systèmes d'Information Géographique. [s.l.] : ISTE, 2014. 190 p. ISBN : 978-1-78405-012-2.



Session 2

**Infrastructures de données
spatiales, observatoires,
géographie participative,
géoprospective**

Infrastructures de Données Géographiques et observatoires de recherche en environnement : un exemple de mise en œuvre

GOURMELON Françoise, ROUAN Mathias, NABUCET Jean

LETG UMR 6554 CNRS (<http://letg.cnrs.fr/>)

francoise.gourmelon@univ-brest.fr, jean.nabucet@uhb.fr, mathias.rouan@univ-brest.fr

RESUME. Cet article décrit une Infrastructure de Données Géographiques (Indigeo) élaborée dans un cadre académique associant un laboratoire de recherche et plusieurs observatoires en environnement. La solution déployée, basée sur des outils libres, repose sur un catalogue de métadonnées et un serveur de données géoréférencées, adossés à un portail web disposant de deux visualiseurs cartographiques dont l'un (geoCMS) prend en charge la dimension temporelle des données. Indigeo met actuellement à disposition en accès libre plus de 500 couches d'informations et leurs métadonnées et permet la co-visualisation des données d'une quinzaine d'IDG généralistes ou thématiques interopérables. Elle offre aussi une plus-value en termes d'analyse de données spatio-temporelles. Néanmoins si Indigeo rend actuellement plusieurs services en termes de stockage, d'analyse, de partage et de porter à connaissances de l'information géographique scientifique, elle demeure sous-utilisée par la communauté concernée, ce qui se traduit par une alimentation timide du catalogue et le peu de projets interdisciplinaires qui la mobilisent. Des solutions sont discutées afin de contrecarrer cette sous-utilisation, relativement commune, de ces dispositifs socio-techniques.

ABSTRACT. We describe a Spatial Data Infrastructure developed in an academic context combining a research laboratory and several environmental observatories. The technical solution is based on free tools that meet the interoperability standards of the OGC. It includes a metadata catalog and a geo-referenced data server, backed by a web portal with two map viewers which one (GeoCMS) supports the time dimension of the data. Indigeo currently offers open access to nearly 500 GIS data layers and metadata, and allows the visualization of data provided by fifteen generalist and thematic SDI. It also includes added value in terms of analysis of space-time data. However this platform is still under-used which means that most data come from a minority of contributors. Solutions are discussed to counteract this relatively common under-used of these socio-technical systems.

MOTS-CLES : Infrastructures de Données Géographiques (IDG), développement technique, métadonnées, observatoire de recherche en environnement, usages, visualisation

SAGEO'2016 – Nice, 6-9 décembre 2016

Infrastructures de Données Géographiques et observatoires de recherche en environnement

KEYWORDS: metadata, Spatial Data Infrastructure (SDI), SDIs use, scientific environmental observatory, technical development, visualization

1. Introduction

Depuis les années 1990, les Infrastructures de Données Géographiques se développent à tous les niveaux de la société (Masser, 2010) afin de faciliter l'accès et l'utilisation des informations géographiques produites par une gamme de plus en plus large de producteurs institutionnels ou « volontaires » (Crompvoets *et al.*, 2004). Ces dispositifs, accessibles sur le web, font référence à des solutions dynamiques qui regroupent des informations, des réseaux informatiques, des normes et standards, des accords organisationnels et des ressources humaines afin de coordonner le partage, l'accès et la gestion des informations géographiques (Rajabifard *et al.*, 2002). Ils répondent à plusieurs besoins stratégiques d'aide à la décision, de coordination, de réduction de l'investissement en matière d'information géographique (Harvey *et al.*, 2012).

En France, sous l'impulsion de la directive européenne INSPIRE et du mouvement international d'ouverture des données publiques, des IDG généralistes et thématiques ont été mises en œuvre par l'Etat et ses services déconcentrés, par des collectivités territoriales, et par divers organismes publics (Afigeo, 2014). Si certains opérateurs¹ gèrent des IDG dédiées à des domaines spécifiques, peu d'organismes nationaux de recherche se sont lancés dans une démarche centralisée les engageant sur le long terme. Même si quelques établissements, à l'image du CNRS (INEE) et du MNHN, proposent à leurs laboratoires des services (UMS BBEES) dans le domaine des bases de données, leur soutien en faveur d'initiatives concernant leur patrimoine informationnel s'est plutôt exprimé à travers différents types d'observatoires scientifiques (OSU de l'INSU, ORE du Ministère de la Recherche, réseau national des zones ateliers du CNRS...) dont certains sont regroupés au sein de systèmes d'observation et d'expérimentation au long terme pour la recherche en environnement (SOERE² de l'Allenvi). Ces observatoires scientifiques ont pour objectif d'observer (suivre, analyser, comprendre), sur une portion d'espace représentative d'une entité territoriale, les dynamiques socio-environnementales résultant d'interactions de systèmes socio-économiques et biophysiques. Ils sont un lieu de production, d'échange et de partage d'informations et de connaissances qui s'inscrit dans la durée (Libourel *et al.*, 2009). La notion de « mémoire environnementale », définie

¹ Par exemple l'IFREMER ou le SHOM pour le domaine marin et littoral (sextant.ifremer.fr/ ; data.shom.fr/)

² «Un Soere est un réseau de dispositifs élémentaires d'observation ou d'expérimentation portant sur le même objet ou partageant des paramètres observables et mesurables communs à tous les sites du réseau. L'acquisition de ces paramètres, leur mesure, le traitement et l'archivage des données sont effectués de manière sécurisée, pérenne et continue, à partir de chaque dispositif. Pour les utilisateurs, membres ou non de la communauté scientifique, ces réseaux facilitent l'accès à toutes les données disponibles. » (<http://www.allenvi.fr/groupestransversaux/infrastructures-de-recherche/les-soere/gouvernance-et-financement>)

SAGEO'2016 – Nice, 6-9 décembre 2016

comme la somme des représentations explicites, persistantes et structurées des données, des connaissances, des modèles et des savoir-faire scientifiques attachés à un système d'observation à long terme de l'environnement en vue d'en faciliter l'accès, le partage et la réutilisation (Guarnieri *et al.*, 2003), s'y applique pleinement. Dans ce système opérant fondé sur les interactions entre organisation, documentation et observation, une IDG répond potentiellement aux objectifs de la mémoire d'observation.

De la récente prise de conscience du patrimoine informationnel détenu par les dispositifs de recherche (laboratoires, observatoires...) et des enjeux nationaux et internationaux d'ouverture des données publiques, résultent quelques initiatives de conception d'IDG dans le domaine académique. Cet article présente le contexte d'élaboration, les cinq composantes (services et technologies, normes et standards, ressources humaines et communautés, information géographique, accords organisationnelles) et deux cas d'utilisation d'une IDG mise en œuvre dans un laboratoire de recherche en géographie de l'environnement. La discussion traite des apports et des limites du dispositif et évoque plusieurs pistes qui pourraient conduire à une meilleure appropriation de la part de ses utilisateurs potentiels, producteurs ou simples consommateurs d'information géographique.

2. L'Infrastructure scientifique de Données et d'Informations GEOspatialisées sur l'environnement (*Indigeo*)

2.1. Contexte de sa mise en œuvre

L'UMR LETG³ multi-site occupe une position originale dans le paysage scientifique de l'ouest de la France de par son implication dans trois Observatoires des Sciences de l'Univers (OSU-IUEM⁴, OSURennes⁵, OSUNantes-Atlantique⁶) et plusieurs systèmes nationaux d'observation (réseau national des zones ateliers du CNRS⁷, système national d'observation Dynalit⁸). Laboratoire de Géographie de l'environnement déployant depuis plusieurs années une expertise géomatique, LETG s'intéresse dès le début des années 2000 aux démarches d'ingénierie des connaissances qui l'amène à mettre en œuvre la notion de mémoire environnementale (Guarnieri *et al.*, 2003) dans le cadre de la conception d'un Système d'Information (Faucher *et al.*, 2009) et d'un catalogue des données dédiés à l'observatoire du domaine côtier de l'IUEM. Ces deux dispositifs exploratoires ne seront pas maintenus

³ Implantation du laboratoire sur les sites universitaires d'Angers, Brest, Caen, Nantes et Rennes

⁴ <http://www-ieuem.univ-brest.fr/observatoire>

⁵ <https://osur.univ-rennes1.fr/>

⁶ <http://www.osuna.univ-nantes.fr/>

⁷ <http://www-ieuem.univ-brest.fr/zabri/fr>

⁸ <http://www.dynalit.fr/fr>

Infrastructures de Données Géographiques et observatoires de recherche en environnement

du fait du peu d'intérêt qu'ils suscitent auprès des membres de l'Institut peu familiers de l'univers de l'information géographique. En dépit de cet échec, le laboratoire développe son expertise dans la mise en œuvre d'IDG notamment en domaine côtier ouest-africain (Gourmelon *et al.*, 2009). A partir de 2010, les injonctions européennes et nationales concernant l'information géographique percolent dans le monde académique. Les opérations de communication relatives à la directive INSPIRE se multiplient, entraînant une prise de conscience croissante des organismes de recherche, même si elle ne se traduit pas directement par des décisions opérationnelles. Profitant de ce contexte renouvelé et de son expérience, le laboratoire LETG propose en 2011 le projet d'élaboration d'une IDG mutualisée avec les observatoires entrant dans son périmètre. L'objectif est de répondre conjointement aux besoins du laboratoire et des OSU, en termes de stockage, de documentation et de mise à disposition des données essentiellement géographiques et des séries temporelles d'observation. Une équipe projet composée de 5 personnes rattachées au laboratoire et aux OSU et d'un prestataire informatique se met rapidement en place.

2.2. Un développement itératif : la méthode Agile

Selon Florent Lothon sur son site l'agiliste⁹, le terme « Agile » fait référence à une gestion de projet itérative et adaptative qui prend le contre-pied des approches traditionnelles prédictives et séquentielles. Caractérisées par une expression détaillée et validée du besoin, suivie d'une phase de développement souvent longue, elles conduisent à un livrable ne répondant pas toujours aux besoins. La méthode Agile propose au contraire d'impliquer le commanditaire tout au long du projet en considérant que le besoin ne peut être figé et qu'il est nécessaire de s'y adapter continuellement. Cela ne se fait pas sans règle, et le « manifeste agile¹⁰ », rédigé en 2001 par 17 experts informatiques, en pose les quatre valeurs et les 12 principes fondateurs. Le développement d'Indigeo s'est donc déroulé selon une méthode Agile comportant plusieurs cycles itératifs mettant à chaque fois à disposition de l'équipe un outil partiel mais utilisable et intégrant de nouvelles fonctionnalités à tester, valider ou à modifier. Au final le produit réalisé ne dispose pas de toutes les fonctionnalités envisagées initialement mais en possède certaines qui ont émergé lors du développement et qui répondent aux besoins des utilisateurs.

2.3. Composantes

Au sens de Rajabifard *et al.* (2002), une IDG repose sur cinq composantes en interaction, représentées dans Indigeo de la manière suivante.

⁹ <http://www.agiliste.fr>

¹⁰ <http://agilemanifesto.org/>

SAGEO'2016 – Nice, 6-9 décembre 2016

2.3.1. Services et Technologies

Le point d'entrée sur Indigeo est un site web (<http://indigeo.fr>) réunissant les accès aux différents services (catalogue de métadonnées, visualiseur et services web OGC¹¹) et à un guide d'aide à la saisie et à la publication des métadonnées. La solution déployée est geOrchestra¹², une suite logicielle libre, interopérable et modulaire, créée en 2009 pour répondre à la directive INSPIRE. Elle connaît un certain succès comme en témoigne son utilisation par une vingtaine d'IDG généralistes mises en œuvre à différents échelons : national (géoBolivia ; <http://geo.gob.bo/>), régional (PIGMA ; <http://www.pigma.org/>, géoBretagne ; <http://cms.geobretagne.fr/>), local (Rennes métropole ; <http://www.data.rennes-metropole.fr/les-donnees/catalogue/>). Cette suite logicielle est composée de modules indépendants et interopérables avec lesquels on compose une infrastructure « à la carte ». Pour Indigeo, trois modules ont été installés dans leurs versions correspondant à geOrchestra 14.01 : un catalogue de métadonnées (geoNetwork), un serveur de données géoréférencées (geoServer) et un visualiseur cartographique avancé (mapfishapp). Afin de répondre aux besoins exprimés par les scientifiques, plusieurs services et outils complémentaires ont été développés. Ils concernent respectivement l'authentification centralisée (geoAuth), la génération de graphiques dynamiques et un visualiseur ergonomique et évolutif (geoCMS¹³) permettant entre autre de prendre en compte la dimension temporelle de certaines données.

Le catalogue de métadonnées (geoNetwork). La gestion des métadonnées est assurée par geoNetwork, applicatif web développé et maintenu par une communauté active, qui assure la mise à disposition des métadonnées de manière standardisée et interopérable. Dans une version test d'Indigeo, une session de formation à la production de métadonnées réunissant une quinzaine de géographes a mis en évidence la complexité de la norme et a conduit à concevoir des modèles de métadonnées simplifiés adaptés aux principaux types d'information géographique et un guide d'aide à la saisie.

Le serveur de données géoréférencées (geoServer). Les données administrées sont principalement des couches d'information géographique vectorielles et des images de télédétection. L'application geoServer permet de gérer leur stockage et leur diffusion à travers des entrepôts. L'accès aux données est réalisé via des services web standardisés et interopérables qui permettent l'affichage et/ou le téléchargement des informations disponibles directement sur Indigeo et/ou via un client SIG (QGIS, ArcGIS®ESRI). geoServer permet non seulement de gérer les droits d'accès à chaque entrepôt de données et ainsi de traiter du caractère privé ou public de la donnée, mais aussi d'intégrer des bases de données à référence spatiale de type PostGIS, ce qui présente un intérêt certain dans le cas de la mise à disposition dynamique de séries de données spatio-temporelles.

¹¹ Open Geospatial Consortium : <http://www.opengeospatial.org/>

¹² <http://www.georchestra.org/>

¹³ CMS pour *Content Management System* (système de gestion de contenu) est un terme utilisé généralement pour décrire les systèmes de mise à jour dynamique de sites web

Infrastructures de Données Géographiques et observatoires de recherche en environnement

Le visualiseur cartographique avancé (mapfishapp). Fourni avec la suite geOrchestra, il se présente comme un logiciel de type SIG accessible via un navigateur web. En termes d'ergonomie, il est adapté à des utilisateurs experts et propose des fonctionnalités telles que la visualisation de couches d'information (flux OGC ou fichiers), la modification de leur symbologie, leur interrogation par des requêtes attributaires et spatiales, l'exportation des résultats, l'édition des données et leur analyse grâce à l'ajout de modules complémentaires, ainsi que la cartographie.

Le visualiseur « découverte et valorisation » (geoCMS). Il est basé sur des technologies relativement récentes (Ruby on Rails, AngularJS et la librairie cartographique Leaflet.js) qui améliorent l'ergonomie et les possibilités de découverte et de valorisation de l'information géographique (navigation spatio-temporelle, visualisation de graphiques et approche projets). Ce visualiseur permet d'une part de structurer l'organisation des couches d'information dans un catalogue administré depuis une interface réservée aux gestionnaires du portail (le backend), et d'autre part de personnaliser la présentation du contenu de la donnée (table attributaire ou valeur) sous la forme d'une info-bulle affichée lors de l'interrogation d'une couche d'information. Cette personnalisation est réalisée à partir d'un éditeur WYSIWYG¹⁴ dans l'interface d'administration et permet de rendre l'information accessible à tout type de public en y associant du contenu textuel mis en forme, des images, des documents ou des applications web (graphiques dynamiques par exemple). Le geoCMS permet également la navigation dans un jeu de données spatio-temporelles en interprétant automatiquement cette caractéristique et en générant un curseur adapté à leur étendue temporelle. Enfin il dispose d'une fonctionnalité de sauvegarde de cartes dans des dossiers permettant d'accéder à des ensembles cohérents de couches d'information en fonction de différents critères (par projet scientifique ou par dispositif scientifique par exemple). Ces cartes peuvent être facilement intégrées à des pages web (un site, un article en ligne, un blog...).

Le dernier module de la suite logicielle mise en œuvre est *un outil d'authentification et de gestion centralisée des utilisateurs (geoAuth)* permettant de créer des comptes, de les affecter à des groupes et de leur attribuer des droits sur les outils (geoServer, geoNetwork, et par la suite geoCMS¹⁵).

¹⁴*What You See Is What You Get* : une interface intuitive qui permet à l'utilisateur de voir à l'écran à quoi ressemblera le résultat final.

¹⁵Il intégrera très prochainement une fonctionnalité d'authentification par le protocole *shibboleth* notamment utilisé par la fédération d'identité recherche et éducation maintenue par RENATER. Cette fonctionnalité permettra à tout membre de la fédération de se connecter à Indigeo à l'aide de ses identifiants institutionnels.

SAGEO'2016 – Nice, 6-9 décembre 2016

2.3.2. Normes et standards

Indigeo repose sur différentes normes :

- pour la description et la recherche des métadonnées (service de découverte INSPIRE), la norme internationale ISO 19115 de définition des métadonnées pour l'information géographique, ainsi que sur les normes ISO 19110 de catalogage des types d'entités (table attributaire) et ISO 19139 de description du schéma d'implémentation XML ;
- pour l'interrogation et le moissonnage de ses métadonnées, le standard Catalogue Service for the Web (CSW) ;
- pour la visualisation des données (service de consultation INSPIRE), les standards de l'OGC et principalement le Web Map Service (WMS) ;
- pour le téléchargement (service de téléchargement INSPIRE), les standards de l'OGC Web Feature Service (WFS) pour les données vectorielles / et Web Coverage Service (WCS) pour les données raster.

2.3.3. Ressources humaines et communautés

En termes de ressources humaines, Indigeo s'appuie actuellement sur une équipe technique de proximité (5 cartographes et 3 administrateurs de données, 2 administrateurs système, le prestataire) en relation avec d'autres équipes techniques impliquées dans la mise en œuvre d'IDG scientifiques dans l'ouest de la France. Les usagers d'Indigeo regroupent les producteurs d'information géographique (42 relevant de 13 laboratoires du domaine des Sciences de l'environnement), et les utilisateurs dont certains sont également producteurs. A la date du 6 avril 2016, 5500 visites ont été comptabilisées depuis juin 2013 sur le site web éditorial <http://indigeo.fr> qui est le point d'entrée vers les outils de l'IDG, les visualiseurs, le catalogue et les services web.

2.3.4. Information géographique

A son ouverture en juin 2013, Indigeo comptabilisait 214 fiches de métadonnées et 106 couches d'informations publiées, essentiellement de nature vectorielle (shapefile), et provenant de l'importation d'une plateforme préexistante (Faucher *et al.*, 2009). En avril 2016, elle met à disposition 534 fiches de métadonnées, 551 couches d'informations (vectorielles et matricielles) et 7 séries temporelles provenant de bases de données (Tab. 1). Toutes les données cataloguées sont en accès libre. 83% du contenu informationnel d'Indigeo provient d'un seul laboratoire (LETG) ; le reste étant produit par les autres laboratoires associés aux observatoires.

En plus des données qu'elle stocke, Indigeo donne accès à la visualisation de plus de 6500 couches d'information grâce au moissonnage d'une quinzaine d'IDG thématiques ou généralistes interopérables parmi celles de l'IGN, du SHOM, d'IFREMER (Sextant), des régions Bretagne, Pays de la Loire et Normandie (géoBretagne, géoPAL, géoNormandie), du BRGM ou du MNHM. Réciproquement

Infrastructures de Données Géographiques et observatoires de recherche en environnement

Indigeo est moissonnée par les IDG régionales citées précédemment du fait d'emprises spatiales communes (Bretagne, Pays de Loire, Normandie).

TABLE 1. Types, formats et nombre de données mises à disposition

Vecteur	Raster	Base de données
Shape : 289	ImageMosaic : 140	PostgreSQL : 6
PostGIS : 15	Geotiff : 93	MySQL : 1
	Arcgrid : 67	
	ECW : 1	

2.3.5. Accords organisationnels

Dans sa phase de développement, Indigeo a été soutenue principalement, en termes de ressources humaines et financières, par le laboratoire LETG, les OSU rennais et brestois, la Zone Atelier Brest-Iroise. Peu de temps après son lancement, l'OSUR a opté pour une IDG autonome (Osuris) reposant sur les mêmes outils mais donnant l'accès aux données scientifiques produites par les laboratoires rennais qui l'alimentent. A travers un conventionnement, Indigeo héberge temporairement des portails thématiques en phase de lancement pour des dispositifs scientifiques auxquels adhère le laboratoire et qui ont vocation à s'autonomiser (OSUNA, SNO Dynalit).

3. Cas d'utilisation

Indigeo offre un point d'accès normalisé à de nombreuses données géographiques dont elle garantit la pérennité et le porter à connaissances, en conformité avec les recommandations en vigueur. Elle offre aussi une gamme de services innovants à plusieurs observatoires et services d'observation en environnement dont l'objectif est la production de connaissances fondées sur l'analyse des données entreposées. Deux exemples sont présentés.

3.1. Observatoire du domaine côtier de l'IUEM : consultation de séries temporelles d'observation de l'environnement

L'observatoire du domaine côtier de l'IUEM (<http://www-iuem.univ-brest.fr/observatoire>) a contribué à la conception d'Indigeo en apportant son soutien logistique par la maintenance des serveurs hébergeant la solution et en participant à l'élaboration du cahier des charges. Un de ses principaux besoins est de visualiser les

SAGEO'2016 – Nice, 6-9 décembre 2016

données de séries temporelles d'observation qu'il gère (analyse des eaux, suivi géomorphologique...).

Données d'entrée. Ces séries produisent des données de type paramètres physico-chimiques (ph, température, salinité, nitrates...) qui sont issus soit de prélèvements manuels in situ puis analyse en laboratoire, soit de capteurs automatiques installés sur des équipements autonomes (bouée).

Stockage en base. Les données acquises en continu sont archivées dans des bases de données relationnelles. Les volumes de données produits dépendent de la fréquence des séries. Par exemple, certains prélèvements sont réalisés manuellement selon une cadence hebdomadaire et leur analyse produit entre 3 et 15 paramètres alors que la série automatisée MAREL¹⁶ mesure 7 paramètres toutes les 20 minutes depuis près de 20 ans, ce qui représente un volume de données considérable pour une visualisation dans un navigateur web.

Services et outils mis en œuvre. Dans sa forme basique, une IDG ne permet pas de représenter graphiquement des données temporelles. Pour répondre à ce besoin, une application web (manageChart¹⁷) couplée au visualiseur geoCMS a été développée à l'aide du framework symfony 2 (PHP) et des bibliothèques graphiques highcharts et highstock¹⁸ (javascript). L'application manageChart exploite les fonctionnalités de ces bibliothèques en proposant une « timeline » permettant de zoomer et de naviguer dans le jeu de données de la série temporelle ainsi que différents formats d'exportation des graphiques et des données (jpeg, pdf, csv...). manageChart permet d'établir des connexions à des bases de données locales ou distantes et, par l'intermédiaire de requêtes paramétrables, de générer dynamiquement différents types de graphiques personnalisables (nuage de points, ligne, histogramme, camembert...). Grâce à la fonctionnalité de personnalisation de l'interrogation des couches d'information du visualiseur geoCMS, il devient alors possible d'afficher le graphique correspondant à une couche d'information dans une info-bulle. Le paramétrage de la requête est réalisé lors de l'interrogation du point sur la carte qui renvoie une information d'identification de l'objet spatial transmise via l'URL à manageChart. La requête avec le paramètre correspondant est exécutée et le graphique associé est affiché. Ainsi un seul graphique prédéfini suffit pour visualiser les données de chaque point de prélèvement d'une série d'observation (Fig. 1).

¹⁶ <https://www-ium.univ-brest.fr/observatoire/observation-cotiere/parametres-physico-chimiques/testpeg>

¹⁷ <https://tucuxi.univ-brest.fr/projects/managechart>

¹⁸ Highstock propose des fonctionnalités de groupage des données adaptées à la visualisation de gros volumes : <http://www.highcharts.com/>

Infrastructures de Données Géographiques et observatoires de recherche en environnement

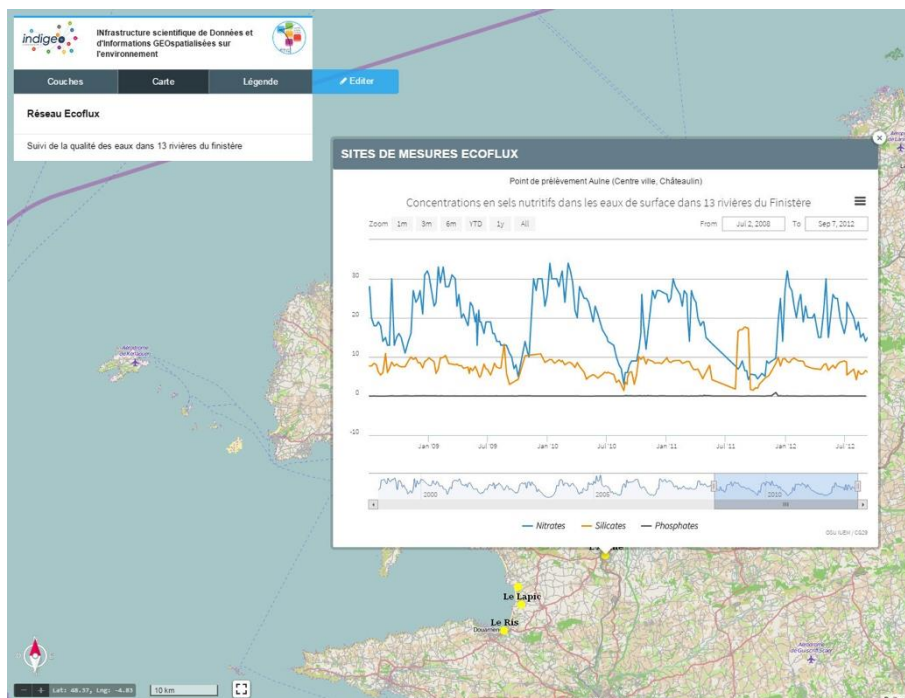


FIGURE 1. Exemple d'affichage d'un graphique relatif à un point de prélèvement du réseau ecoflux de l'OSU IUEM (<http://portail.Indigeo.fr/geocms/maps/portail-jhiqpymc#project>).

3.2. Zone Atelier Brest-Iroise : cartographie spatio-temporelle du trafic maritime

Dans le cadre d'un projet de recherche¹⁹ mené sur la Zone Atelier Brest-Iroise, l'exploitation des données de positionnement des navires issues de l'Automatic Identification System (AIS) a été réalisée pour mettre en évidence les territoires de pratique des activités maritimes. L'automatisation de la visualisation des couloirs de navigation sur Indigeo a fait l'objet d'un développement spécifique.

Données d'entrée. L'AIS fournit des données de positionnement toutes les 2 à 10 secondes pour différents types de navires (Le Guyader *et al.*, 2012). Les données sont transmises à terre par voie hertzienne (VHF) de messages pré-formatés et numérotés.

Stockage en base. Une fois décodés, ces messages délivrent une information géographique ponctuelle caractérisée par plusieurs attributs (date/heure, identification et type du navire, vitesse, direction...) qui est archivée dans une base de données spatio-temporelle (PostgreSQL/Postgis).

¹⁹ <https://www-iuem.univ-brest.fr/pops/projects/DACTARI>

SAGEO'2016 – Nice, 6-9 décembre 2016

Services et outils mis en œuvre. Un des objectifs est d'automatiser l'analyse et la visualisation des données de positionnement. Le concept de « vue matérialisée » en SQL a été mobilisé pour y parvenir. Une vue est une table virtuelle, c'est-à-dire que ses données ne sont pas stockées dans une table de la base de données, dans laquelle il est possible de rassembler des informations provenant de plusieurs tables ou d'une requête qui est exécutée dynamiquement au moment de l'appel de la vue. La vue matérialisée étant exécutée à une fréquence de rafraîchissement définie ou manuelle, les performances d'une table sont conservées tout en automatisant l'analyse des données. Dans un esprit didactique, il a été défini une vue par étape d'analyse ce qui permet de générer des couches d'informations intermédiaires à certaines étapes (Fig. 2). Les fonctionnalités du visualiseur geoCMS permettent d'exploiter les couches d'information produites par cette chaîne de traitement en prenant en charge par exemple le caractère spatio-temporel des données (Fig. 3) et l'affichage de statistiques agrégées sur une zone géographique déterminée.

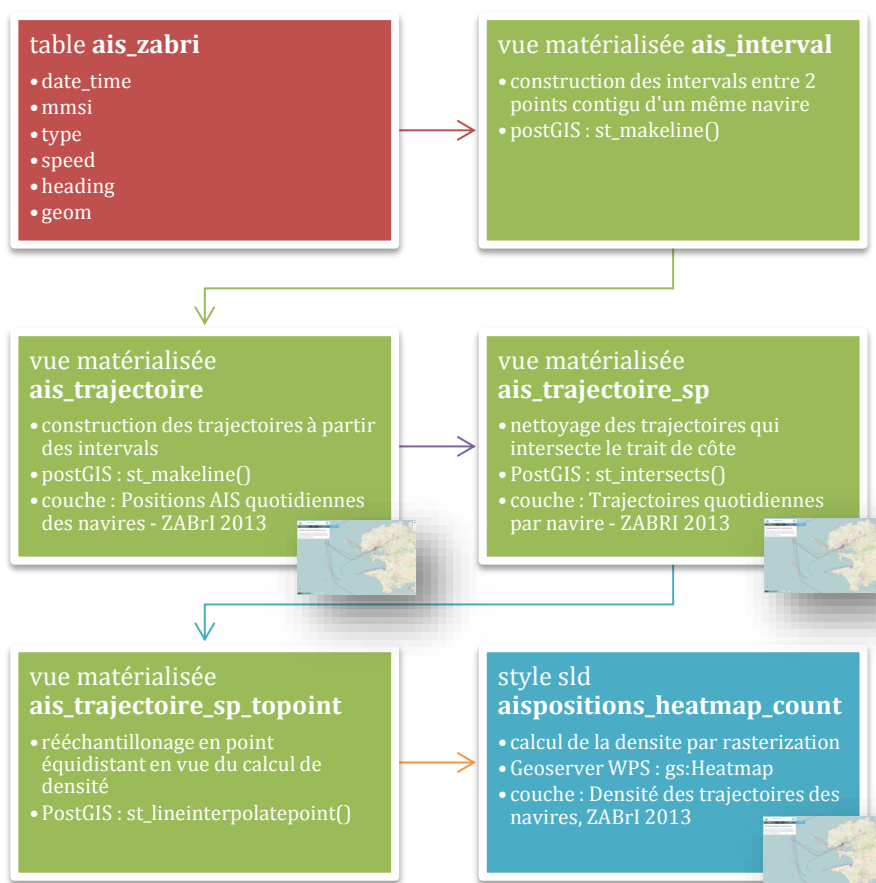


FIGURE 2. Les étapes de la chaîne automatisée d'analyse des données AIS pour la visualisation spatio-temporelle (cliquez sur les vignettes pour visualiser en ligne)

Infrastructures de Données Géographiques et observatoires de recherche en environnement

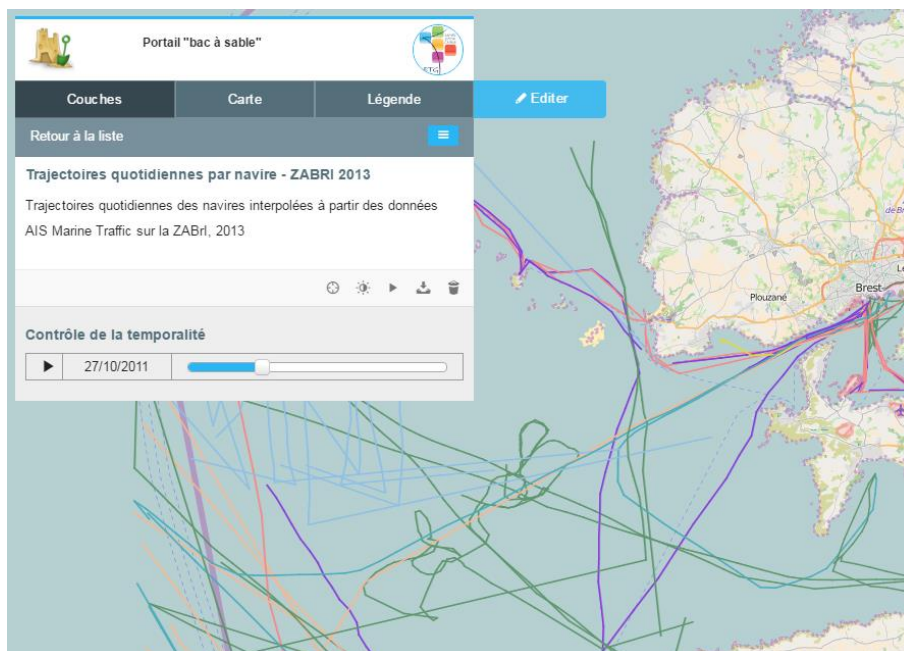


FIGURE 3. Navigation spatio-temporelle dans la couche d'information des trajectoires des navires sur le territoire de la Zone Atelier Brest-Iroise

4. Discussion et conclusion

L'IDG scientifique Indigeo est active depuis 2013, après plusieurs années de maturation. Sa première version a tiré profit de la suite logicielle geOrchestra et des fonctionnalités avancées du geoCMS. La démarche itérative adoptée a conduit à une plateforme proposant une gamme de services étendue mais dont la prise en main par les utilisateurs, en consultation et en alimentation du catalogue, s'est avérée difficile. Par exemple, les portails thématiques générés grâce au geoCMS ont complexifié non seulement la navigation dans les données mais aussi leur administration. Cette fonctionnalité a donc été abandonnée dans la version 2 d'Indigeo qui mise sur la simplicité d'exploration en adoptant une posture orientée utilisateur afin de favoriser l'évolution du consommateur d'information géographique vers un statut de « producteur » (Budhathoki *et al.*, 2008). Néanmoins, en dépit de ces évolutions techniques, le profil des contributeurs reste relativement homogène puisque l'alimentation du catalogue repose essentiellement sur un petit nombre de géographes ayant des compétences dans le domaine des SIG. Ce constat est relativement classique dans le monde académique y compris à l'international (Ellul *et al.*, 2012) dans des pays qui mènent une politique volontariste de mise en œuvre d'infrastructures au niveau des organismes scientifiques (JISC, 2011). Il est aussi équivalent dans d'autres contextes tels qu'Open Street Map par exemple (Neis and Zipf, 2012). Même pour ces contributeurs avertis, la génération de métadonnées normalisées est jugée encore

SAGEO'2016 – Nice, 6-9 décembre 2016

trop complexe et chronophage, ce qui explique en grande partie le peu d'implication de producteurs relevant d'autres disciplines non familiers du domaine. S'ajoute à cette difficulté, le statut relativement flou de la donnée scientifique dans le contexte actuel d'ouverture des contenus numériques.

La plus-value d'Indigeo peut être également considérée comme relativement limitée en termes scientifiques. Ses perspectives pourraient donc consister à la connecter à d'autres types de « mémoire » (Guarnieri *et al.*, 2003), de manière à replacer l'information géographique dans son « écosystème », c'est-à-dire dans un système d'information intégré et accessible (Faucher *et al.*, 2009 ; Loireau *et al.*, 2014) fournissant des données sur la documentation et l'organisation auxquelles les observations (l'information géographique cataloguée) se rattachent.

Dans le cas d'Indigeo, la faible adhésion des autres disciplines potentiellement impliquées dans les observatoires bénéficiaires du dispositif tient aussi probablement à sa forte identité « laboratoire ». Pour des raisons d'affichage et de valorisation, il semble évident que le développement d'IDG interopérables à différents niveaux académiques ou à un niveau fédérateur serait la solution à privilégier de manière à favoriser leur appropriation par le public concerné et lever certains blocages liés à des enjeux de pouvoir au sein de la sphère scientifique comme il en existait il y a quelques années entre collectivités territoriales au sujet des SIG (Roche, 2000). Il est en effet probable que les IDG dans leur ensemble, quel que soit leur contexte de mise en œuvre, révèlent les stratégies de leurs promoteurs pour conserver ou acquérir une place dans différentes arènes publiques. Le niveau de déploiement de ce type d'outils est donc à questionner et il n'est pas certain qu'un laboratoire de recherche soit le plus légitime et efficace à porter cette démarche.

Enfin pour comprendre le rôle des IDG académiques dans la construction de l'interdisciplinarité inhérente aux recherches en environnement et dans la circulation des connaissances, il serait intéressant de les étudier en tant qu'objet de recherche et non en tant qu'outil de recherche (Noucher, 2013) en s'intéressant notamment à leurs usages effectifs (Georis-Creuseveau *et al.*, 2015) et aux facteurs individuels qui les déclenchent (Venkatesh *et al.*, 2003).

Bibliographie

- AFIGEO, (2014). Catalogue des IDG 2014, <http://www.afigeo.asso.fr/voir-toutes-les-news/972-3-juin-2014-la-nouvelle-version-du-catalogue-des-idg.html>
- Budhathoki N., Bruce B., Nedovic-Budic Z., (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal*, 72(3), p. 149-160.
- Crompvoets J., Bregt A., Rajabifard A., Williamson I., (2004). Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science*, 18(7), p. 665-689.
- Ellul C., Winer D., Mooney J., Foord J., (2012). Bridging the gap between traditional metadata and the requirements of an academic SDI for interdisciplinary research. In Rajabifard A. &

Infrastructures de Données Géographiques et observatoires de recherche en environnement

- Coleman D. (eds), *Spatially enabling government, industry and citizens. Research and development perspectives*, GSDI Association Press, ISBN 978-0-9852444-0-8, p. 57-77.
- Faucher C., Gourmelon F., Lafaye J.Y., Rouan M., (2009). Mise en oeuvre d'une mémoire environnementale adaptée aux besoins d'un observatoire du domaine côtier : MenIr. *Revue internationale de Géomatique*, 19 (1/2009), p.7-26.
- Georis-Creuseveau J., Gourmelon F., Claramunt C., (2016). Infrastructures de données géographiques : quelle contribution à la Gestion intégrée des zones côtières françaises ? *VertigO - la revue électronique en sciences de l'environnement* [En ligne], Volume 15 Numéro 1 | mai 2015, mis en ligne le 12 mai 2015, consulté le 11 juin 2015. URL : <http://vertigo.revues.org/15857> ; DOI : 10.4000/vertigo.15857
- Gourmelon F., Georis-Creuseveau J., Rouan M., N'Diaye S., (2009). Le CIAO, première étape vers une infrastructure de données spatiales en Afrique de l'Ouest. *Revue internationale de Géomatique*, 19 (3), p. 351-372.
- Guarnieri F., Garbolino E., Houllier F., Cuq F., Lévêque C., Weill A., Matarasso P., (2003). Contribution à la définition opérationnelle et à la modélisation de la mémoire environnementale des zones ateliers. In Lévêque C., Leeuw S. (eds.), *Quelles natures voulons-nous ? Pour une approche socio-écologique du champ de l'environnement*, Paris, Elsevier, p. 296-307.
- Harvey F., Iwaniak A., Coetzee S., Cooper A., (2012). SDI past, present and future: a review and status assessment. In Rajabifard A. & Coleman D. (eds), *Spatially enabling government, industry and citizens. Research and development perspectives*, GSDI Association Press, ISBN 978-0-9852444-0-8, p. 23-38.
- JISC, (2011). JISC Geospatial working group. Terms of reference (on line) available from :<http://www.jisc.ac.uk/aboutus/howjiscworks/committees/workinggroups/geospatial>.
- Le Guyader D., Brosset D., Gourmelon F., (2012). Exploitation de données AIS pour la cartographie du transport maritime. *M@ppemonde*, 104 (2011.4), <http://mappemonde.mgm.fr/num32/articles/art11405.html>.
- Libourel T., Passouant M., Loireau M., (2009). Introduction du chapitre « Systèmes d'observation et observatoires ». *Dossier Agropolis International* 9, p. 27
- Loireau M., Fargette M., Desconnets J.C., Mougenot I., Libourel T., (2014). Observatoire scientifique en appui à la gestion du territoire (OSAGE). *Entre espaces, temps, milieux, sociétés et informatique. Actes de la conférence SAGEO 2014*.
- Masser, I., (2010). *Building European spatial data infrastructures*, ESRI Press, Redlands, California.
- Neis P., Zipf A., (2012). Analyzing the contributor activity of a volunteered geographic information project, the case of OpenStreetMap. *ISPRS Int. J. Geo-Inf.* 2012, 1, p. 146-165 ; doi :10.3390/ijgi1020146
- Noucher M., (2013). Infrastructures de données géographiques et flux d'information environnementale : de l'outil à l'objet de recherche. *Netcom, réseaux, communication et territoires*, 26/4
- Rajabifard A., Feeney M.-E., Williamson I. P., (2002). Future directions for SDI development. *International Journal of Applied Earth Observation and Geoinformation*, 4 (1), p. 11-22.
- Roche S., 2000. *Les enjeux sociaux des systèmes d'information géographique : le cas de la France et du Québec*. Paris, L'Harmattan (coll. Géographie sociale), 158 p. (ISBN 2-73840-8648-7)
- Venkatesh V., Morris M.G., Davis G.B., Davis F.D., (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27(3), p. 425-48.

Intérêt de la simulation paysagère en modélisation géoprospective

Christine Voiron-Canicio¹, Jean-Christophe Loubier², Dominique Genoud³, Daniel Hunacek³, Florian Sant¹

1. Université Côte d'Azur, CNRS, ESPACE, France
98 Boulevard Edouard Herriot, BP 3209-06204 Nice, France
christine.voiron@unice.fr ; sant.florian@gmail.com
2. Institut IEM, Université des sciences appliquées (HES-SO Vs)
Techno-pôle 3, 3960 Sierre, Suisse
Jchristophe.loubier@hevs.ch,
3. Institut IIG, Université des sciences appliquées (HES-SO Vs)
Techno-pôle 3, 3960 Sierre, Suisse
Daniel.hunacek@hevs.ch ; Dominique.genoud@hevs.ch

RÉSUMÉ. Cet article présente la démarche, et les premiers retours d'expérience, d'une recherche destinée à susciter les réflexions et les réactions de différents types d'acteurs sur les transformations futures de leur territoire de vie. La démarche se décompose en deux temps, tout d'abord, réalisation de différents scénarios de changements de l'occupation du sol, avec, associée à chacun d'eux, une représentation paysagère en 3D. Puis, travail en petit groupe dans un atelier participatif. La réflexion collective partagée porte sur les représentations paysagères en 3D des transformations de l'espace, choisies, testées, par les participants, à partir d'une application conçue à cet effet. Le territoire d'étude, situé sur le littoral des Alpes-Maritimes, est un territoire très menacé par l'urbanisation diffuse galopante.

ABSTRACT. This article presents a new approach and the initial feedback in a research that seeks to provoke reflections and reactions to different stakeholders on future changes in their living territory. This project is divided into two stages. The first is a production of different scenarios of changes in land use, associated with a 3D virtual representation of landscape. The second stage is a participatory workshop in small groups of stakeholders. The shared collective thinking is based on 3D virtual representations of transformations of space that were chosen by the participants from an application designed for this project. The study area,

Simulation paysagère et modélisation géoprospective

located on the coastal zone of Alpes-Maritimes, is very threatened by fast and diffuse urbanization.

MOTS-CLES : modèle spatial, automate cellulaire, machine learning, simulation, paysage, 3D, géoprospective, démarche participative,

KEYWORDS: spatial modelling, cellular automata, machine learning, simulation, landscape, 3D, geo-prospective, participatory approach,

1. Introduction

Si déterminer le devenir d'un territoire n'est pas chose aisée, appréhender les changements qui se produiront à travers l'espace, l'est encore moins (Voiron-Cancio, 2012). Les grandes tendances d'évolution sont assez facilement perçues, diffusion du bâti individuel, enrichissement, fragmentation des espaces naturels, par exemple. En revanche, localiser ces changements et les transformations paysagères associées, imaginer le paysage de demain produit par cette dynamique de changement, plus ou moins forte selon les lieux, est difficile sans l'assistance « d'outils » appropriés. Pour ce faire, plusieurs types de méthodes sont traditionnellement mobilisés. Les futurs d'un territoire sont déterminés en choisissant entre deux familles de méthodes, l'une quantitative « le modèle numérique » (Lajoie et Hagen-Zanker, 2007 ; Houet et Hubert-Moy, 2006 ; Voiron-Cancio, 2006), l'autre, qualitative, empirique, « la mise en récit du futur » (Sgard, 2008 ; Vanier 2010). Ces deux méthodes peuvent être couplées, par exemple, dans les démarches de modélisation d'accompagnement (Etienne, 2010 ; Gourmelon *et al.*, 2008). Il en découle des localisations plus ou moins précises et une représentation du changement qui peut prendre la forme de cartes, de dessins, de blocs diagrammes, de maquettes en dur ou numériques, d'objets 3D symbolisant des éléments spatiaux sur une scène (Michelin, 2000 ; Piveteau et Lardon, 2002 ; Chardonnel *et al.*, 2003 ; Loubier, 2006 ; 2013). Une approche participative peut être intégrée au processus de formalisation du changement spatial, la réflexion collective peut accompagner la démarche de modélisation prospective, ou bien intervenir à la phase de restitution des premiers résultats (Etienne, 2010 ; Dodane *et al.*, 2014).

La démarche proposée mobilise ces grandes familles de méthodes mais se distingue des travaux habituels par le parti retenu. Le premier choix concerne la modélisation. Les changements spatiaux futurs ont été déterminés à l'aide d'un modèle spatialement explicite, et prédits par chaînes de Markov (Takada *et al.*, 2010). Mais ce futur projeté est totalement dépendant des trajectoires antérieures, or, d'une part, le futur n'est pas totalement contenu dans le passé, d'autre part, il est à imaginer, à inventer, à tester. Aussi, à côté de la carte du scénario tendanciel, nous avons créé des scénarios alternatifs en introduisant localement des modifications spatiales « à dire d'expert ». Le second choix concerne la représentation des futurs projetés. Les représentations cartographiques en 2 D sont abstraites, peu expressives pour des personnes non familiarisées avec les codes cartographiques de l'occupation

SAGEO'2016 – Nice, 6-9 décembre 2016

du sol. Nous avons alors choisi de représenter les changements sur une scène en 3D, et d'entrer par le paysage dans la discussion sur les futurs possibles. Enfin, le troisième parti retenu est de donner la possibilité à des personnes réunies dans un atelier de proposer elles-mêmes les changements d'occupation du sol et d'explorer les scènes 3D résultant de leur choix, de naviguer dans la scène et de modifier certains de ses éléments.

Les paragraphes qui suivent exposent les grandes étapes de cette démarche : la modélisation géoprospective et les scénarios des futurs possibles ; l'application embarquée à l'usage du public, couplée aux scénarios ; et enfin, les retours d'expérience des premiers ateliers participatifs. La démarche est appliquée à une zone située à cheval sur deux communes des Alpes-Maritimes, Antibes et Vallauris, qui, du fait de leur forte attractivité, connaissent une urbanisation diffuse mal maîtrisée. Le secteur analysé couvre une zone de 3,2 km², en position rétro-littorale, où les espaces agricoles non encore urbanisés sont extrêmement convoités.

2. La modélisation géoprospective 2D et 3D

La géoprospective est une démarche heuristique d'exploration des changements spatiaux qui repose sur deux principes fondamentaux, d'une part, l'intégration de la dimension spatiale aux différents stades de la démarche, et d'autre part, le couplage de scénarios prospectifs et d'une démarche participative, dont les modalités dépendent du but assigné à la prospective (Gourmelon *et al.*, 2012 ; Voiron-Canicio, 2008 ; Houet, 2015). La modélisation retenue est une modélisation spatialement explicite dont la finalité est de proposer différentes dynamiques de l'occupation du sol de la zone d'étude.

2.1. Matrice de transition 1990-2013 et chaîne de Markov

La dynamique récente de l'occupation du sol de la zone d'étude a été déterminée par la photo-interprétation des orthophotos de l'IGN correspondant aux années 1990 et 2013. Dix natures différentes d'occupation du sol ont été relevées sur la zone d'étude, et expliquent la ventilation en dix classes retenue¹. Puis une matrice de transition a été extraite de la comparaison des 2 années par un simple calcul de fréquence de changement en divisant le total de chaque croisement par le total de chaque ligne (Tableau 1).

¹ Bâti dense ; Bâti discontinu ; Bâti diffus ; Entrepôts, espaces industriels ; Infrastructures publiques et loisirs ; Oliviers, vignes, agrumes ; Agriculture sous serre ; Forêt ; Végétation rase ; Végétation arbustive et arborée.

Simulation paysagère et modélisation géoprospective

Tableau 1. Matrice de transition pour la période 1990-2013

	Cla 1	Cla 2	Cla 3	Cla 4	Cla 5	Cla 6	Cla 7	Cla 8	Cla 9	Cla 10
Cla 1	0.48	0.00	0.00	0.21	0.00	0.22	0.00	0.00	0.02	0.07
Cla 2	0.01	0.57	0.00	0.01	0.04	0.05	0.03	0.04	0.18	0.07
Cla 3	0.01	0.26	0.36	0.00	0.01	0.15	0.01	0.01	0.18	0.02
Cla 4	0.41	0.02	0.00	0.21	0.00	0.09	0.00	0.03	0.02	0.21
Cla 5	0.02	0.36	0.00	0.01	0.13	0.12	0.03	0.01	0.26	0.04
Cla 6	0.13	0.04	0.01	0.02	0.00	0.60	0.00	0.01	0.13	0.06
Cla 7	0.01	0.27	0.00	0.00	0.03	0.02	0.28	0.01	0.34	0.03
Cla 8	0.00	0.20	0.00	0.00	0.01	0.01	0.00	0.44	0.08	0.25
Cla 9	0.03	0.41	0.00	0.01	0.04	0.08	0.04	0.04	0.26	0.08
Cla 10	0.00	0.22	0.00	0.00	0.01	0.03	0.01	0.20	0.14	0.39

Cette matrice de transition est une chaîne de Markov homogène à états finis d'ordre 1 car :

- elle décrit un processus aléatoire à états finis (les types d'occupation du sol) ;

- et que « si l'instant présent est k , la probabilité pour que le système soit dans l'état j à l'instant $k+1$ ne dépend que de l'état du système à l'instant k » (Coquillard *et al.*, 1997).

L'analyse du graphe d'état de cette matrice montre que celle-ci est ergodique car elle est irréductible et le PGDC des longueurs de circuit est 1. Elle conduit donc à un état global stationnaire. Cette propriété peut être observée expérimentalement en élevant la matrice à des puissances successives. À partir de la 6^{ème} puissance, la matrice devient pseudo-stationnaire avec une précision de 2 décimales. Le rapport de la plus haute valeur propre (3.953) avec la seconde (2.254) donne une vitesse de convergence vers la stationnarité de 1.753.

Ces informations sont intéressantes pour identifier une dynamique de l'évolution du territoire. Le caractère stationnaire montre une tendance d'ajustement optimal du territoire vers un point temporel d'équilibre global au-delà duquel tout serait cristallisé dans un état précis. Il ne serait plus possible de modifier quoi que ce soit sans risquer de perturber gravement l'équilibre du territoire. Cette matrice est donc un bon outil pour expérimenter des déviations qui seraient induites par des changements de probabilités de transition. C'est ce que nous avons fait pour construire les scénarios d'évolution possible.

Les chaînes de Markov sont par nature a-spatiales. Or, la probabilité qu'une portion de territoire soit dans un certain état est influencée par son état initial ainsi que celui des portions de territoire qui lui sont proches. Les automates cellulaires (AC) ont la capacité de permettre cela car l'une des trois propriétés des AC (outre le

SAGEO'2016 – Nice, 6-9 décembre 2016

fait de discrétiser l'espace sous la forme de mailles) est le principe de localité². C'est-à-dire que le nouvel état d'une cellule ne dépend que de son état initial et celui de ses voisins. Il devient donc intéressant de coupler les chaînes de Markov à des AC pour proposer des cartes prospectives d'occupation du sol (Caloz et Collet, 2011).

Certains Systèmes d'Information Géographique en mode raster proposent maintenant des modules de simulation Markov AC. Le plus connu est sans doute IDRISI qui est également le module que nous avons utilisé (Eastman *et al.*, 1998 ; Eastman, 2009).

2.2. Le scénario tendanciel 2013-2026

Le modèle tendanciel a été obtenu à partir des probabilités de transition tirées de la matrice de transition de l'occupation du sol, entre 1990 et 2013. Ces treize années ont été marquées par l'expansion du bâti diffus au détriment des espaces agricoles et des serres. Les changements prévus pour les treize années suivantes (2013-2026) poursuivent *de facto* la tendance à l'urbanisation diffuse observée au cours de la période précédente, et qui se fait principalement au détriment de l'agriculture et de la végétation (Fig. 1).

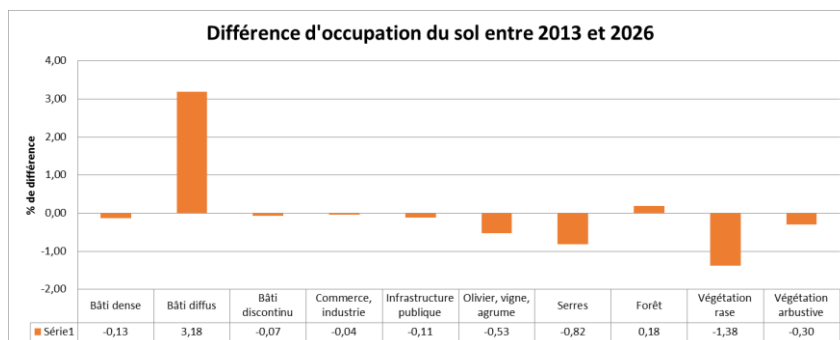


FIGURE 1. Variations des surfaces des types d'occupation du sol entre 2013 et 2026, selon le modèle tendanciel

2.3. Les scénarios alternatifs 2013-2026

L'objectif est ici de se libérer du poids du passé et d'imaginer d'autres dynamiques de changements non seulement plausibles mais aussi possibles, y

2. Les deux autres propriétés sont le parallélisme et l'homogénéité.

Simulation paysagère et modélisation géoprospective

compris des changements radicaux reposant sur des politiques d'aménagement opposées aux politiques passées et actuelles.

La démarche retenue a consisté à prendre la carte de l'occupation des sols de 2013 comme point de départ et à modifier directement les types d'occupation du sol en fonction de différents scénarios imaginés. L'objectif est de construire des images (cartes) du futur en amplifiant ou renversant des tendances, et/ou en opérant localement des transformations de l'occupation du sol par suppression ou introduction de protections en certains lieux, et création d'aménagements sur d'autres secteurs.

Une fois la carte de chaque scénario réalisée, la matrice de transition est calculée entre la carte actuelle et celle du scénario. L'objectif est d'obtenir la matrice de probabilité de transition de chaque scénario. La spatialisation s'étant faite sur le logiciel ArcGIS, la même méthode que pour le scénario tendanciel est utilisée. Ce qui revient à extraire la chaîne de Markov correspondant à chaque futur en utilisant les algorithmes d'IDRISI, en comparant les différentes cartes du futur avec celle de 2013. Ainsi, en partant d'une image du futur, on fait ressortir la chaîne qui correspond à ce futur. Cette méthode permet de « créer » aisément des futurs possibles. Pour des non spécialistes, il est plus facile de dessiner des cartes du futur que d'agir sur une matrice de transition. Par ailleurs comparer les probabilités des chaînes issues de ces différentes cartes permet également de vérifier si les propriétés markoviennes sont conservées.

Onze scénarios prospectifs ont ainsi été élaborés. Le parti retenu n'est pas de présenter un scénario souhaitable, ni même des scénarios probables, mais des scénarios correspondant à des devenirs qui pourraient être envisagés. Ces scénarios sont « libres », dans le sens où ils ne sont pas contraints par les documents d'urbanisme existants. En effet, nous considérons que ceux-ci peuvent être remis en question par une politique volontariste. Ils ne le sont pas non plus par les tendances observées par le passé. Aussi, les facteurs de changement introduits peuvent-ils être radicalement différents de la tendance dominante observée, telle une politique en faveur de l'agriculture ou bien l'interdiction du bâti diffus.

Les options pour chacun des 11 scénarios sont les suivantes :

- S1 Tendanciel A : disparition des serres au profit du bâti diffus ;
- S2 Tendanciel B : disparition de toutes les parcelles agricoles (sauf les jardins protégés) au profit du bâti diffus ;
- S3 Protection de l'agriculture et densification des espaces bâtis ;
- S4 Sauvegarde des espaces agricoles, et politique en faveur de nouveaux espaces naturels ;
- S5 Coup d'arrêt de l'urbanisation et renaissance de l'agriculture par une politique volontariste de compensation écologique ;

SAGEO'2016 – Nice, 6-9 décembre 2016

- S6 Maintien de l'agriculture mais diffusion du bâti sur les parcelles à végétation rase ;
- S7 Envahissement du territoire par du bâti diffus et discontinu, excepté dans les espaces forestiers ;
- S8 Densification du bâti, maintien de l'agriculture et compensation écologique en faveur des espaces boisés ;
- S9 Extension du bâti dense, et propagation du bâti diffus excepté dans les espaces forestiers classés ;
- S10 Extension du bâti dense, et propagation du bâti diffus excepté dans les espaces agricoles et forestiers protégés ;
- S11 Densification du bâti et politique volontariste de reprise de l'agriculture.

La série des scénarios a été construite par l'équipe de chercheurs, puis soumise à des personnes appartenant au milieu associatif, ainsi qu'à l'analyse du responsable de la planification territoriale d'Antibes, qui n'a souhaité se prononcer que sur la partie antiboise de la zone d'étude.

2.4. Construction des scènes 3D de chaque scénario

Le parti de cette recherche est d'anticiper le devenir possible de ce territoire, selon différents scénarios, en ne se limitant pas à la cartographie 2D de l'occupation du sol projetée mais en construisant les paysages virtuels associés à chacun d'eux. En effet, la cartographie des types d'occupation, même avec une nomenclature très détaillée ne peut parvenir à représenter le paysage résultant, la diversité des éléments bâtis et non bâtis, leur volume, leur agencement dans l'espace, et la composition d'ensemble. La figure 2 en est la démonstration. Le paysage de 2013 visible sur la photo est très hétérogène, voire complexe, les objets sont imbriqués.

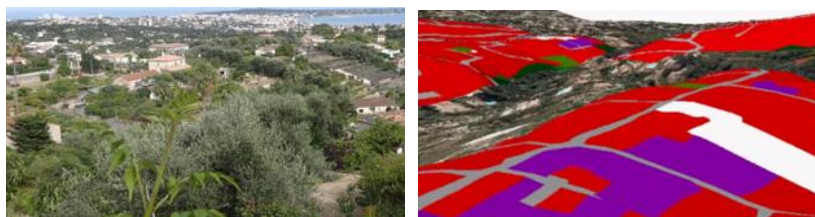


FIGURE 2. *Paysage observé et cartographie de l'occupation du sol drapée sur le MNT.*

Le niveau d'abstraction représenté par la sémiologie graphique conventionnelle, utilisant des plages de couleur homogène pour représenter le bâti diffus dans un milieu végétalisé, les serres, et les vergers, ne peut rendre compte du paysage correspondant.

Simulation paysagère et modélisation géoprospective

La modélisation prospective, telle qu'elle est envisagée, a vocation à être intégrée dans une démarche de géogouvernance³. Nous faisons l'hypothèse qu'un paysage de concertation s'avère grandement utile à tout public cherchant à imaginer la transformation de son espace de vie, et peut aider non seulement les habitants à appréhender le paysage futur de leur territoire, mais également les élus et les gestionnaires à mesurer les impacts visuels des transformations attendues. Chaque scénario cartographié a donc été couplé à une image 3D (Fig. 3). Une fois les scènes 3D construites, il devient possible de se déplacer dans chacune d'elles en temps réel.



FIGURE 3. Exemples de scènes 3D

3. « Démarche s'appuyant sur l'utilisation des méthodes et des outils de l'analyse spatiale, et destinée à mettre à la portée de tous les acteurs concernés une information territoriale pertinente, tout au long de la chaîne de construction d'un projet de territoire. Son objectif est de contribuer à rendre intelligible la complexité territoriale, à faire émerger les enjeux socio-spatiaux du territoire, ainsi que ses évolutions à différents horizons temporels » (Masson-Vincent *et al.*, 2012).

SAGEO'2016 – Nice, 6-9 décembre 2016

3. Conception d'une interface multi supports de sélection de scénarios

L'objectif du système de sélection de scénarios est triple. En premier, il s'agit de permettre à des utilisateurs d'effectuer facilement des choix quantifiés sur les types d'occupation du sol établis par photo interprétation de manière à déterminer un projet individuel de ce que l'utilisateur souhaite voir réaliser sur son territoire (par exemple, plus de bâti dense ; plus de forêts, moins de bâti individuel....) puis de transmettre le choix à un serveur qui va le considérer comme un vote. En second, les choix individuels sont fusionnés pour permettre, en dernier, de sélectionner le scénario pré-construit (parmi les 11 qui ont été élaborés) le plus proche des choix de tous les acteurs qui ont participé au vote. Le but est de visualiser le plus vite possible la maquette 3D qui correspond le mieux aux choix des acteurs de manière à faciliter le processus de concertation. En pratique, il s'agissait donc de développer :

- l'interface logicielle multi supports (ordinateurs, tablettes ; smartphones) et l'infrastructure de gestion des données sur un serveur centralisé,
- la partie logicielle de calcul/machine learning de matrice collaborative,
- l'algorithme de sélection des maquettes virtuelles.

3.1. Interface multi supports

Du point de vue de l'application logicielle exécutée sur les appareils mobiles (Heitkötter *et al.*, 2012), le système se décompose en deux parties : une partie administrateur et une partie utilisateur. L'administrateur peut créer et gérer les différentes sessions ainsi que les droits des utilisateurs, par exemple, autoriser un individu à participer à une session. La partie utilisateur se compose d'un système d'inscription et d'un menu de choix de sessions. Quand il choisit une session, l'utilisateur voit apparaître la fenêtre représentée sur la figure 4.

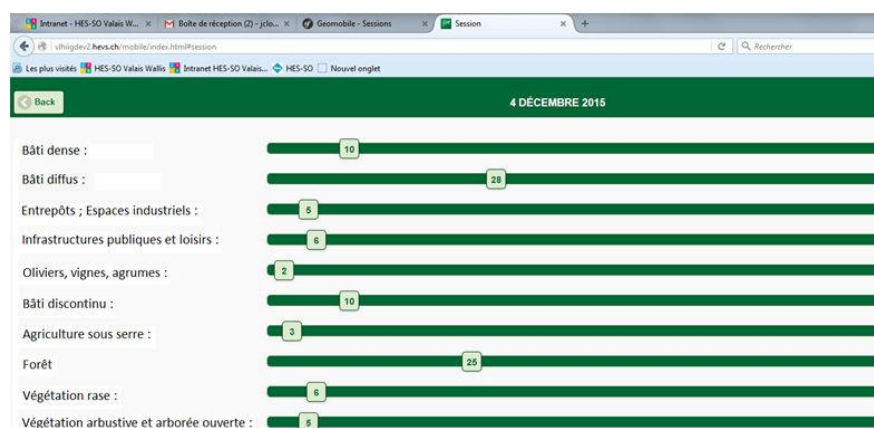


FIGURE 4. L'interface de quantification des types d'occupation du sol. Les curseurs peuvent être déplacés horizontalement.

Simulation paysagère et modélisation géoprospective

Le mécanisme est simple. Plus l'utilisateur déplace un curseur vers la droite, plus il souhaite que ce type d'occupation du sol soit valorisé dans l'aménagement du territoire. Le % d'influence est écrit dans chaque curseur. Bien entendu, comme le territoire est borné, les autres curseurs sont influencés en sens inverse de façon à ce que leur somme soit égale à 100%. La valeur inscrite dans le bouton de chaque curseur correspond à la quantité d'occupation du sol (en %) liée à la position du curseur. Dans la partie basse de l'interface, un graphique représente les variations des pourcentages d'occupation du sol choisis par l'utilisateur. Les droites qui vont de T0 (2013) à T1 (2026), représentent la tendance temporelle attribuée à chaque classe (Fig. 5).

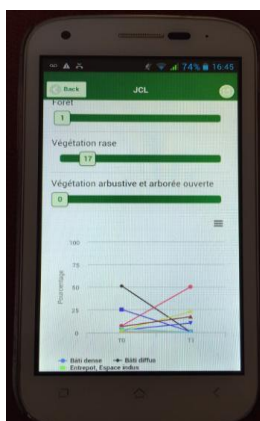


FIGURE 5 : Capture d'écran du graphe situé dans la partie basse de l'interface-utilisateur.

Quand l'utilisateur a validé ses choix, ils sont transmis à l'application de calcul, sur un serveur, où les algorithmes de fusion des choix et de sélection des scénarios seront exécutés.

3.2. Algorithme de calcul du meilleur scénario

Le système embarqué crée une position moyenne - celle des votants - à partir d'un vecteur d'états - les types d'occupations du sol - dans un espace multidimensionnel constitué par la matrice de transition de chacun des scénarios. Il mesure ensuite la distance qui sépare cette position de celle de chacun des scénarios. De multiples possibilités de mesurer ces distances s'offrent à nous comme par exemple celle proposée par Boutsidis *et al.* (2009).

SAGEO'2016 – Nice, 6-9 décembre 2016

L'algorithme que nous avons développé pour cela est présenté ci-dessous. Il comporte 7 étapes.

1. Soit $M1 \dots n$ les matrices du catalogue de scénarios d'aménagement
2. $T0$: le vecteur initial: transmis aux mobiles (situation de départ)
3. $T0 * M1 \dots n = TM1 \dots 0$: calcul des vecteurs de relation aux matrices selon le vecteur initial
4. $T0$ modifié sur chaque mobile devient $T0mob1 \dots k$
5. Calcul du vecteur résultant $Tres =$ moyenne ($T0mob1 \dots k$)
6. Donc on calcule $Tres * Mi = Tmires$ puis distance euclidienne ($Tmires, Tmi$) = d_i

[Preuve : $Tres = T0$ et $m1$ matrice de départ donc $Tres * m1 = Tm1res$ donc $Tm1res - Tm1 = 0$]

7. On sélectionne le plus petit $d_i \implies i$ l'indice de la matrice la plus proche m_i .

La stabilité des calculs matriciels est assurée par la plateforme de calcul utilisée (Berthold *et al.*, 2007).

Au final, l'administrateur de la session reçoit dans l'interface de gestion, un message lui indiquant quel scénario est choisi. Il ne lui reste plus qu'à l'activer dans l'environnement de visualisation 3D.

De multiples tests ont été réalisés et ils montrent que le système est robuste et qu'il renvoie une solution dans un intervalle de temps compris entre 30 secondes et 2 minutes.

4. Retour d'expérience des premiers ateliers participatifs

Deux ateliers participatifs ont été organisés à Antibes et à Vallauris, avec des ingénieurs de la ville, et avec des membres d'une association de défense de l'environnement et une élue en charge de l'environnement.

Les scènes étaient projetées sur un grand écran. L'animateur de l'atelier opérait les déplacements sur la scène, selon les indications des participants. Pour faciliter la comparaison entre le paysage actuel et le paysage du scénario retenu, un assistant effectuait le même déplacement sur la scène de 2013. Les deux paysages 3D étaient visualisés côte à côte. Les simulations 3D ont été bien reconnues par les acteurs et les navigations se sont effectuées sans problème. Ils reconnaissaient leur territoire et commentaient les maquettes (reconnaissance de certains lieux, discussion et justification sur le futur développement urbain des zones concernées). Cependant, les participants n'ont pas cherché à modifier les maquettes et se sont satisfaits des visualisations de leurs votes. Ceci pourrait être dû au degré de vérisme des scènes. Lors de précédents travaux, il avait été remarqué que le niveau de vérisme des scènes 3D jouait un rôle dans l'interaction proposée aux acteurs. Cela avait conduit

Simulation paysagère et modélisation géoprospective

l'un des auteurs de cet article à mener une réflexion sur la notion de perception spatiale dans un environnement 3D dans un contexte de prise de décision (Loubier, 2004). Dans le cadre de cette étude, les scènes proposées étaient assez réalistes. Seuls, les bâtiments restaient relativement symboliques. Comme l'essentiel de l'interaction avec les scènes se faisait par des survols, il est probable que le message délivré par ces scènes ait été suffisant pour que les acteurs puissent construire un point de vue sans entrer dans un processus de négociation, d'autant que les participants présents à chaque séance partageaient un corpus idéologique identique.

Par ailleurs, deux faits sont apparus. Le premier concerne la grande difficulté à appréhender les occupations du sol de type mixte habitat/végétation, caractérisant par exemple le bâti diffus, à partir des cartes résultant des scénarios. En effet, en complément de la projection à l'écran du paysage simulé, la carte correspondant au scénario a été distribuée. Or, cette dernière a davantage dérouté qu'aidé à la compréhension des transformations de l'occupation du sol. Les plages de couleur uniforme, et rouge pour le bâti, ne permettant pas de représenter l'imbrication, à des degrés divers, de la végétation et du bâti. L'avantage de la représentation 3D des scénarios par rapport aux représentations 2D est apparu de façon flagrante.

Le second fait concerne la mise en évidence de certains effets de bord dans la combinaison des critères, notamment dans la perception logique des différentes combinaisons. Les utilisateurs ont voté à un moment donné en privilégiant le bâti dense pour dégager selon leur dire, de l'espace destiné à la forêt. Ce lien qualitatif d'interaction entre les critères de façon à dégager de la place pour un autre critère ne peut pas être pris en compte dans l'algorithme que nous avons développé. Comme il s'agit d'une sélection construite sur l'évaluation d'une distance dans un espace de dimension 10, les critères interviennent comme des mesures quantitatives pures et indépendantes entre elles d'un point de vue mathématique. Cela a engendré la sélection d'un scénario urbain plutôt qu'un scénario forestier, ce qui a surpris les participants. Cette observation nous a amenés à envisager d'intercaler un système expert entre le système de calcul des positions des acteurs et les positions des scénarios dans l'espace. Il est difficile de qualifier mathématiquement ou par des méthodes de traitement systématique un raisonnement logique et qualitatif portant sur l'interaction entre les types d'occupation du sol, et surtout sur le fait que l'utilisateur suppose que la place dégagée pour un critère est destinée à un autre critère bien précis (comme le bâti dense réduit au profit de la forêt). En effet, en supposant que les curseurs se déplacent par cran de 1% comme cela se passe dans notre fenêtre de vote, les combinaisons possibles à traiter par des règles de sélection s'élèvent à $2.6341E+23$ ce qui est impossible à programmer par des méthodes classiques. Il faut donc développer une stratégie de raisonnement et utiliser un langage de programmation logique comme Prolog qui nous permettra à partir d'une base de faits (les positions de différents curseurs des critères) et d'une base de règles (les façons d'utiliser les faits) de reproduire le raisonnement logique d'un humain qui manipule les critères (Blackburn *et al.*, 2007). Cette modification est en cours et ne sera pas présentée dans cet article.

À l'issue de l'atelier, une discussion s'est engagée autour de la pertinence de l'outil. Les deux publics ont réagi différemment vis-à-vis de la démarche.

SAGEO'2016 – Nice, 6-9 décembre 2016

Globalement, les gestionnaires ont eu beaucoup de mal à faire abstraction du droit du sol existant. La démarche consistant à créer librement un paysage sans contraintes juridiques a gêné certains d'entre eux, alors que d'autres se sont plus vite adaptés à la situation. Aux yeux de certains, l'outil représente un danger pour les pouvoirs publics (augmentation du nombre de recours possibles sur des projets). Ils doutent qu'il puisse aider la concertation, compte tenu des égoïsmes locaux. Parallèlement, ils ont souligné le fait que cet outil pourrait aider le maire à comprendre les enjeux du territoire pour la construction du PLU, avec, toutefois, quelques hésitations sur les personnes-cibles. Ainsi, un ingénieur a ajouté que cet outil ne devait pas sortir du bureau du maire, alors qu'un autre l'a perçu comme un instrument intéressant destiné à l'éducation territoriale, et qui devrait être diffusé dans les écoles pour modifier les mentalités.

Les membres de la société civile se sont montrés plus ouverts à l'égard de la démarche, appréciant le fait qu'elle permette de comprendre les enjeux territoriaux et paysagers sans avoir de connaissances scientifiques ou techniques particulières. Ils ont déclaré que l'outil devrait être employé dans toutes les séances portant sur les aménagements envisagés du territoire communal ; qu'il devait être proposé comme service sur le site web de la commune, et enrichi de fonctions plus fines comme des simulations d'inondations. À l'inverse des ingénieurs territoriaux, ils battent en brèche l'idée d'un égoïsme local et pensent que cet outil pourrait renforcer la notion d'identité de la commune.

D'une façon générale, le recours aux scènes 3D a aidé les acteurs à mesurer les enjeux spatiaux du territoire étudié. On peut donc estimer que ce type de représentation apporte un avantage au niveau de l'aide à la compréhension des phénomènes territoriaux, comme l'ont déjà souligné d'autres auteurs (Jacquinod et Langumier, 2010). Toutefois, le débat reste ouvert autour de l'apport effectif de ces nouveaux moyens de représenter l'espace qui utilisent massivement la puissance informatique dans le but de produire des expériences les plus immersives possible. Il semble que l'aide apportée par la représentation 3D ne soit pas équivalente, dans l'approche expérientielle, à celle déployée dans les jeux vidéo, notamment quand il s'agit de construire interactivement des paysages négociés entre acteurs aux opinions différentes. Dans ce cas, une maquette respectant les volumes mais pas les textures semble plus facile à modifier qu'une maquette avec un haut niveau de réalisme. Des expérimentations seront effectuées dans de prochains travaux pour tenter d'en mesurer les contours et valider cette hypothèse.

5. Conclusion

La démarche présentée appartient à la famille des méthodes de géoprospective et de géogouvernance. En effet, d'une part, elle est basée sur une modélisation spatialement explicite dont la finalité n'est pas de prédire le futur mais d'anticiper le devenir du territoire, d'autre part, la modélisation est considérée comme un médium facilitant la réflexion et la construction collectives, dans le cadre d'ateliers

Simulation paysagère et modélisation géoprospective

participatifs. Cependant, cette démarche se distingue des modélisations prospectives habituelles. Son originalité réside dans trois points. Tout d'abord, elle s'inscrit volontairement dans une pratique de modélisation prospective « libérée » pour reprendre l'expression d'Arnaud Banos (2013), qui se situe entre la modélisation exploratoire et la modélisation normative, entre la perception qualitative et l'appréhension quantitative du changement spatial futur. Elle prend ainsi ses distances vis-à-vis des canons de la modélisation prospective. Les scénarios spatiaux contrastés ne cherchent pas à être probables mais simplement possibles. L'élaboration des scénarios n'est ni contrainte par le passé, ni dépendante des règlements d'urbanisme, néanmoins, elle respecte les principes de vraisemblance, de pertinence et de cohérence. Par ailleurs, la cartographie de changement n'est pas le résultat de la modélisation mais le point de départ. À l'instar du récit par lequel est « apprivoisé le futur » (Sgard, 2008), le tableau du cadre de vie de demain est brossé en deux temps, cartographie de l'occupation du sol, tout d'abord, puis, représentation 3D du paysage fictif qui lui serait associé.

Dans ce contexte, la maquette 3D remplit un rôle important et n'est pas réduite à un simple volet spectaculaire, moderne et technologique. Par sa capacité holophrastique et visuelle, elle s'apparente à une machine à présenter un « énoncé occasionnel sur lequel les membres... [d'une]... communauté peuvent s'accorder sur le champ au vu de la situation » (Van Orman Quine, 1993, 27).

Grace à cela, le paysage du futur n'est pas l'objet de la modélisation mais l'outil de mise en débat du territoire de demain. Le paysage virtuel couplé au scénario a pour intérêt majeur de permettre l'immersion des participants dans le territoire imaginaire de demain. Le protocole adopté couplant la modélisation 3D et l'interface de choix de scénarios est un compromis entre l'interactivité recherchée et la faisabilité de l'exercice en temps réel. Ce protocole est encore au stade expérimental mais les enseignements tirés des tests préliminaires menés lors des deux premiers ateliers, et l'intérêt des participants nous confortent dans la poursuite de cette démarche prospective par la simulation paysagère.

Bibliographie

- Banos A. (2013). *Pour des pratiques de modélisation et de simulation libérées en SHS*. HDR, Université de Paris I.
- Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Wiswedel, B. (2009). KNIME-the Konstanz information miner: version 2.0 and beyond. *AcM SIGKDD explorations Newsletter*, 11(1), 26-31.
- Blackburn, P., Bos, J., Striegnitz, K. (2007). *Prolog, tout de suite!*. College publications.
- Boutsidis C., Mahoney M., Dieneas P. (2009). An improved approximation algorithm for the column subset selection problem. *SODA '09 Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, p. 968-977.
- Caloz R., Collet C. (2011). *Analyse spatiale de l'information géographique*. PPUR Presses polytechniques.

SAGEO'2016 – Nice, 6-9 décembre 2016

Chardonnel, S., Feyt G., Loubier J. C. (2003). La maquette virtuelle comme fond de carte: une vision commune du territoire. *Les figures du projet territorial*, 157-170.

Coquillard P., Hill D. R., Sime-Ngando T. (1997). *Modélisation et simulation d'écosystèmes: des modèles déterministes aux simulations à événements discrets*. Paris, Masson.

Dodane C., Joliveau T., Rivière-Honegger A. (2014). Simuler les évolutions de l'utilisation du sol pour anticiper le futur d'un territoire. Analyse critique d'une expérience de géoprospective dans un bassin versant périurbain de l'agglomération lyonnaise, *Cybergeo*, <http://cybergeo.revues.org/26483>.

Eastman J. R., Jiang H., Toledano J. (1998). Multi-criteria and multi-objective decision making for land allocation using GIS. *Multicriteria analysis for land-use management*. Springer Netherlands, p. 227-251.

Eastman J. R. (2009). *IDRISI Taiga guide to GIS and image processing*. Clark Labs, Clark University. Worcester.

Etienne M. (2010). La modélisation d'accompagnement : une démarche adaptative ? *La Modélisation d'accompagnement. Une démarche en appui au développement durable*. Versailles, Quae coll. « Update Sciences & Technologie », p. 295-303.

Gourmelon F., Houet T., Voiron-Canicio C., Joliveau T. (2012). La géoprospective, apport des approches spatiales à la prospective. *L'Espace géographique*, n° 2/2012, p. 97-98.

Gourmelon F., Étienne M., Rouan M., Kerbiriou C., Charles M., Bioret F., Chlous-Ducharme F., Guermeur Y., Levrel H. (2008). Éléments de prospective environnementale dans une réserve de biosphère. *Cybergeo*, article 429. <http://www.cybergeo.eu/index20343.html>.

Heitkötter H., Hanschke S., Majchrzak T. A. (2012). Evaluating cross-platform development approaches for mobile applications. *Web information systems and technologies*, Springer Berlin Heidelberg, p.120-138.

Houet T. (2015). Usages des modèles spatiaux pour la prospective. *Revue Internationale de géomatique*, n° 1/2015, p. 123-143.

Houet T., Hubert-Moy L. (2006). Modelling and projecting land-use and land-cover changes with a cellular automaton considering landscape trajectories: an improvement for simulation of plausible future states. *EARSel eProceedings*, n° 5, p. 63-76

Jacquino F., Langumier J. (2010). Géovisualisations 3D en action dans l'aménagement du territoire. *Géocarrefour*, Vol. 85/4 2010, p. 303-311.

Lajoie G., Hagen-Zanker A. (2007). La simulation de l'étalement urbain à La Réunion : apport de l'automate cellulaire Metronamica® pour la prospective territoriale. *Cybergeo : European Journal of Geography*, <http://cybergeo.revues.org/11882>

Loubier J-C. (2013). L'intérêt des maquettes virtuelles dans la compréhension des enjeux spatiaux. *Géogouvernance: Utilité sociale de l'analyse spatiale*. Versailles, Quae, p. 119-127.

Loubier J-C. (2006). Construction interactive de maquettes numériques du territoire. Un moyen d'aider la concertation en aménagement. *Revue Internationale de Géomatique*, 16(1), p. 93-113

Loubier, J. C. (2004). *Perception et simulation des effets du changement climatique sur l'économie du ski et la biodiversité (Savoie et Haute-Savoie)* (Doctoral dissertation).

Simulation paysagère et modélisation géoprospective

- Masson-Vincent M., Dubus-Viossat N., Bley D., Voiron-Canicio C., Helle C., *et al.*, (2012). La Géogouvernance : un concept novateur ? *Cybergeo*, article 587, <http://cybergeo.revues.org/25086>
- Michelin Y. (2000). Le bloc-diagramme : une clé de compréhension des représentations du paysage chez les agriculteurs ? Mise au point d'une méthode d'enquête préalable à une gestion concertée du paysage en Artense (Massif central français). *Cybergeo*, <http://cybergeo.revues.org/1992>
- Piveteau V., Lardon S. (2002). Chorèmes et diagnostics de territoire : une expérience de formation. *Mappemonde*, n° 4, p. 1-6.
- Sgard A. (2008). Entre rétrospective et prospective, *EspacesTemps.net*, <http://www.espacestemp.net/articles/entre-retrospective-et-prospective/>
- Sgard A. (2010). Le paysage dans l'action publique : du patrimoine au bien commun. Dossier Paysage et développement territorial, *Revue Développement durable & Territoires*, <https://developpementdurable.revues.org/8565>
- Takada, T., Miyamoto, A., and Hasegawa, S.F., (2010) Derivation of a yearly transition probability matrix for land-use dynamics and its applications, *Landscape Ecology*, 25, p. 561–572
- Vanier M. (2010). Le choix d'une prospective in Délégation à l'aménagement du territoire et à l'Action Régionale, *Territoires 2040*, Paris, La documentation française, p. 27-31.
- Van Orman Quine W., Clavelin M. (1993). *La poursuite de la vérité*. Ed. du Seuil.
- Voiron-Canicio C. (2012). L'anticipation du changement en prospective et des changements spatiaux en géoprospective », *L'Espace géographique*, n° 2012/2, p. 99-110.
- Voiron-Canicio C. (2008). A spatio-morphological modelling for spread predicting, *Computational Science and Its Applications, ICCSA 2008, International Conference, Proceedings, Part I, LNCS, Subseries: Theoretical Computer Science and General Issues*, Vol. 5072, p. 210-220.
- Voiron-Canicio C. (2006). *L'espace dans la modélisation des interactions nature-société*, https://www.researchgate.net/publication/267631509_L%27espace_dans_la_modelisation_de_s_interactions_nature-societe.

Un modèle et une plateforme pour l'évaluation et la comparaison de services de reconnaissances d'entités nommées spatiales

Christophe Marquesuzaà , Patrick Etcheverry ,
Christian Sallaberry

Laboratoire LIUPPA - Equipe T2I, EA3000
Université de Pau et des Pays de l'Adour, France

christophe.marquesuzaa@iutbayonne.univ-pau.fr,patrick.etcheverry@iutbayonne.univ-pau.fr,christian.sallaberry@univ-pau.fr

RÉSUMÉ. Cet article s'intéresse à l'évaluation des systèmes de reconnaissance d'entités nommées spatiales. Nous présentons un état de l'art des principaux systèmes connus et nous mettons en évidence la difficulté et le coût pour évaluer et comparer la qualité des résultats produits par ces systèmes. Pour répondre à ce problème nous nous appuyons sur le cadre d'évaluation SNERBM et nous proposons un modèle dédié à l'évaluation des entités nommées. Nous présentons ensuite une plateforme intégrant ce modèle et permettant de mettre en oeuvre un processus d'évaluation conforme au cadre SNERBM. Nous concluons cet article en présentant les premiers résultats d'expérimentation et nous ouvrons quelques perspectives pour faire évoluer le processus d'évaluation et la plateforme support.

ABSTRACT. This paper focuses on the evaluation of spatial named entities recognition systems (SNERS). We present an overview of the main known systems and we highlight the difficulty and the cost to evaluate and to compare the quality of the results produced by these systems. To overcome this problem, we use an evaluation framework named SNERBM and we propose a model dedicated to the evaluation of the named entities recognized by these different SNERS. Next, we present a platform integrating this model and carrying out an evaluation process according to the SNERBM framework. We conclude this paper by presenting the first experimentation results and we open perspectives for both the evaluation process and the underlying platform.

MOTS-CLÉS : Entités nommées spatiales, cadre d'évaluation, modèle de représentation d'entités nommées, plateforme d'évaluation.

KEYWORDS: Spatial named entities, evaluation framework, named entities representation model, evaluation platform

1. Introduction

La reconnaissance d'entités nommées (REN) dans des textes (Grishman, Sundheim, 1996 ; Chinchor, 1998), consiste à identifier des syntagmes appelés entités nommées (noms propres, expressions de temps et expressions numériques, par exemple) catégorisables dans des classes telles que noms de personnes, noms d'organisations ou d'entreprises, noms de lieux, dates, quantités, distances, valeurs, acronymes, abréviations, etc. Parmi les entités nommées, les noms de lieux, que nous appellerons entités nommées spatiales (ENS), désignent des objets géographiques tels que des entités administratives (commune, par exemple), des éléments du relief, des éléments hydrographiques, etc.

Des systèmes de reconnaissance d'entités nommées spatiales (SRENS) sont très largement utilisés, mais souvent sans en connaître les forces et faiblesses. C'est pourquoi nous utilisons le cadre d'évaluation SNERBM (*Spatial Name Entity Recognition BenchMark*) comme référentiel commun (Palacio *et al.*, 2016). Ce cadre permet l'évaluation de l'efficacité (*effectiveness*) de SRENS. Il permet également d'envisager le choix d'un système, ou encore la combinaison de différents systèmes, particulièrement adaptés aux catégories d'entités nommées spatiales (ville, barrage, montagne, par exemple) majoritairement présentes dans un corpus donné.

Dans cet article, nous proposons un modèle dédié à l'évaluation d'entités nommées retournées par des SRENS. Nous proposons également une plateforme s'appuyant sur ce modèle et permettant l'évaluation de SRENS de manière semi-automatique. L'objectif à terme est de pouvoir invoquer différents SRENS pour analyser un même texte et obtenir automatiquement une évaluation de la qualité des résultats retournés.

Pour faciliter cette évaluation, nous proposons un processus semi-automatique qui permet, à partir d'un fichier résultat de référence construit par des experts, de mesurer automatiquement les taux de rappel, de précision et de F1-Mesure (Manning *et al.*, 2008) des entités reconnues par divers SRENS. Au delà d'une évaluation chiffrée, le modèle et la plateforme proposée doivent également permettre aux experts de visualiser graphiquement les résultats obtenus afin de faciliter leur interprétation.

Cet article est organisé comme suit. La section 2 recense différents SREN, le type d'entités qu'ils sont capables d'annoter, les campagnes d'évaluation associées et le coût de mise en place de ces campagnes. Après avoir rappelé les principes de l'évaluation SNERBM en section 3, nous proposons en section 4 un modèle étendu d'entités nommées dédié à la visualisation d'annotations et à l'évaluation automatique de SREN spatiales et de SREN temporelles. Dans la section 5, nous proposons ensuite une plateforme intégrant le modèle précédent et permettant d'évaluer de manière semi-automatique des SRENS. La section 6 présente les premières expérimentations menées sur la plateforme. Enfin, nous concluons par un bilan et des perspectives sur nos prochains travaux.

SAGEO'2016 – Nice, 6-9 décembre 2016

2. Etat de l'art

Un grand nombre de SREN ont été proposés dans la littérature lors des vingt dernières années (Lieberman *et al.*, 2010 ; Lingad *et al.*, 2013). Le Tableau 1 propose une liste non exhaustive de ces systèmes.

Tableau 1. Liste non exhaustive de SREN et de campagnes d'évaluation

Système de REN	Annotation			URL	Campagnes d'évaluation
	Spat.	Temp.	Thém.		
Abner			✓	pages.cs.wisc.edu/~bsettles/abner/	(Marrero <i>et al.</i> , 2009)
Alchemy	✓		✓	www.alchemyapi.com/	(Rizzo <i>et al.</i> , 2012)
Annie	✓	✓	✓	gate.ac.uk/sale/tao/splitch6.html#x9-1300006	(Marrero <i>et al.</i> , 2009)
CiceroLite	✓	✓	✓	www.languagecomputer.com/products/text-annotation/cicerolite.htm	
Clavin	✓			clavin.bericotechnologies.com/	(D'Ignazio, 2013)
Dbpedia Spotlight	✓		✓	github.com/dbpedia-spotlight/dbpedia-spotlight/wiki/Web-service	(Rizzo <i>et al.</i> , 2012)
FreeLing	✓		✓	nlp.lsi.upc.edu/freeling/node/1	(Marrero <i>et al.</i> , 2009)
FOX	✓		✓	139.18.2.164:4444/demo/index.html#!/home	
Geolocator	✓			github.com/geoparser/geolocator-3.0	(Gelernter, Zhang, 2013)
Gisgraphy	✓			www.gisgraphy.com/documentation/user-guide.htm#geocodingwebservice	
Heidelttime		✓		code.google.com/p/heidelttime/	
LingPipe	✓		✓	alias-i.com/lingpipe/	(Marrero <i>et al.</i> , 2009)
Metacarta	✓			www.metacarta.com/	
Mondeca	✓	✓	✓	www.mondeca.com/	
NERD	✓	✓	✓	nerd.eurecom.fr/documentation	
OpenCalais	✓		✓	www.opencalais.com	(D'Ignazio, 2013)
Poolparty	✓		✓	www.poolparty.biz/	
Saplo	✓		✓	saplo.com/	(Rizzo <i>et al.</i> , 2012)
Supersense	✓	✓	✓	sourceforge.net/projects/supersensetag/	(Marrero <i>et al.</i> , 2009)
TextPro	✓		✓	textpro.fbk.eu/	(Marrero <i>et al.</i> , 2009)
TextRazor	✓		✓	www.textrazor.com/named_entity_recognition	
Time Bliography	✓	✓		spaceandtime.wsiabato.info/tGIS.html	
Unlock	✓		✓	edina.ac.uk/unlock/texts/getstarted.html	(Tobin <i>et al.</i> , 2010)
Yago Naga Aida	✓		✓	www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/aida/webservice/	
Yahoo! Content Analysis	✓		✓	developer.yahoo.com/search/content/V2/contentAnalysis.html	(Rizzo <i>et al.</i> , 2012)
Yahoo! PlaceSpotter	✓			developer.yahoo.com/boss/geo/docs/key-concepts.html	(D'Ignazio, 2013)
YooName	✓		✓	yooname.wordpress.com/	(Marrero <i>et al.</i> , 2009)
YQL	✓			developer.yahoo.com/yql/guide/	
Zemanta	✓	✓	✓	www.zemanta.com/	(Rizzo <i>et al.</i> , 2012)

Aussi, de nombreuses campagnes d'évaluation proposent des plateformes pour évaluer des systèmes de REN ou de recherche d'information géographique (RIG) (Leidner, 2007 ; Andogah, 2010 ; Nouvel, 2012) :

- MUC : Message Understanding Conference (Chinchor, 1998),
- MET : Multilingual Extraction Task (Chinchor, 1998),

Un cadre pour l'évaluation du SRENS

- IREX : Information Retrieval and Extraction eXercise (Sekine, Eriguchi, 2000),
- CoNLL : Computational Natural Language Learning conference (Tjong Kim Sang, De Meulder, 2003),
- HAREM : Avaliação de sistemas de Reconhecimento de Entidades Mencionadas (Santos *et al.*, 2006),
- GeoClef : Geographic Cross Language Evaluation Forum (Mandl *et al.*, 2009),
- ACE : Automatic Content Extraction program (Strassel *et al.*, 2008),
- EVALITA : Evaluation of NLP and Speech Tools for Italian (Lenzi *et al.*, 2013),
- TREC-CS : Text Retrieval Conference – Contextual Suggestion (Voorhees, 2001 ; Dean-Hall *et al.*, 2013).

Cependant, les cadres d'évaluation mis en œuvre reposent sur des corpus et des métriques différents et, en définitive, peu de SREN du Tableau 1 y ont participé. Par conséquent, il est impossible à ce jour de connaître les performances relatives de SREN suivant un même référentiel.

D'autres approches, dédiées à l'évaluation d'un SREN particulier ou d'un groupe de SREN de même domaine, utilisent des corpus plus petits dont les ressources annotées ne sont que rarement mises à disposition. Bucher *et al.* (2005) ont ainsi proposé d'adapter des techniques d'évaluation existantes pour évaluer le système de RIG SPIRIT. Marrero *et al.* (2009) présentent une plateforme qui a permis l'évaluation des systèmes de REN Supersense, Abner, Annie, FreeLing, TextPro, YooName, ClearForest et Lingpipe. Tobin *et al.* (2010) décrivent une approche dédiée à l'évaluation de système de RENS et plus particulièrement des modules de désambiguïsation et de géolocalisation des systèmes Unlock et Yahoo!PlaceMaker. Anastácio *et al.* (2010) ciblent également l'évaluation de systèmes de RENS. Les auteurs comparent les méthodes de calcul de portée spatiale supportées respectivement par les systèmes Yahoo!PlaceMaker, GIPSY, Web-a-Where et GraphRank. D'Ignazio (2013) compare les services d'analyse spatiale des systèmes de RENS OpenCalais, Clavin et Yahoo!PlaceSpotter. Enfin, Gelernter et Zhang (2013) évaluent le système de RENS Geolocator. Les auteurs mesurent la qualité du module de reconnaissance de toponymes de Geolocator ainsi que de son module d'analyse et de géolocalisation.

En ce qui concerne les systèmes de RENS, aucun *benchmark* existant ne s'est imposé comme référentiel commun. Le principal frein à l'adoption d'un *benchmark* est certainement le verrouillage (la non diffusion) des corpus et du code du logiciel d'évaluation. Nous avons pu nous procurer et travailler sur les *benchmarks* TREC-CS (Dean-Hall *et al.*, 2013) et GeoparsingQT (Gelernter, Zhang, 2013). Le premier ne propose que des ENS de type nom de grandes villes d'Amérique du nord. Le second est plus intéressant car il traite de quinze catégories différentes d'ENS. Ainsi, un jeu d'ENS hétérogènes, constitué par

SAGEO'2016 – Nice, 6-9 décembre 2016

des géographes, permet de mesurer les éventuels effets de bord engendrés par chaque passage à une version supérieure du système de RENS Geolocator.

Au même titre que l'évaluation de la qualité de l'annotation des EN, il est important de proposer des scénarios de visualisation des entités ainsi annotées (Luong *et al.*, 2015). La visualisation des ENS se fait dans le texte mais ces dernières peuvent également être projetées sur un fonds de carte. Il en va de même pour les entités temporelles que l'on peut représenter sur une frise du temps. Les autres entités que nous qualifions de thématiques (personnes, organisation, par exemple) peuvent être représentées sous forme de graphes ou d'arbres.

Il n'existe, à notre connaissance, qu'une plateforme intégrant des SREN. Il s'agit de la plateforme NERD¹ (*Named Entity Recognition and Disambiguation*) qui intègre les SREN Alchemy, Dandelion, DBPedia Spotlight, Lupe-dia, OpenCalais, Saplo, SemiTags, TextRazor, THD, Wikimeta, Yahoo!Content Analysis et Zemanta (les plus utilisés figurent dans le Tableau 1). Chacun de ces systèmes peut être invoqué via l'API de la plateforme. Les résultats sont retournés au format JSON (Rizzo *et al.*, 2012) comme illustré ci-dessous après un appel à la version 0.5 de l'API NERD :

Listing 1 – Exemple d'entité reconnue par NERD au format JSON

```

1  [{
2      "idEntity": 35312,
3      "label": "Tunisia",
4      "extractorType": "Country",
5      "nerdType": "http://nerd.eurecom.fr/ontology#Location",
6      "uri": "http://www.cometotunisia.co.uk",
7      "confidence": 0.317654,
8      "relevance": 0.5,
9      "extractor": "alchemyapi",
10     "startChar": 1445,
11     "endChar": 1452
12  }]

```

Suite à cet état l'art nous soulignons la difficulté et le coût engendré par la mise en place d'une campagne d'évaluation. Ce coût est essentiellement induit par la nécessité de disposer d'experts capables, d'une part, de créer une solution de référence pour un texte donné et, d'autre part, d'évaluer la qualité des résultats retournés par les SREN afin de mesurer leur pertinence, au regard de la solution.

A ce titre, nous proposons dans la suite de cet article un modèle étendu d'EN et une plateforme permettant d'automatiser l'évaluation d'un SREN à partir d'un fichier de correction établi par un expert.

1. <http://nerd.eurecom.fr/>

3. SNERBM : Un cadre d'évaluation de SRENS

N
dié à
d'uti

Cadre d'évaluation de systèmes de REN spatiales 11/10/2016 18 016), dé-
lémarche

Reconnaissance d'entités nommées spatiales

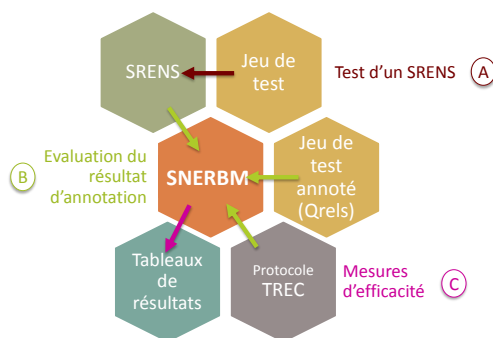


Figure 1. Le « benchmark » SNERBM

L'évaluation SNERBM consiste, pour un texte donné, à comparer les entités spatiales retournées par un SRENS (A) avec les entités spatiales listées dans un fichier *Qrels* (cf. section 5) de correction associé à ce texte et créé par des experts. La comparaison des entités détectées par le SRENS avec les entités listées dans le fichier de correction est réalisée sur la représentation géocodée des entités.

Pour le fichier *Qrels*, les représentations géocodées sont fournies avec précision par les experts sous forme de polygones et de polygones. Pour le fichier retourné par un SREN, deux cas de figure se présentent : soit le SRENS fournit nativement le géocodage des entités, soit ce travail est réalisé par les experts.

Le benchmark SNERBM (B) applique ensuite le protocole TREC pour la mesure d'efficacité (calcul de métriques telles que le rappel, la précision et la F1-Mesure (C) (Manning *et al.*, 2008)). Une synthèse des résultats de mesure de la qualité des annotations spatiales, produites par le SRENS sur le jeu de test, est ainsi produite.

La mise en œuvre du benchmark SNERBM nécessite l'intervention d'experts humains (Figure 2 - parties gauche et centrale) visant :

- la création du fichier de correction dans lequel chaque entité à reconnaître est géocodée l'une après l'autre à l'aide d'un service appelé manuellement ;
- le géocodage des entités retournées par le SRENS et la mise en forme de ces résultats afin de pouvoir les comparer au fichier *Qrels* de correction ;
- l'invocation du service TREC pour comparer le fichier précédemment mis en forme avec le fichier *Qrels*.

SAGEO'2016 – Nice, 6-9 décembre 2016

Nous proposons une plateforme capable d'enchaîner automatiquement ces différentes étapes (Figure 2 - partie droite) de telle sorte que le géocodage des ENS reconnues et leur évaluation via le protocole TREC soit automatisés.

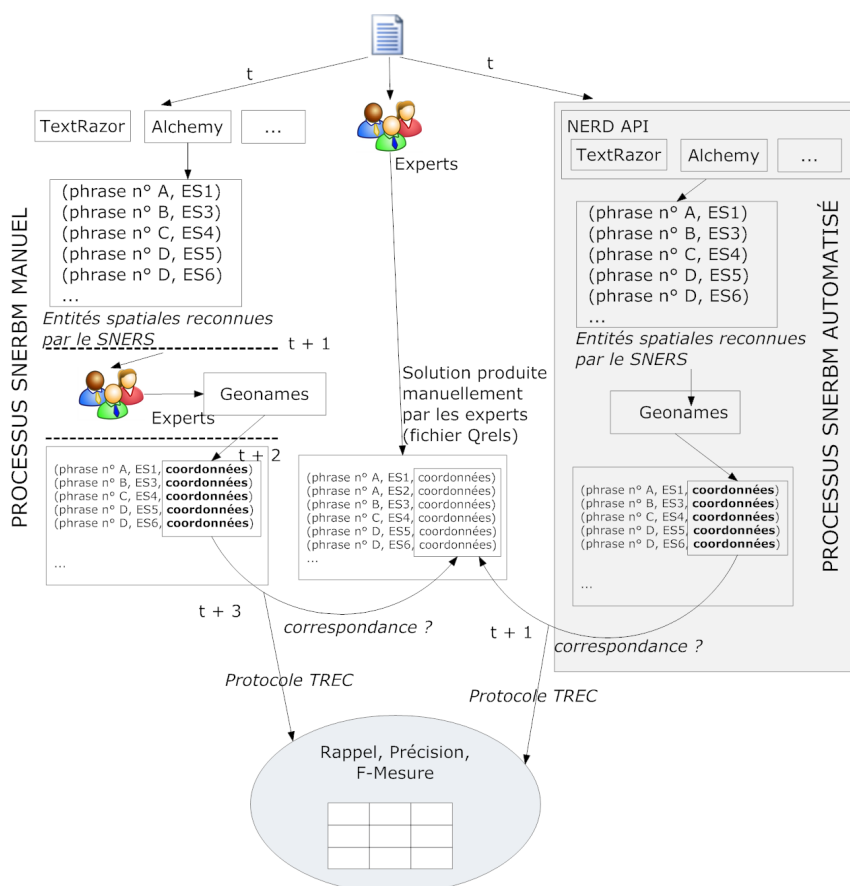


Figure 2. Automatisation du processus d'évaluation de SRENS

Dans la section suivante, nous présentons un modèle dédié à l'évaluation de SREN compatible avec le «benchmark» SNERBM.

4. Un modèle pour l'évaluation de SRENS

Conformément au cadre d'évaluation SNERBM (section 3), nous proposons un modèle permettant de faire une évaluation semi-automatique de SRENS.

Cette évaluation se base sur la comparaison des géolocalisations des entités spatiales reconnues par le SRENS (Figure 2) - il s'agit généralement de points - avec les géolocalisations des entités spatiales correspondantes dans le fichier

Un cadre pour l'évaluation du SRENS

de correction (Fichier *Qrels* - il s'agit de géométries). Nous souhaitons réaliser cette comparaison de manière automatique en vérifiant que les géométries correspondant aux entités reconnues par le SRENS sont en intersection avec les celles relatives aux entités présentes dans le fichier de correction.

A cette fin, mais aussi à des fins de visualisation et de comparaison d'entités nommées, il est nécessaire d'étendre le modèle de données de NERD (Listing 1) repris sur la Figure 3 via la classe *Entity* (A).

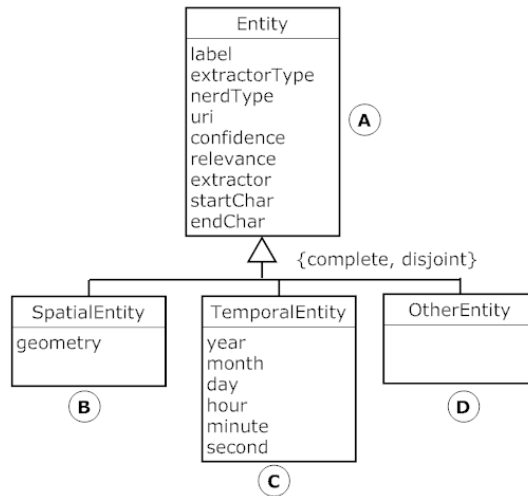


Figure 3. Modèle pivot pour l'évaluation de SRENS

Cette extension permet de :

1. catégoriser les entités reconnues en entités spatiales (B), temporelles (C) et autres (D);
2. d'enrichir chaque entité spatiale par une représentation numérique (attribut *geometry*).

Nous envisageons à terme d'appliquer la même démarche pour l'évaluation des SREN temporelles.

5. Une plateforme pour l'évaluation de SRENS dans le cadre de SNERBM

Pour mettre en place la visualisation des EN et l'évaluation des SRENS, nous proposons une plateforme modulaire présentée sur la Figure 4.

A partir d'un texte (A) sélectionné par l'utilisateur, la plateforme interroge un SRENS particulier via l'API NERD (B). Suite à cet appel, la plateforme récupère une liste d'entités nommées (C) conformément au modèle de données

SAGEO'2016 – Nice, 6-9 décembre 2016

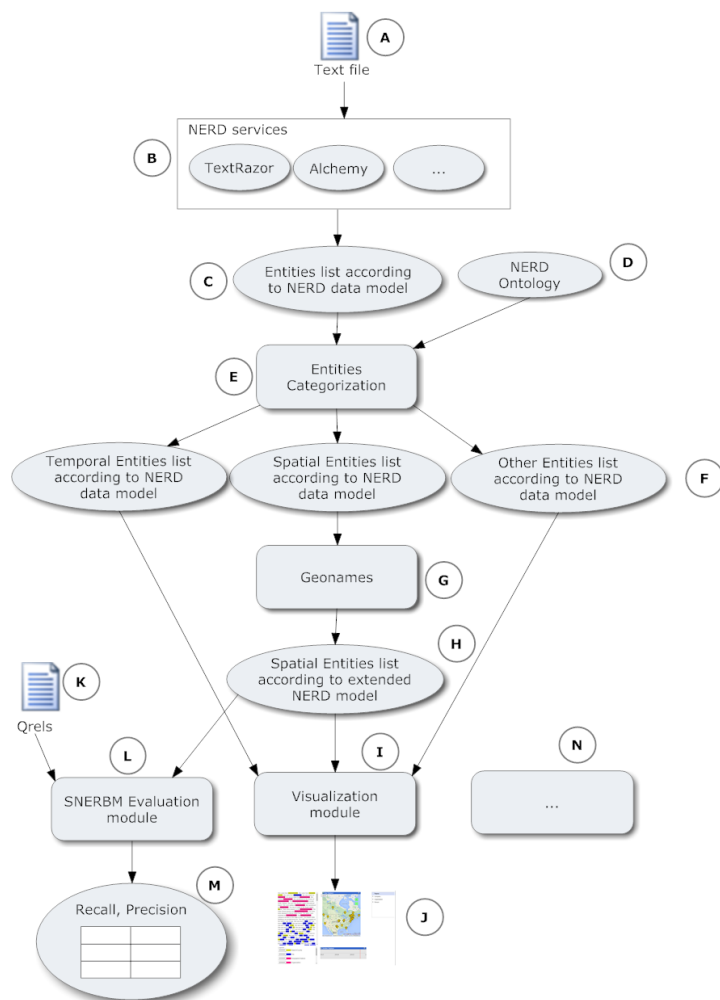


Figure 4. Architecture de la plateforme (évaluation, visualisation, fusion)

(Rizzo, Troncy, 2012) de NERD (Listing 1). En s'appuyant sur l'ontologie de NERD² (D), la plateforme catégorise (E) les entités retournées selon trois familles (F) : les entités spatiales, temporelles et autres.

Pour évaluer la pertinence des entités spatiales reconnues, la plateforme calcule les représentations géographiques de chaque entité via le service *Geonames* (G). Les entités spatiales reconnues par le SRENS sont donc, à ce stade,

2. <http://nerd.eurecom.fr/ontology>

Un cadre pour l'évaluation du SRENS

enrichies de leur géolocalisation (H). Cette géolocalisation permet :

- d’obtenir une représentation cartographique des résultats retournés par le SRENS (J) via un module de visualisation (I);
- d’évaluer (L) automatiquement (sur la base des inclusions des points - relatifs aux entités spatiales géolocalisées - dans les polygones résultats - relatifs aux *Qrels* de correction (K)) les taux de précision et de rappel (M);
- de lancer d’autres modules permettant de (N) fusionner, désambigüiser... les entités reconnues par un ou plusieurs SRENS.

Le fichier *Qrels* de correction (K) intègre, pour chaque entité nommée à reconnaître, le numéro de phrase et la phrase où l’entité apparaît, le type de cette entité et sa représentation géocodée précise sous forme de polygone. Le code ci-dessous (Listing 2) montre un extrait de fichier *Qrels* :

Listing 2 – Extrait d’un fichier *Qrels* de correction

1	1	She was born in Massachusetts.	ADM1	POLYGON((-73.234863 42.747012 ...))
2	2	She came from Belchertown.	PPL	POLYGON((-72.419643 42.300167 ...))
3	3	She climbed Mount Watatic.	Mountain	POLYGON((-71.921310 42.708426 ...))

Cette plateforme, en cours de développement, propose d’ores et déjà des modules opérationnels permettant par exemple :

- de visualiser (I) les entités reconnues par un SREN (H) via une interface graphique présentant (J) les entités spatiales sur une carte, les entités temporelles sur une frise chronologique et les autres entités sous forme arborescente classées selon leur type (Figure 5);
- de fusionner (N) des entités spatiales reconnues par divers SRENS (H) en tenant compte du nombre (*Nbr* sur la Figure 6) et de la pertinence (*R* sur la Figure 6) des entités spatiales reconnues par chaque SRENS (Figure 6). Le degré de pertinence (*R*) est fourni par les SRENS invoqués; chaque SREN ayant sa propre échelle pour exprimer le degré de pertinence, la plateforme ramène cette valeur sur un intervalle de 0 à 1 pour homogénéiser ces résultats et les rendre comparables entre eux.

6. Expérimentation de la plateforme

Nous prenons comme cas d’application le SREN Alchemy. Il est intégré à notre plateforme et nous a permis d’expérimenter les modules d’annotation et de visualisation. La Figure 5 illustre la visualisation d’EN nommées annotées par Alchemy conformément au modèle pivot présenté en section 4. De plus, nous avons évalué l’efficacité d’Alchemy, au sein de la plateforme, en utilisant le « benchmark » SNERBM (Palacio *et al.*, 2016) sur un texte composé de 200 phrases disponibles à l’adresse suivante : http://eott.univ-pau.fr/snerbm_eval/evaluationText.txt. Nous présentons les premiers résultats d’évaluation dans le Tableau 2.

SAGEO'2016 – Nice, 6-9 décembre 2016

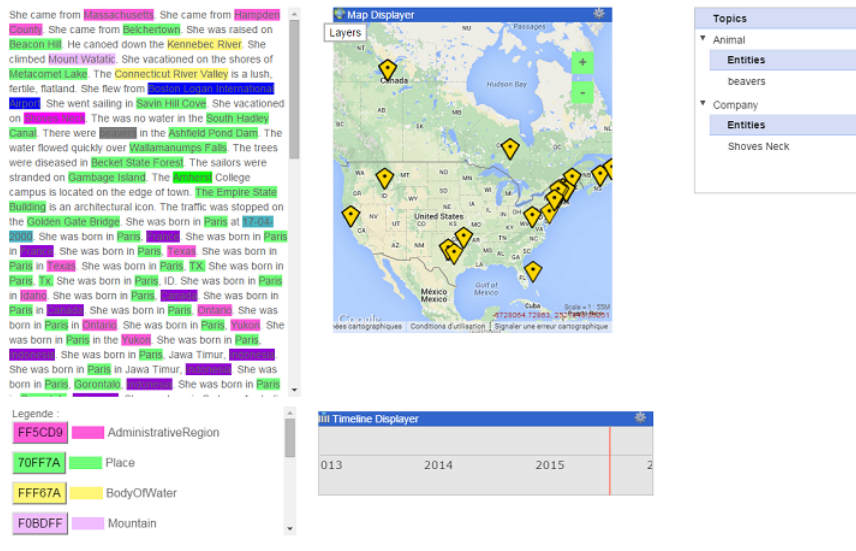


Figure 5. Illustration du module de visualisation

		PivotAlchemy.xml	PivotTextRazor.xml	fusion_07-02-2016_23-41.xml
Spatial	+	Nbr: 3 R:0.17	Nbr: 10 R:0	Nbr: 6 R:0.06
Country	+	Nbr: 2 R:0.17	Nbr: 6 R:0	Nbr: 2 R:0.17
Maroc		Nbr: 1 R:0.17	Nbr: 3 R:0	Nbr: 1 R:0.17
France		Nbr: 1 R:0.17	Nbr: 3 R:0	Nbr: 1 R:0.17
City	+	Nbr: 1 R:0.17	Nbr: 0 R:0	Nbr: 0 R:0
Amherst		Nbr: 1 R:0.17	Nbr: 0 R:0	Nbr: 0 R:0
University	+	Nbr: 0 R:0	Nbr: 4 R:0	Nbr: 4 R:0
Amherst College		Nbr: 0 R:0	Nbr: 4 R:0	Nbr: 4 R:0
Other	+	Nbr: 0 R:0	Nbr: 0 R:0	Nbr: 0 R:0
Temporal	+	Nbr: 0 R:0	Nbr: 0 R:0	Nbr: 0 R:0

Figure 6. Fusion d'entités spatiales reconnues par plusieurs SRENS

Tableau 2. Evaluation d'Alchemy via la plateforme et le cadre SNERBM

Outil de REN	Alchemy
Nombre total d'ENS	317
Nombre total d'ENS reconnues	197
Nombre total d'ENS pertinentes	93
Rappel	0.47
Précision	0.29
F1-Mesure	0.35

Un cadre pour l'évaluation du SRENS

Le rappel (pourcentage d'annotations spatiales pertinentes par rapport au nombre total d'ENS dans le document) est de 47%. La précision (pourcentage d'annotations spatiales pertinentes par rapport au nombre total d'annotations spatiales) est de 29%. Enfin, la F1-Mesure correspondante est de 35%. Ces premiers résultats sont encourageants et nous allons procéder à l'évaluation d'autres SRENS intégrés dans la plateforme.

Nous envisageons par la suite d'exploiter les indicateurs de rappel, précision et F1-Mesure dans le module de fusion présenté Figure 6. Ce module permet de fusionner des entités reconnues par différents SREN sur un même texte. Les critères de fusion proposés se basent actuellement uniquement sur le nombre (pour privilégier la quantité) ou la pertinence (pour privilégier la qualité) des entités reconnues. Il sera intéressant d'étendre ces indicateurs en intégrant les retours proposés par le module d'évaluation pour, par exemple, fusionner des entités en fonctions des valeurs de précision, rappel ou F1-Mesure calculées.

7. Conclusion et perspectives

Dans cet article, nous avons proposé une plateforme et un modèle dédiés à l'évaluation semi-automatique de SRENS conformément au cadre d'évaluation SNERBM. Le modèle proposé est une extension du modèle NERD permettant d'homogénéiser les retours des SRENS. L'extension consiste à enrichir automatiquement la représentation des entités spatiales reconnues en leur adjoignant une représentation numérique sous forme de coordonnées géographiques. L'ajout de ces coordonnées permet entre autres :

- d'identifier de manière unique chaque entité reconnue,
- de comparer automatiquement les entités reconnues par le SREN avec les entités listées dans la solution des experts,
- de représenter chaque entité sur un support cartographique afin de faciliter une évaluation visuelle des résultats retournés par le SRENS.

Cette extension du modèle NERD apporte une plus-value pour mettre en oeuvre le benchmark SNERBM car elle permet d'automatiser le processus d'évaluation. Les experts doivent toujours produire manuellement un fichier de correction type pour un texte donné mais une fois ces fichiers créés le processus d'évaluation peut être lancé automatiquement pour obtenir directement une évaluation selon des critères de rappel, précision et F1-Mesure. Il devient alors peu coûteux de lancer une série d'évaluations impliquant différents SRENS afin de les situer les uns par rapport aux autres.

Nous envisageons plusieurs perspectives à court terme concernant l'évaluation des SRENS :

- Désambigüiser les retours de la plateforme Geonames qui propose parfois plusieurs géocodages pour une même ENS (par défaut, la plateforme conserve uniquement le premier géocodage retourné) ;

SAGEO'2016 – Nice, 6-9 décembre 2016

- Intégrer dans le fichier de correction les types d'entités spatiales à reconnaître (commune, pays, fleuve ...) et prendre en compte une correspondance des types lors de la comparaison ;

- Mesurer l'éventuelle perte d'efficacité des SREN due à leur intégration dans une plateforme qui impose la transformation des annotations dans un modèle pivot générique (Figure 3). Pour cela, nous envisageons d'évaluer deux versions de chaque SRENS selon le « benchmark » SNERBM : (i) le SRENS est invoqué directement pour annoter le jeu de données ; (ii) le SRENS est invoqué via notre plateforme support (avec usage du modèle pivot) pour annoter le même jeu de données.

A moyen terme, nous envisageons d'étendre ce processus d'évaluation aux entités temporelles en appliquant ce principe d'intersection sur des données temporelles formalisées sous forme calendaire telle que déjà définie dans le modèle pivot présenté Figure 3 ©.

A plus long terme nous visons à étendre la plateforme d'évaluation pour intégrer des modules permettant entre autres :

- d'intégrer d'autres SRENS que ceux proposés par l'API NERD ;
- d'ajouter un module permettant de désambigüiser les propositions de certains SREN qui retournent plusieurs entités candidates pour une même annotation textuelle.

Bibliographie

- Anastácio I., Martins B., Calado P. (2010). Using the geographic scopes of web documents for contextual advertising. In *GIR'10: Proceedings of the 6th Workshop on Geographic Information Retrieval*, p. 18:1–18:8. New York, NY, USA, ACM.
- Andogah G. (2010). *Geographically Constrained Information Retrieval*. Thèse de doctorat, University of Groningen, Netherlands.
- Bucher B., Clough P., Joho H., Purves R., Syed A. K. (2005). Geographic IR Systems: Requirements and Evaluation. In *ICC'05: Proceedings of the 22nd International Cartographic Conference*. Global Congressos.
- Chinchor N. A. (1998). MUC/MET evaluation trends. In *Proceedings of the TIPSTER text program: Phase III*, p. 235–239. Baltimore, Maryland, USA, Association for Computational Linguistics.
- Chinchor N. A. (1998). Overview of MUC-7. In *MUC-7: Proceedings of the 7th Message Understanding Conference*.
- Dean-Hall A., Clarke C. L. A., Kamps J., Thomas P., Simone N., Voorhees E. (2013). Overview of the TREC 2013 Contextual Suggestion Track. In *TREC'13: Proceedings of the 22nd text retrieval conference*. Gaithersburg, MD, NIST.
- D'Ignazio C. (2013). *Big data, news and geography: Research update*. Consulté sur <https://civic.mit.edu/blog/kanarinka/big-data-news-and-geography> (MIT Center for Civic Media)

Un cadre pour l'évaluation du SRENS

- Gelernter J., Zhang W. (2013). Cross-lingual geo-parsing for non-structured data. In *GIR'13: Proceedings of the 7th Workshop on Geographic Information Retrieval*, p. 64–71. New York, NY, USA, ACM.
- Grishman R., Sundheim B. (1996). Message understanding conference- 6: A brief history. In *16th international conference on computational linguistics, proceedings of the conference, COLING 1996, center for sprogteknologi, copenhagen, denmark, august 5-9, 1996*, p. 466–471.
- Leidner J. L. (2007). *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names*. Thèse de doctorat, Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh, Scotland.
- Lenzi V. B., Speranza M., Sprugnoli R. (2013). Named entity recognition on transcribed broadcast news at EVALITA 2011. In *Revised Papers from EVALITA'11: International Workshop on the Evaluation of Natural Language and Speech Tools for Italian*, vol. 7689, p. 86-97. Springer.
- Lieberman M. D., Samet H., Sankaranarayanan J. (2010). Geotagging with local lexicons to build indexes for textually-specified spatial data. In F. Li *et al.* (Eds.), *Proceedings of the 26th international conference on data engineering, ICDE 2010, march 1-6, 2010, long beach, california, USA*, p. 201–212. IEEE.
- Lingad J., Karimi S., Yin J. (2013). Location extraction from disaster-related microblogs. In L. Carr *et al.* (Eds.), *22nd international world wide web conference, WWW '13, rio de janeiro, brazil, may 13-17, 2013, companion volume*, p. 1017–1020. International World Wide Web Conferences Steering Committee / ACM.
- Luong T. N., Marquesuzaà C., Etcheverry P., Nodenot T., Laborie S. (2015). Facilitating the design/evaluation process of web-based geographic applications: A case study with windmesh. In T. K. Dang, R. Wagner, J. Küng, N. Thoai, M. Takizawa, E. J. Neuhold (Eds.), *Future data and security engineering - second international conference, FDSE 2015, ho chi minh city, vietnam, november 23-25, 2015, proceedings*, vol. 9446, p. 259–271. Springer.
- Mandl T., Carvalho P., Nunzio G. M. D., Gey F. C., Larson R. R., Santos D. *et al.* (2009). GeoCLEF 2008: The CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview. In C. Peters *et al.* (Eds.), *Revised Selected Papers of CLEF'08: Proceedings of the 9th Workshop of the Cross-Language Evaluation Forum*, vol. 5706, p. 808–821.
- Manning C. D., Raghavan P., Schütze H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Marrero M., Sánchez-Cuadrado S., Lara J. M., Andreadakis G. (2009). Evaluation of named entity extraction systems. *Advances in Computational Linguistics. Research in Computing Science*, vol. 41, p. 47–58.
- Nouvel D. (2012). *Reconnaissance des entités nommées par exploration de règles d'annotation*. Thèse de doctorat, Université François Rabelais de Tours, France.
- Palacio D., Sallaberry C., Cabanac G., Hubert G. (2016). Cadre d'évaluation de systèmes de reconnaissance d'entités nommées spatiales. In *Actes du XXXIVe congrès INFORSID, Grenoble, France, may 31 - june 3, 2016.*, p. 149–164. Consulté sur http://inforsid.fr/actes/2016/INFORSID2016_paper_9.pdf

SAGEO'2016 – Nice, 6-9 décembre 2016

- Rizzo G., Troncy R. (2012, 04/2012). Nerd: A framework for unifying named entity recognition and disambiguation web extraction tools. In *European chapter of the association for computational linguistics (EACL'12)*. Avignon, France.
- Rizzo G., Troncy R., Hellmann S., Brümmer M. (2012). NERD meets NIF: lifting NLP extraction results to the linked data cloud. In C. Bizer, T. Heath, T. Berners-Lee, M. Hausenblas (Eds.), *WWW2012 workshop on linked data on the web, lyon, france, 16 april, 2012*, vol. 937.
- Santos D., Seco N., Cardoso N., Vilela R. (2006). HAREM: An Advanced NER Evaluation Contest for Portuguese. In *LREC'06: Proceedings of the 5th International Conference on Language Resources and Evaluation*, p. 1986–1991.
- Sekine S., Eriguchi Y. (2000). Japanese Named Entity Extraction Evaluation – Analysis of Results. In *COLING'00: Proceedings of the 18th conference on Computational linguistics*, p. 1106-1110. Association for Computational Linguistics.
- Strassel S., Przyboccki M. A., Peterson K., Song Z., Maeda K. (2008). Linguistic Resources and Evaluation Techniques for Evaluation of Cross-Document Automatic Content Extraction. In *LREC'08: Proceedings of the International Conference on Language Resources and Evaluation*, p. 2706–2709.
- Tjong Kim Sang E. F., De Meulder F. (2003). Introduction to the CoNLL-2003 Shared Task Language-Independent Named Entity Recognition. In W. Daelemans, M. Osborne (Eds.), *CoNLL-2003: Proceedings of the 7th conference on Natural language learning at HLT-NAACL 2003*, p. 142–147. Association for Computational Linguistics.
- Tobin R., Grover C., Byrne K., Reid J., Walsh J. (2010). Evaluation of georeferencing. In R. Purves, P. Clough, C. B. Jones (Eds.), *GIR'10: Proceedings of the 6th Workshop on Geographic Information Retrieval*. New York, NY, USA, ACM.
- Voorhees E. M. (2001). Overview of TREC 2001. In *TREC'01: Proceedings of the 9th Text REtrieval Conference*. Gaithersburg, MD, USA.

Conception logique et physique des hiérarchies spatiales non strictes dans les entrepôts de données spatiales

Ibtisam Ferrahi ¹, Sandro Bimonte ², Kamel Boukhalfa ³

1. *UMBB, Université M'hamed Bougara- Faculté des sciences
Avenue de l'indépendance, 35000, Boumerdès, Algérie.
ferrahi.ibtisam@gmail.com*

2. *Irstea, TSCF
9 Avenue Blaise Pascal CS20085,
63178 Aubière, France
sandro.bimonte@irstea.fr*

3. *USTHB, Université des sciences et de la technologie-Houari Boumedienne,
Alger- Algérie.
kboukhalfa@usthb.dz*

RÉSUMÉ. L'optimisation des performances des systèmes SOLAP ne cesse de susciter l'intérêt des chercheurs ces dernières années. Plusieurs techniques d'indexation ont été proposées pour atteindre cet objectif. La plupart des travaux effectués dans ce domaine s'est intéressée aux dimensions spatiales avec des hiérarchies spatiales strictes et peu de travaux se sont focalisés sur les hiérarchies spatiales non strictes. Nous montrons dans le présent article que les schémas logiques proposés dans la littérature ainsi que les techniques d'indexation ne sont pas adaptés aux hiérarchies spatiales non strictes. Nous proposons un nouveau schéma logique supportant les hiérarchies spatiales non strictes. Nous étendons ainsi l'index bitmap afin d'optimiser les requêtes définies sur des dimensions spatiales qui incluent ce type de hiérarchies spatiales. Nous montrons à travers une étude expérimentale l'efficacité de notre solution.

ABSTRACT. Performance optimization of SOLAP systems continues to attract the interest of researchers in recent years. Several indexing techniques have been proposed to achieve this goal. Most work in this domain have focused on the spatial dimensions with strict hierarchies and little work has focused on non strict hierarchies. We show, in this paper, that the logical models proposed in the literature and indexing techniques are not suitable to non strict hierarchies. We propose a new logical schema supporting the non strict hierarchies and a bitmap index to optimize queries defined on spatial dimensions with several non strict hierarchies. We show through an experimental study the effectiveness of our solution.

MOTS-CLES : Olap spatial, entrepôt de données spatiales, index, hiérarchies non strictes.

KEYWORDS: Spatial OLAP, Spatial Data Warehouses, Indexes, non strict hierarchies.

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

Aujourd'hui de plus en plus des données spatiales sont collectées via les réseaux des capteurs, les réseaux sociaux, etc. (Lee et Kang 2015). Ces données alimentent les systèmes de GeoBusiness Intelligence comme le Spatial Data Mining, le Reporting spatial, et l'OLAP Spatial. Les Entrepôts de Données (EDs) et les systèmes OLAP sont des technologies d'aide à la décision, dont l'objectif est de permettre aux décideurs de formuler des requêtes complexes par le biais d'une interface visuelle et interactive (Kimball 1996). Afin de bénéficier de ces technologies également dans le contexte des données spatiales, certains travaux introduisent le concept d'OLAP Spatial (SOLAP), qui intègre les fonctionnalités des systèmes OLAP et des Systèmes d'Information Géographique (SIGs) dans un environnement unique (Bédard et Han 2001). Le SOLAP a été utilisé avec succès dans différents domaines d'applications comme l'agriculture, l'environnement, le marketing, etc. Le SOLAP étend les concepts principaux d'OLAP comme les dimensions et les mesures, en introduisant la composante spatiale (Bédard et Han 2001). Les données spatio-multidimensionnelles sont stockées dans les Entrepôts de Données Spatiales (EDSs). Un EDS a été défini comme « Une collection de données spatiales et non spatiales, orientées-sujet, intégrées, variantes dans le temps, et non volatiles dédiées à la prise de décision spatiale » (Stefanovic, Han et al. 2000). Si dans l'âge du Spatial Big Data de nouvelles technologies pour le stockage et le requêtage d'un volume important de données spatiales sont apparues récemment (ex. MongoDB, CouchDB, Neo4j, etc.), les Systèmes de Gestion de Données Spatiales (SGDBs) relationnels comme PostGIS, Oracle Spatial restent des solutions efficaces pour le Small Data (Kitchin et Lauriault, 2015). Dans le contexte du Small Data et des EDSs relationnels, nous étudions l'impact des hiérarchies spatiales non strictes sur la modélisation logique et les performances des EDSs. Un EDS relationnel hérite de plusieurs composantes des entrepôts de données classiques, tels que la notion du fait, mesure et dimension, etc. Une dimension correspond à l'axe de l'analyse, elle est décrite par une ou plusieurs hiérarchies. Une hiérarchie permet d'agréger les mesures selon plusieurs niveaux de granularité. Les dimensions spatiales peuvent aussi être organisées selon plusieurs niveaux de granularité ou selon différents niveaux de détails pour permettre les opérations SOLAP comme roll-up et drill-down. Une hiérarchie spatiale non stricte (Malinowski et Zimanyi, 2008) est une hiérarchie spatiale dont certains membres possèdent plusieurs membres pères, ou en d'autres termes une relation plusieurs à plusieurs (n-n) existe entre les niveaux spatiaux. Ces hiérarchies spatiales posent des problèmes d'analyse car elles sont à l'origine des problèmes de double comptage (Lechtenborger et Vossen 2003). Ces hiérarchies spatiales caractérisent nombreuses données spatiales comme les bassins versants, les réseaux sans fil, etc. Par conséquent, plusieurs solutions à différents niveaux (conceptuel (Malinowski et Zimanyi, 2008) et visualisation (Mansmann et Scholl, 2006)) ont été proposées dans la littérature permettant une correcte agrégation des mesures le long des hiérarchies spatiales non strictes.

Cependant, les modèles spatio-multidimensionnels logiques existants ne sont pas adaptés aux hiérarchies spatiales non strictes définies sur plusieurs niveaux, car ils ne traitent pas le problème du double comptage dans ce cas. Ainsi, au meilleur de notre connaissance, aucun travail n'étudie les performances associées à ce type de hiérarchies spatiales en proposant une optimisation au niveau physique. Pour remédier aux insuffisances citées ci-dessus et motivés par l'importance des hiérarchies spatiales non strictes dans les applications SOLAP réelles, nous présentons dans cet article une extension du modèle logique en flocon (Kimball, 1996). Nous étendons ainsi l'index bitmap pour optimiser les requêtes qui incluent des hiérarchies spatiales non strictes dans les EDSs.

L'article est organisé de la manière suivante : la section 2 présente les motivations de notre travail à travers un cas d'étude. La section 3 introduit notre nouveau modèle logique. La section 4 détaille le nouvel index proposé. La section 5 présente l'étude des performances. Nous présentons l'état de l'art dans la section 6.

2. Motivation

L'EDS étudié est composé d'un fait avec une mesure *Nb_Accident* et deux dimensions : «Temps» et «Localisation» (figure 1a). La dimension «Temps» est une dimension temporelle classique composée des niveaux : «Heure», «Jour», «Mois» et «Année». La dimension «Localisation» est une dimension spatiale qui inclut une hiérarchie spatiale non stricte. Elle est composée de trois niveaux spatiaux «Segment», «Route» et «Région». Un segment de route peut appartenir à plusieurs routes (exemples de routes qui se joignent sur un ou plusieurs segments) (figure 1b). Une route peut englober plusieurs segments. De la même manière, une route peut traverser plusieurs régions et une région peut avoir plusieurs routes. Un exemple de faits est représenté dans la figure 1c. En utilisant cet EDS, il est possible de répondre aux requêtes SOLAP comme : «Quel est le total d'accidents par segment de route et année ?», «Quel est le total d'accidents par route et année ?», ou encore « Quel est le nombre total d'accidents par région et année ? ». Comme montré en figure 1b, la hiérarchie spatiale présentée est non stricte. Par exemple, le segment *SI* appartient à deux routes *R1* et *R2*, et aussi la route *R1* traverse les trois régions *Region1*, *Region2* et *Region3*. En particulier, si l'on considère le prédicat topologique d'intersection, alors le segment *SI* est inclus à 50% dans *R1* et à 50% *R2*. Ainsi la route *R1* est incluse à 20% dans *Region1*, à 40% dans *Region2* et à 40% dans *Region3*. Cette valeur d'inclusion nous permet d'éviter le problème du double comptage comme décrit dans la suite de l'article.

2.1 Les solutions existantes

A notre connaissance seulement (Malinowski et Zimanyi, 2008) propose un modèle logique pour gérer les hiérarchies spatiales non strictes. Le modèle proposé permet de normaliser les niveaux dont la relation est plusieurs à plusieurs. Il s'agit d'une extension du modèle proposé par (Kimball, 1996) pour les EDs classiques. En particulier, une table, appelée Bridge Table, est ajoutée entre les niveaux. Cette table contient un attribut numérique (Distribution Factor) représentant le degré

SAGEO'2016 – Nice, 6-9 décembre 2016

d'inclusion d'un membre dans son père (comme montré en figure 1b pour la hiérarchie spatiale).

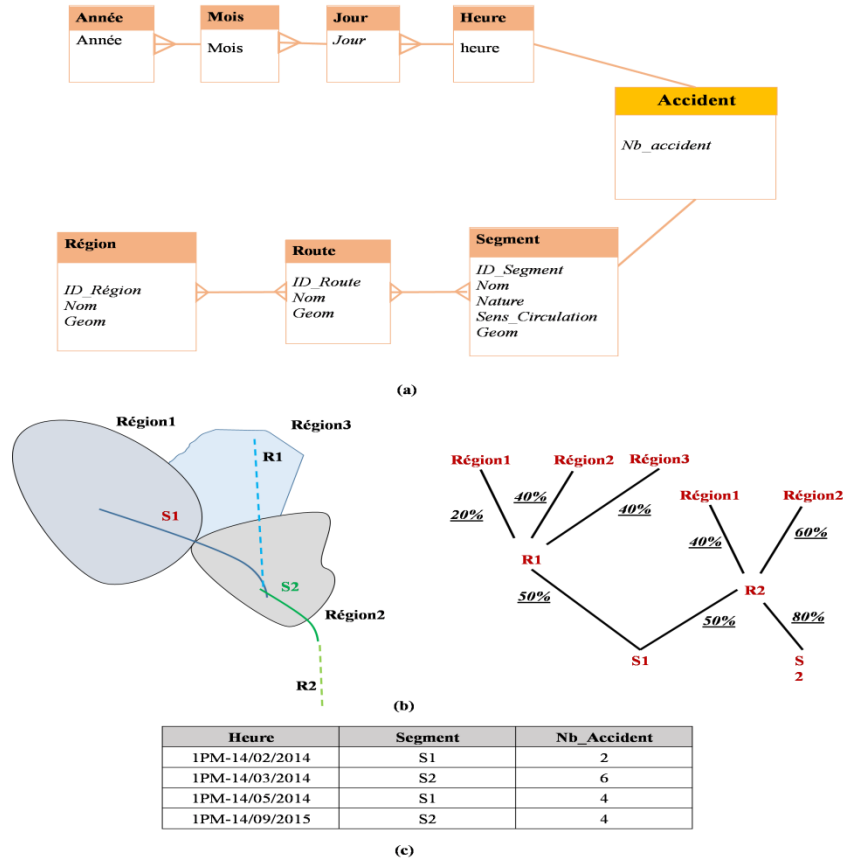


FIGURE 1: a) *Modèle conceptuel spatio-multidimensionnel*, b) *Instance de la hiérarchie spatiale non stricte*, c) *Faits*

En appliquant cette approche sur notre EDS, le modèle logique résultat est montré en figure 2. En utilisant ce modèle logique, la requête «*Quel est le total d'accidents par route ?*» peut être traduite en SQL par la requête de la figure 3a. Le résultat de cette requête est montré en figure 3b. Ce résultat est correct, car l'utilisation d'un facteur de distribution permet d'éviter le double comptage. Par exemple, seuls 50% des accidents survenus sur le segment *S1* sont comptabilisés pour la route *R1*. Par contre, sans l'utilisation du facteur de distribution, tous les accidents survenus sur *S1* sont comptabilisés pour *R1* et *R2* en même temps, ce qui donne le résultat erroné représenté dans la figure 3c.

Conception logique et physique des hiérarchies spatiales non strictes dans les EDSs

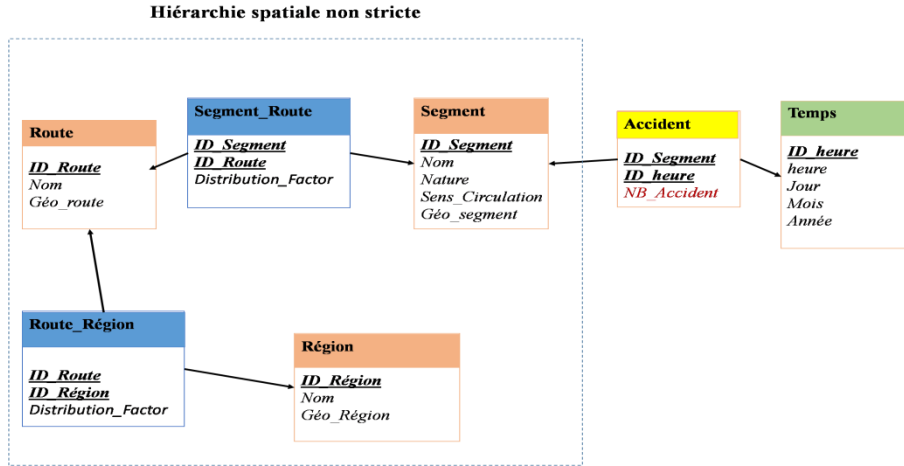


FIGURE 2. Schéma logique avec « Bridge Table »

2.2 Limites des solutions existantes

Les limites des travaux effectués sur les hiérarchies spatiales non strictes peuvent se présenter dans deux niveaux : logique et physique.

```

SELECT SUM (Nb_accident * SR.distribution_factor), R.nom
FROM Accident I, Segment S, Segment_Route SR, Route R WHERE I.ID_segmet= S. ID_segmet AND
S. ID_segmet= SR.ID_segmet AND R. ID_route= RR.ID_route
GROUP BY R.nom
    
```

(a)

Route	Nb_Accident
R1	3
R2	11

(b)

Route	Nb_Accident
R1	6
R2	16

(c)

FIGURE 3. a) Requête SQL sur une hiérarchie spatiale non stricte, b) Résultat correct avec le facteur de distribution, c) Résultat incorrect sans le facteur de distribution

Au niveau logique, le modèle logique de Zimanyi (Malinowski et Zimanyi, 2008) ne permet pas gérer le problème de double comptage lorsque plusieurs Bridges Tables sont utilisées par les requêtes SOLAP. La présence de Bridges Tables alourdit le processus de calcul des requêtes SOLAP classiques, car elles nécessitent plusieurs opérations de jointure entre la table de faits, les niveaux spatiaux et les Bridges Tables.

SAGEO'2016 – Nice, 6-9 décembre 2016

Considérons la requête suivante « *Quel est le total d'accidents par région ?* » qui peut être traduite dans la requête SQL de figure 4a et dont le résultat est montré dans la figure 4b. Ce résultat est incorrect, car le segment *S1* est inclus à 50% dans *Région1*, et non pas à 10% ($50\% * 20\%$). Aussi, le segment *S2* n'appartient pas à *Région1*, alors que la requête retourne 32% ($40\% * 80\%$). En d'autres termes, la prise en compte des facteurs de distribution de tous les niveaux spatiaux dans le calcul de l'agrégation peut porter à des résultats incorrects. Ce problème a été défini formellement comme « *partial containment* » par (Jensen et al, 2004), qui, par contre, ne détaille pas la représentation logique des hiérarchies complexes.

```

SELECT SUM (Nb_accidents * SR.distribution_factor*RR.distribution_factor), Re.nom
FROM Accident I, Segment S, Segment_Route SR, Route R, Route_Region RR, Region Re
WHERE I.ID_segment= S.ID_segment AND S.ID_segment= SR.ID_segment
      AND R.ID_route= RR.ID_route AND RR.ID_region = Re.ID_region
GROUP BY Re.nom

```

(a)

Route	Nb_Accident
R1	3
R2	11

(b)

FIGURE 4. a) Requêtes SQL sur une hiérarchie spatiale non stricte avec plusieurs niveaux, b) Résultat incorrect avec le facteur de distribution

Au niveau physique, le traitement des requêtes qui utilisent des niveaux non stricts nécessite des opérations de jointure entre la table de faits, les niveaux spatiaux et les Bridge Tables. Par conséquent, plus ces tables sont volumineuses plus le temps de traitement de ces requêtes devient important. Ainsi, l'association *n-n* qui existe entre le fait et les membres spatiaux rend impossible l'usage des méthodes existantes d'indexation des jointures comme par exemple le bitmap (O'Neil et Graefe, 1995).

3. Le modèle logique spatial-NN pour les hiérarchies spatiales non strictes

Vu les inconvénients rencontrés en utilisant le modèle avec « *Bridge Table* » (Malinowski et Zimanyi, 2008), nous présentons dans cette section un nouveau modèle logique pour la prise en compte des hiérarchies spatiales non strictes dans les EDSs. Le modèle que nous proposons est une extension du modèle décrit dans la section précédente 2.1. Notre proposition permet de déplacer les tables « *Bridge Table* » au niveau spatial le plus fin. De cette façon, les problèmes liés à la distributivité des facteurs de distribution sont éliminés. Le modèle est montré en détails dans la figure 5.

Conception logique et physique des hiérarchies spatiales non strictes dans les EDSS

«*Spatial Bottom Level Table*» représente la table utilisée pour implémenter le niveau spatial le plus détaillé. Cette table est associée via une clé primaire à une ou plusieurs tables de faits «*Fact Table*» (s’il s’agit d’un schéma en constellation ou pas). Elle est associée ainsi via une clé primaire à m bridge tables «*Bridge Table*». Une table bridge «*Bridge Table*» est définie par niveau spatial «*Spatial Bottom Level Table*» avec une association non stricte (n-n) dans la hiérarchie spatiale.

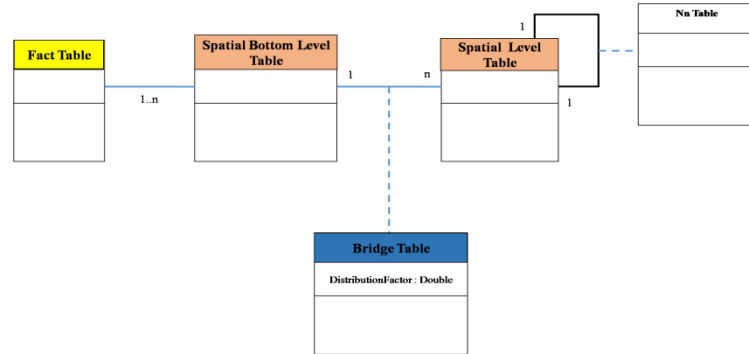
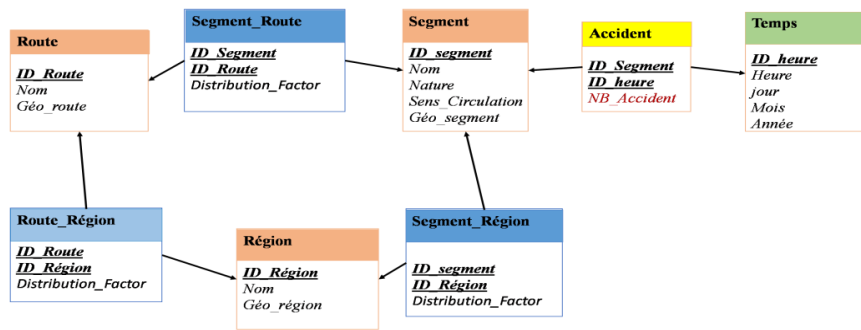


FIGURE 5. Méta modèle logique «*spatial-NN*»



(a)

```

SELECT SUM (Nb_accident * SR.Facteur_distribution), Re.nom
FROM Accident I, Segment S, Segment_Region SR, Region Re
WHERE I.ID_segment= S. ID_segment AND S. ID_segment= SR. ID_segment AND SR. ID_région=
Re.ID_région
GROUP BY Re.nom
    
```

(b)

FIGURE 6. a) Exemple du modèle logique spatial-NN, b) Requête SQL

Dans notre exemple, la table «*Spatial Bottom Level Table*» est représentée par la table «*Segment*» (figure 6a). Nous avons défini ainsi deux bridges tables associées à

SAGEO'2016 – Nice, 6-9 décembre 2016

«*Segment*» : «*Segment-Région*» et «*Segment-Route*». Ces bridges tables contiennent un attribut numérique, appelé *DistributionFactor*. Il représente le facteur de distribution calculé entre le membre spatial du niveau spatial de plus haut niveau (par exemple *Région1*) et chaque membre spatial du niveau spatial le plus détaillé (par exemple *SI*), qui est son descendant. Par conséquent, chaque bridge table est associée via une clé étrangère à la table «*Spatial Level Table*». Cette dernière permet de représenter le niveau spatial impliqué dans la relation n-n.

Dans notre exemple, «*Segment-Région*» est associée à «*Région*», et «*Segment-Route*» est associée à «*Route*». En considérant cette nouvelle politique de calcul du facteur de distribution, le segment *SI* de notre exemple sera associé à 50% à *Region1*, qui, comme expliqué précédemment, est le bon facteur de distribution topologique.

```

SELECT SUM (Nb_accident * SR.Facteur_Distribution), Re.nom
FROM Accident I, Segment S, Segment_Région SR, Région Re
WHERE I.ID_segment= S. ID_segment AND S. ID_segment= SR. ID_segment
AND SR. ID_région= Re.ID_région and Re.ID_région
  IN ( SELECT Re.ID_région
        FROM Région Re, Route_Région RR, Route R
        WHERE Re.ID_région = RR.ID_région AND R.ID_route = RR.ID_route
        AND R.nom='R1 )
GROUP BY Re.nom

```

FIGURE 7. Requête SQL nécessitant l'utilisation des NN tables

En utilisant notre nouveau modèle logique, la requête «Quel est le total d'accidents par région ?» peut être traduite dans la requête SQL de la figure 6(b). Il est aussi important de souligner que dans notre modèle nous avons gardé les tables («*NN Table*») représentant les associations n-n entre les niveaux spatiaux (figure 5), car elles permettent de représenter les liens de paternité qui existent entre les membres spatiaux. Dans la figure 6(a), la table «*Route_Region*» permet de représenter les membres pères au niveau «*Région*» des membres du niveau «*Route*». Par exemple, la route *R1* a trois membres pères *Région1*, *Région2* et *Région3* (figure 1b). Ces tables nous permettent de répondre aux requêtes utilisant les relations hiérarchiques entre les membres comme par exemple : «*Quel est le total d'accidents par région traversée par la route R1?*» peut être traduite dans la requête SQL de la figure 7.

```

SELECT {[Measures].members} ON COLUMNS {[Temps]. [All Temps]} ON ROWS
FROM [Accident] WHERE [Localisation]. [Auvergne]

```

FIGURE 8. Requête de type «*Spatial Slice*» en MDX

4. Un index bitmap pour les hiérarchies spatiales non strictes (Index-NN)

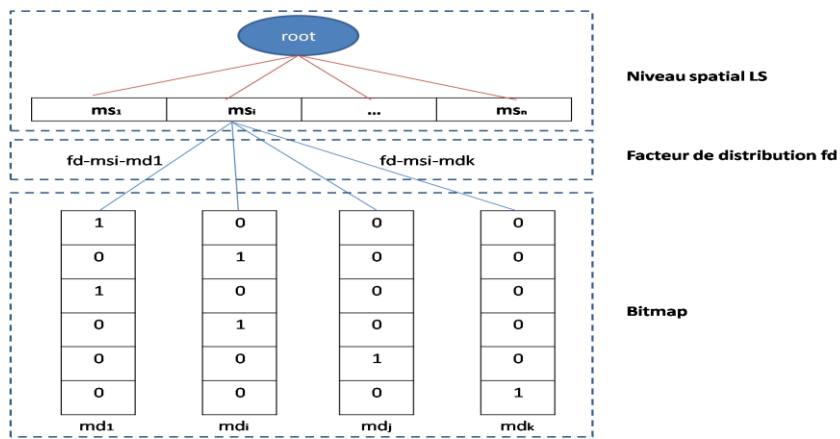
Afin de réduire le temps d'exécution des requêtes de type «*Spatial Slice*» sur les hiérarchies spatiales non strictes, nous avons étendu l'index bitmap (O'Neil et Graefe, 1995) pour définir un nouvel index nommé *Index-NN*. Classiquement, un opérateur *Spatial Slice* dans les serveurs SOLAP permet de couper l'EDS en utilisant un membre de la dimension spatiale. Au niveau du serveur SOLAP, il est souvent représenté par la clause WHERE de la requête MDX. Par exemple, la requête «*Quel est le total d'accidents par heure dans la région Auvergne ?*» peut être traduite par la requête MDX représentée dans la figure 8. La traduction SQL de cette requête engendre de nombreuses jointures comme décrit dans la section 3 (voir figure 11(a)). Les requêtes de type «*Spatial Slice*» permettent donc l'agrégation d'une mesure numérique (*m*) (par exemple le nombre d'accidents) sur un ensemble de niveaux non spatiaux et un membre spatial (par exemple la région *Auvergne*) appartenant à un niveau spatial (*LS*) (par exemple le niveau spatial *Région*) impliqué dans une relation spatiale non stricte.

Conceptuellement, l'index-NN est constitué de deux parties. La première partie est un arbre, et la deuxième est une structure de correspondance (figure 9). L'arbre est composé de deux niveaux. Le premier niveau est la racine, le deuxième représente chaque membre spatial (*ms*) du niveau spatial *LS*. En fait, index-NN est créé par niveau spatial. La structure de correspondance utilise un index bitmap composé de *k* vecteurs de bits, où *k* représente le nombre de membres spatiaux (*md*) du niveau spatial le plus détaillé (*LD*). Chaque vecteur est attaché aux feuilles de l'arbre par un ou plusieurs arcs. L'arc représente le facteur de distribution entre les membres spatiaux *ms* et *md* (*fd-ms-md*). Dans chaque vecteur, chaque n-uplet de la table de faits est associé à un bit qui prend la valeur 1 si le tuple correspond au membre du niveau spatial le plus détaillé, 0 dans le cas contraire. En fait, ce type d'indexation permet le calcul préalable des jointures qui peuvent exister entre la table des faits et la dimension spatiale via son niveau le plus fin. Par exemple, dans notre cas d'étude s'il y a 500 segments alors, chaque région sera associée aux 500 vecteurs avec un facteur de distribution différent. La valeur du facteur de distribution est égale à 0, si un membre du niveau le plus bas n'est en aucun cas associé au membre (*ms*). Afin de pouvoir traiter les requêtes spatiales de type «*Spatial Slice*», il faut parcourir l'index-NN et trouver pour chaque membre spatial *ms* de *LS*, les facteurs de distribution (*fd*). Par la suite, il faut pondérer les valeurs de mesures (*m*) par rapport aux membres spatiaux du *LD* et *ms*. Les membres de *LD* utilisé par la requête doivent avoir un facteur de distribution différent de 0. Cela permet de récupérer seulement les tuples de la table de faits dont les membres de *LD* sont associés à *ms*.

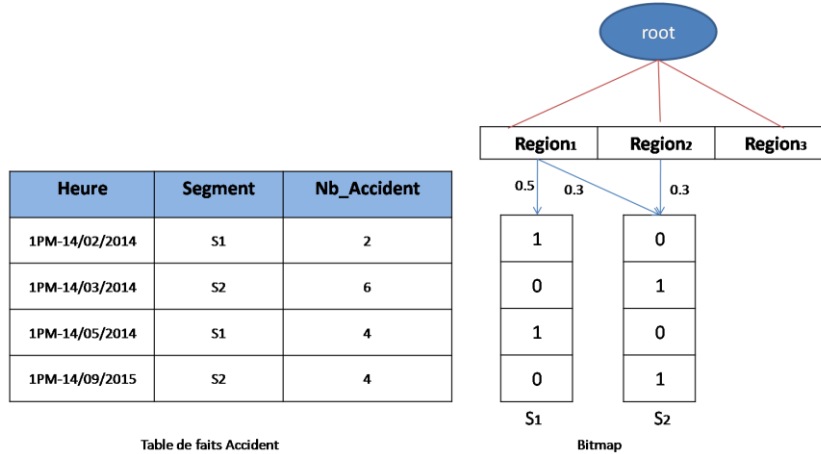
Afin de pouvoir utiliser l'index-NN proposé, une réécriture de la requête «*Spatial Slice*» est indispensable. Dans notre exemple, la requête SQL de la figure 10a est traduite par la requête de la figure 10b de la façon suivante. L'utilisation de l'index-NN permet d'éviter les jointures entre la table de faits et le niveau LD (lignes 2-5 de la Figure 10a) sont remplacées par le IN de la ligne 7 de la Figure 10b), et l'utilisation des facteurs de distribution dans la fonction d'agrégation (ligne 1 Figure 10a sera

SAGEO'2016 – Nice, 6-9 décembre 2016

remplacée par le CASE des lignes 2-7 de la Figure 10b). En d'autres termes, l'idée principale est d'utiliser seulement la table de faits (éviter donc les nombreuses jointures – Figure 4a) pour calculer rapidement l'agrégation avec l'utilisation du bitmap. Pour ce faire, la partie haute de l'index-NN nous permet de sélectionner les membres spatiaux du niveau le plus détaillé (ex. Segment) (ligne 7 Figure 10b) directement de la table de faits, et agréger leur mesures avec les facteurs de distribution issus de l'index-NN (lignes 2-5 Figure 10b).



(a)



(b)

FIGURE 9. a) Structure de l'index-NN, b) Instance de l'index-NN

L'implémentation de l'index et la réécriture des requêtes est montré dans la section suivante.

Conception logique et physique des hiérarchies spatiales non strictes dans les EDSs

```

1 SELECT SUM (Nb_accident * SR.distribution_factor* RR.distribution_factor), Re.nom
2 FROM Accident I, Segment S, Segment_Route SR, Route R, Route_Région RR, Région Re
3 WHERE
4   I.ID_segment= S.ID_segment   AND   S.ID_segment= SR.ID_segment   AND
5   R.ID_route= RR.ID_route       AND   RR.ID_région = Re.ID_région
6 AND Re.nom = 'Auvergne' GROUP BY Re.nom

```

(a)

```

1 SELECT sum (Nb_accident)*
2 CASE
3   WHEN Accident.ID_segment=1   THEN 0.3
4   WHEN Accident.ID_segment =2   THEN 0.2
5   else 0.1
6 END
7 FROM Accident WHERE Accident.ID_segment IN ( 1,2,3,4,5,6,7,8,9,12,15)

```

(b)

FIGURE10. Requête Spatial Slice : a) originale, b) en utilisant le Index-NN.

5. Implémentation et test d'évaluation

Nous avons implémenté nos solutions dans le SGBD *Oracle 12c Entreprise Edition* avec une machine i5 et une RAM de 4Go. Pour l'implémentation de l'index-NN nous avons utilisé l'index Bitmap d'Oracle avec une table pour le stockage de la partie supérieur de notre index composée de trois colonnes qui représentent respectivement : le niveau spatial LS (ex. *Région*), la partie CASE de la requête (i.e. les facteurs de distribution *fd*), et la partie IN de la requête (i.e. les membres *md*). Un exemple de cette table est présenté dans la table ci-dessous. Pour la réécriture de la requête, il s'agit donc simplement de lire cette table.

Region	CASE	IN
R1	WHEN Accident.ID_segment=1 THEN 0.3 Accident.ID_segment =2 THEN 0.2 else 0.1	WHEN 1,2,3,4,5,6,7,8,9,12,15
R2	WHEN Accident.ID_segment=1 THEN 0.4 Accident.ID_segment =2 THEN 0.6 else 0.1	WHEN 1,2,3,4,5,7,8 ,15, 17

Nous présentons, dans la section suivante, une étude de performance de notre approche d'indexation. Cependant, vue l'absence d'un benchmark pour les hiérarchies spatiales non strictes (le seul existant pour les EDS gère seulement les hiérarchies spatiales strictes (Siqueira et al, 2008)), nous avons utilisé pour notre étude, un EDS dont le schéma est présenté par la figure 1. Cet EDS correspond à notre modèle logique proposé dans la section 3. Nous avons évalué l'index-NN selon quatre facteurs : (1) le nombre de membres des dimensions non-spatiales, (2) le nombre de membres de la table de faits (3) le nombre de membres du niveau spatial le plus détaillé ainsi que le nombre de membres du niveau spatial le moins détaillé et (4) le nombre de membres de la Bridge table.

SAGEO'2016 – Nice, 6-9 décembre 2016

Plusieurs tests ont été effectués où nous avons comparé, dans chacun d'entre eux, le temps d'exécution des requêtes sans et avec l'index-NN, tout en vidant les buffers à chaque fois pour éviter l'utilisation des données temporaires. Les figures 11 et 12, montrent les résultats obtenus. Le premier test consiste à montrer le rapport entre le temps d'exécution des requêtes et la taille de la dimension non spatiale *Temps*. Pour cela, nous avons fixé le nombre de tuples de la table de faits à 78 801 172 tuples, et nous avons fait varier la taille de la dimension temporelle *Temps*. Les résultats obtenus sont représentés dans la figure 12a. Nous observons que le temps d'exécution des requêtes est nettement réduit avec l'utilisation de l'Index-NN. Cela est dû au fait que l'index proposé permet d'optimiser l'accès à la table de faits. Cependant, lorsqu'il s'agit de requêtes invoquant plusieurs dimensions non spatiales, il serait intéressant d'utiliser d'autres types d'indexation sur ces dimensions afin d'obtenir des résultats plus intéressants.

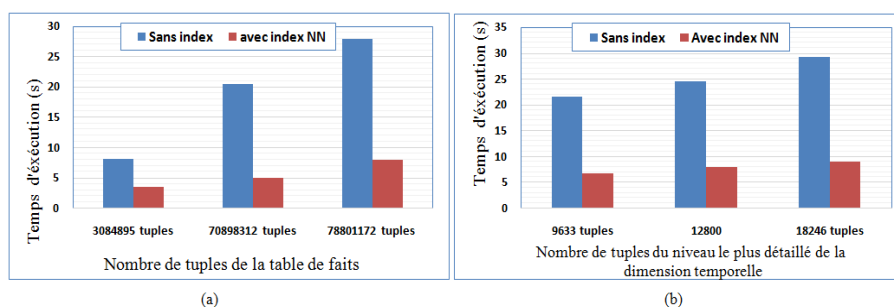


FIGURE 11. Tests de performance par rapport à la taille de : a) la dimension non spatiale, b) la table de faits.

Le deuxième test consiste à mettre en évidence l'effet de l'utilisation de l'index-NN sur la taille de la table de faits. Pour cela nous avons fait varier le nombre de tuples de la table de faits. Les résultats obtenus sont présentés dans la figure 12b. En effet, plus la taille de la table de faits augmente, plus le gain en temps d'exécution devient de plus en plus intéressant. Le test suivant consiste à montrer l'impact du nombre des membres des niveaux spatiaux sur les performances. Dans ce cas, nous avons fixé la taille de la table de faits à 78 801 172 tuples et nous avons fait varier la taille du niveau spatial le plus détaillé ainsi que le niveau spatial le moins détaillé (*Segment* et *Région*). Les résultats obtenus sont montrés dans la figure 12c-1 et la figure 12c-2. Le temps d'exécution des requêtes de type Spatial Slice dépend aussi de la taille des niveaux spatiaux. En fait, plus le nombre de tuples d'un niveau spatial augmente plus l'utilisation de l'index-NN est efficace, car ce dernier permet d'éviter les jointures existantes entre les différents niveaux de la dimension spatiale. Enfin, le dernier test consiste à faire varier la taille de la bridge table. Dans ce cas, nous avons fixé la taille de la table de faits à 78 801 172 tuples, la taille de la table *Segment* à 100 000 tuples et celle de la table *Région* à 500 000 tuples. La figure 12c-3 montre que plus la taille de la bridge table est élevée plus l'utilisation de l'index non strict

devient intéressante. En effet, l'index proposé permet d'éviter non seulement les jointures entre les différents niveaux de la dimension spatiale, mais aussi la jointure entre ces niveaux spatiaux et la bridge table. Plus sont petites les cardinalités de ces tables et plus sont réduits les temps d'exécution.

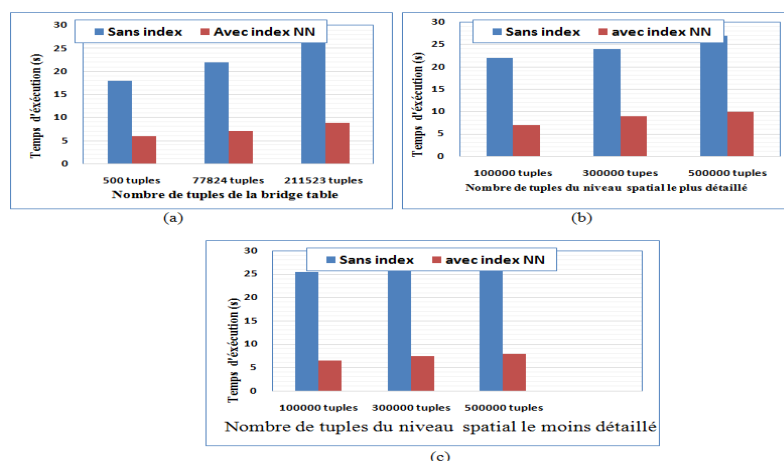


FIGURE12. Tests de performance par rapport à la taille de la dimension spatiale

En conclusion, nous pouvons dire que notre proposition améliore, par rapport aux quatre différents facteurs, les performances des requêtes Spatial slice dans les hiérarchies spatiales non strictes.

6. Etat de l'art

Dans cette section, nous présentons les travaux existants concernant les modèles logiques des EDSs et les index.

Plusieurs travaux existants ont étudié la modélisation conceptuelle et logique pour des EDSs. Ainsi, plusieurs modèles logiques ont été proposés dans ce contexte (Malinowski et Zimanyi, 2008) (Bédard et Han, 2001) (Cuzzocrea et al, 2012) (Siqueira et al, 2008). Certains travaux s'intéressent aux mesures spatiales complexes (Bimonte et al, 2010) (Stefanovic et al, 2000), ou aux données spatiales complexes comme par exemple les données raster (Gomez et al, 2010), les trajectoires (Vaisman et al, 2013) etc. A notre connaissance, seuls (Malinowski et Zimanyi, 2008) proposent un modèle logique pour la prise en charge des hiérarchies spatiales non strictes dans les EDSs. Il s'agit d'ajouter une dimension avec le niveau de la relation n-n. Cependant, ce modèle contient une seule relation n-n dans la hiérarchie spatiale avec les limites décrites dans la Sec 2.1. Pour pallier ces limites, (Jensen et al, 2004) définissent un modèle multidimensionnel formel avec le concept de «partial containment» qui correspond au facteur de distribution de notre approche. Cependant, ils ne présentent pas le modèle logique utilisé pour

SAGEO'2016 – Nice, 6-9 décembre 2016

l'implémentation de leur modèle complexe, ainsi que la manière dont les opérateurs de l'algèbre définis sont implémentés en relationnel.

Plusieurs techniques d'optimisation ont été proposées, vu la taille importante de données spatiales dans les EDSs et la complexité des requêtes spatiales qu'ils exploitent. Nous nous sommes intéressés dans cet article à l'indexation spatiale. Les premiers travaux apparus considèrent un seul niveau dans la dimension spatiale telle que *R-tree* (Guttman 1984), *R*-tree* (Beckmann et al. 1990), etc. L'index *aR-tree* (Papadias et al. 2001) est basé sur des hiérarchies spatiales qui ne sont pas connues au préalable. Les index *SB-index* et *HSB-index* (Siqueira et al. 2011) sont basés sur les hiérarchies spatiales prédéfinies. Ces index combinent la stratégie de filtrage des objets spatiaux avec un index bitmap de jointure pour accélérer le traitement des requêtes spatiales. Dans le contexte des EDs classiques, (O'Neil et Graefe, 1995) propose un index bitmap de jointure définis sur des jointures précalculées entre les hiérarchies. Ainsi, un index binaire hiérarchisé, appelée HOBI (Hierarchically Organized Bitmap Index for Indexing Dimensional Data) est proposé (Chmiel et al, 2009). Il vise à optimiser les requêtes de jointure définies sur des hiérarchies importantes des tables de dimension. Il est basé sur l'utilisation de plusieurs index binaires hiérarchiquement organisés. Cependant, cet index ne considère que les hiérarchies strictes. A notre connaissance, aucune méthode d'indexation pour les EDSs n'a été proposée pour les opérations d'agrégation pour les hiérarchies spatiales non strictes.

7. Conclusion

Dans cet article, nous nous sommes intéressés à la modélisation logique et physique des hiérarchies spatiales non strictes dans les EDSs. En particulier, nous avons proposé une solution composée d'un nouveau schéma logique, et un nouvel index appelé index-NN afin de permettre un calcul correct des agrégations définies sur les hiérarchies spatiales non strictes définies sur plusieurs niveaux. Le modèle logique et l'index proposés permettant de rapprocher les tables des niveaux spatiaux de la dimension de la table des faits en pré-calculant des jointures intermédiaires. Afin de montrer l'efficacité de la solution proposée, nous avons comparé le temps d'exécution de ces requêtes sans et avec l'utilisation de l'index proposé. Nos travaux futurs sont : l'extension de l'index-NN pour supporter d'autres opérateurs spatiaux, la définition d'un benchmark adapté aux hiérarchies spatiales non strictes, et l'étude faite dans cet article sur les solutions NoSQL.

Bibliographie

- Beckmann, N., H.-P. Kriegel, et al. (1990). "The R*-tree: an efficient and robust access method for points and rectangles." SIGMOD Rec. 19(2): 322-331.
- Bédard Y., and Han J. (2001). Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery, Geographic Data Mining and Knowledge Discovery, Taylor & Francis.

Conception logique et physique des hiérarchies spatiales non strictes dans les EDSs

- Bimonte S., Miquel M. (2010). When Spatial Analysis Meets OLAP: Multidimensional Model and Operators. *IJDWM* 6(4): 33-60
- Chmiel J., Morzy T., and Wrembel R.(2009). Hobi : Hierarchically organized bitmap index for indexing dimensional data. In 11th International Conference on Data Warehousing and Knowledge Discovery (DaWaK), pp. 87–98.
- Gomez L., Vaisman A. and Zimanyi E. (2010). Physical Design and Implementation of Spatial Data Warehouses Supporting Continuous Fields. 12th international conference on Data warehousing and knowledge discovery, Springer-Verlag, Berlin, Heidelberg, p. 25-39
- Guttman, A. (1984). R-trees: a dynamic index structure for spatial searching. Proceedings of the 1984 ACM SIGMOD international conference on Management of data. Boston, Massachusetts, ACM: 47-57.
- Jensen C. S., Kligys A., Pedersen T. B. and Timko I. (2004). Multidimensional data modeling for location-based services. *Vldb Journal* 13(1): 1-21.
- Kimball R. (1996). The data warehouse toolkit: Practical techniques for building dimension data warehouse. *Library Journal* 121(13): 102-102.
- Kitchin R. and Lauriault T. P. (2015). Small data in the era of big data. *Geojournal* 80(4): 463-475.
- Lechtenborger J. and Vossen G. (2003). Multidimensional normal forms for data warehouse design. *Information Systems* 28(5): 415-434.
- Lee J. G. and Kang M. (2015). Geospatial Big Data: Challenges and Opportunities. *Big Data Research* 2(2): 74-81.
- Malinowski, E. and Zimanyi E. (2008). *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications*, Springer.
- Mansmann, S. and Scholl M. H. (2006). Extending visual OLAP for handling irregular dimensional hierarchies. *Data Warehousing and Knowledge Discovery, Proceedings* 4081: 95-105.
- O'Neil P., Graefe G. (1995). Multi-Table Joins Through Bitmapped Join Indices, *SIGMOD Record* 24(3), 8-11, 1995.
- Papadias D., Kalnis P., Zhang J. and Tao Y. (2001). Efficient OLAP operations in spatial data warehouses. *Advances in Spatial and Temporal Databases, Proceedings* 2121: 443-459.
- Stefanovic N., Han J. and Koperski K. (2000). Object-based selective materialization for efficient implementation of spatial data cubes. *IEEE Transactions on Knowledge and Data Engineering* 12(6): 938-958.
- Siqueira TLL., Ciferri RR., Times VC., Ciferri C d D A. (2008): Investigating the Effects of Spatial Data Redundancy in Query Performance over Geographical Data Warehouses. *GeoInfo* 2008: 1-12
- Siqueira TLL, Ciferri C. d. D. A., et al. (2011). "The SB-index and the HSB-index: efficient indices for spatial data warehouses." *GeoInformatica* 16(1): 165-205.
- Vaisman A. A, Zimányi E (2013): Trajectory Data Warehouses. *Mobility Data* 2013: 62-82
- Cuzzocrea A, Fidalgo R. do N., (2012): SDWM: An Enhanced Spatial Data Warehouse Metamodel. *CAiSE Forum* 2012: 32-39

Modélisation des activités viticoles sous contraintes climatiques

Cyril Tissot¹, Céline Le Coq¹, Etienne Neethling², Mathias Rouan¹, Hervé Quenol³

1. UMR LETG/Brest Géomer

2. UMR VINITERA

3. UMR LETG/Rennes COSTEL

RESUME. Cet article est dédié à la simulation de l'impact de la variabilité des conditions climatiques sur la dynamique de la vigne et les stratégies de production viticole. La méthodologie présentée exploite des technologies issues de la géomatique, des modèles orientés objets et des bases de données spatio-temporelles. L'approche se base sur un environnement de modélisation multi-agents capable de simuler la croissance de la vigne et d'intégrer des stratégies de production et des règles d'adaptation de ces stratégies en fonction de l'évolution de ces contraintes.

ABSTRACT. This article is consecrated to the simulation of climatic conditions variability impact's on wine-growing dynamic and wine-making production strategies. Displayed methodology operates GIS technology, agent-based modeling and spatiotemporal databases. The approach rests on an agent-based modeling environment, able to simulate wine-growing and to include production strategies and strategies adaptation rules depending on the tie's evolution.

MOTS CLES : stratégies de production viticoles, changement climatique, modèle multi-agents, adaptation, dynamique de la vigne, scénarios.

KEYWORDS: wine-making production strategies, global warming, agent-based modeling, adaptation, wine-growing dynamic, scenarios

1. Introduction

Si la communauté scientifique s'interroge massivement sur les effets du changement climatique à l'échelle planétaire, les initiatives de recherche abordant les conséquences des variations climatiques aux échelles fines restent marginales. Dans le contexte de la viticulture cette question est cruciale dans la mesure où la croissance de la vigne, le rendement et la qualité des baies récoltées sont fortement liés aux conditions et contraintes environnementales locales (Jackson et Lombard, 1993 ; Tesic *et al.*, 2002 ; Van Leeuwen *et al.*, 2004 ; Carey *et al.*, 2008). Les variations spatiales et temporelles du climat jouent un rôle majeur sur le rythme saisonnier de la phénologie de la vigne et le niveau de maturité des raisins (Huglin, 1986). Une augmentation significative des températures peut donc entraîner une modification du cycle phénologique et du risque pathogène (Pugliese *et al.*, 2011). Il en résulte une forte demande de la profession viticole qui doit sans cesse s'adapter à des conditions locales modifiées. Fort de ce constat, il apparaît pertinent de développer une démarche de modélisation permettant de simuler l'impact des différents scénarios sur la dynamique de la vigne et les caractéristiques des zones de production.

La démarche proposée vise à lier les observations climatiques et agronomiques menées sur le terrain (Bonney *et al.*, 2009, Neethling *et al.*, 2011) et les questionnements liés à la recherche d'une adaptation optimale de la vigne à la spécificité des terroirs viticoles (Barbeau *et al.*, 2001 ; Asselin *et al.*, 2003). Elle mobilise des technologies relevant de la géomatique et des modèles multi-agents.

L'objectif est double :

- il s'agit d'une part de répondre aux questionnements des professionnels de la viticulture qui doivent adapter leurs méthodes de production à la variabilité des conditions climatiques sans connaître l'horizon temporel de ces évolutions. Le développement de technologies permettant d'aborder dynamiquement et dans une dimension prospective cette problématique représente donc un fort enjeu à l'échelle nationale (maintien de la spécificité des terroirs viticoles) et internationale (identification de zones d'implantation en adéquation avec un cépage et des pratiques viticoles associées) ;

- il s'agit d'autre part de proposer une méthodologie de modélisation capable d'aborder une problématique complexe intégrant des contraintes spatio-temporelles multi-échelles permettant de restituer la croissance de la vigne à partir d'indices bioclimatiques et de déterminer des zones homogènes au plan des caractéristiques agroclimatiques et dans différents contextes climatiques (approche rétrospective et prospective).

A travers une expérience menée dans les Coteaux du Layon (Val de Loire) un premier prototype a été développé dans l'environnement GAMA (GIS & Agent-based Modeling Architecture). Il vise à analyser la variabilité de la croissance de la vigne et à identifier des zones d'aptitudes en fonction de l'évolution des conditions d'environnement. L'ensemble de la méthodologie (de la conceptualisation aux premières simulations) est exposée dans cet article.

SAGEO'2016 – Nice, 6-9 décembre 2016

2. Approche méthodologique

Aborder dynamiquement l'évolution des conditions d'environnement et leurs implications sur la phénologie de la vigne et les caractéristiques des terroirs viticoles nécessite de mettre en relation des informations hétérogènes sur le plan spatial et temporel. Elles concernent principalement des processus biophysiques, des observations agronomiques et des sorties des modèles climatiques. La conduite d'une démarche de modélisation pose dans ce contexte des questions spécifiques relatives à la définition d'une granularité spatio-temporelles adaptée à la problématique viticole.

Plusieurs auteurs ont montré que l'introduction de plusieurs niveaux d'échelles dans un modèle spatio-temporel était une condition préalable indispensable pour simuler des interactions complexes (Pereira *et al.*, 2004 ; Schindler, 2010) et l'un des principaux écueils en matière de formalisation (Pelekis *et al.*, 2005 ; Spaccapietra *et al.*, 2007) et de développement (Gotway et Young, 2002 ; Duboz, 2004). Dans les modèles abordant spécifiquement la problématique viticole cette difficulté cloisonne généralement les approches à un niveau d'échelle donné (Tissot *et al.*, 2014) ou à l'étude d'un processus spécifique (Shanmuganathan *et al.*, 2011).

Pour contourner ces limites, le modèle présenté dans cet article se base sur un environnement de simulation permettant d'intégrer des données spatio-temporelles multi-sources et multi-échelles au sein d'un modèle à base d'agents contraints. Cette approche répond au besoin de formaliser explicitement les relations spatio-temporelles entre des entités spatiales représentant les unités fonctionnelles des zones de production viticoles (unités de terroir, parcelles, zones d'appellation) et la variabilité des conditions climatiques affectant les vignobles.

Pour parvenir à connecter les différents niveaux scalaires, le prototype développé mobilise des procédures d'agrégation et de désagrégation visant à mettre en adéquation les données avec les échelles de simulation et de restitution des résultats, choisis lors des phases d'initialisation et d'expérimentation du modèle. Compte tenu de la diversité des données manipulées (données statistiques, données matricielles et vectorielles, données temporelles) et de l'hétérogénéité des niveaux d'organisation spatiale impliqués, la méthodologie développée est basée sur une approche bottom-up. L'objectif est de relier plusieurs niveaux d'abstraction en se basant sur la résolution la plus élevée des objets. L'ensemble des développements ont été réalisés au sein de la plateforme multi-agents GAMA (Grignard *et al.*, 2013) en raison de sa capacité à intégrer pleinement les données géographiques (continues, grilles, graphes). Cette spécificité permet d'associer chaque agent à une géométrie simple (un point, une ligne, un polygone) ou complexe (composée de plusieurs sous-géométries) et d'intégrer la topologie de la couche d'objets source. GAMA permet ainsi de définir plusieurs niveaux d'objets synchronisés à un environnement continu de référence (Taillandier *et al.*, 2014).

2.1 Elaboration d'une base de connaissance

La démarche de modélisation s'appuie sur une base de connaissance (données des enquêtes, du milieu naturel, etc.) regroupant l'ensemble des informations permettant de décrire les pratiques et d'expliquer la relation entre la vigne et le climat. L'objectif est de produire une description fine du déroulement des activités viticoles en les replaçant dans un contexte de changement climatique.

Dans cette perspective, des enquêtes approfondies et indépendantes ont été réalisées durant deux années auprès des viticulteurs afin de bien comprendre les interactions entre la variabilité climatique et les pratiques culturales (annuelles et pérennes, TABLE 1). Une première enquête semi-dirigée abordait les questions et les thèmes relatifs aux périodes favorables et défavorables de travail, aux techniques et aux machines impliquées et à l'influence des conditions d'environnement sur la période de travail (Barbeau *et al.*, 2014). Les données récoltées ont ainsi permis d'évaluer la sensibilité de 21 pratiques annuelles aux conditions météorologiques.

TABLE 1 : Caractéristiques des enquêtes approfondies et indépendantes réalisées auprès des viticulteurs.

	Enquêtes 1	Enquêtes 2
Objectifs	<i>Etude de la sensibilité des pratiques annuelles aux conditions climatiques</i>	<i>Etude de la capacité d'adaptation des viticulteurs</i>
Méthode	Semi-dirigée	Semi-dirigée
Questions et thèmes abordés	<ul style="list-style-type: none"> • Périodes de travail • Techniques et machines impliquées • Variables climatiques favorables et défavorables 	<ul style="list-style-type: none"> • Comment les pratiques ont évoluées • Facteurs menant aux changements • Conditions climatiques caractérisant « bon » et « mauvais » millésimes • Stratégies d'adaptation

Une seconde enquête a traité plus spécifiquement de la capacité d'adaptation des vigneron. Les vigneron ont été interrogés sur l'évolution des pratiques, les facteurs menant à des changements, les conditions climatiques qui caractérisent les «bons» et les «mauvais» millésimes et les stratégies d'adaptation adoptées au cours de ces millésimes. Ces informations ont permis de caractériser l'organisation et la stratégie de production des viticulteurs en fonction des conditions d'environnement et des contraintes dans la mise en œuvre des pratiques agronomiques à l'échelle de la parcelle. Les données recueillies ont été utilisées pour construire une typologie basées sur des critères fonctionnels (structure d'entreprise, stratégie de vente), de perception et de sensibilité au changement climatique (niveau de sensibilité, temps de réaction, etc.).

SAGEO'2016 – Nice, 6-9 décembre 2016

Ces données d'enquêtes ont été complétées par des informations spécifiques relatives aux caractéristiques pédologiques (texture des sols, profondeur du sol colonisée par le système racinaire, réserve hydrique maximale), topographiques (pentes, orientation et exposition), et phénologiques (observations lors des différents stades de la vigne) des parcelles de la zone d'appellation des coteaux du Layon. L'intégration des données climatiques s'effectue à plusieurs niveaux d'échelles. A l'échelle régionale, les données du réseau météo France sont mobilisées (température, pluviométrie, ETP, vent). A l'échelle plus locale, le modèle s'appuie sur un réseau de mesures spécifiques, combinant station multi-paramètres et capteurs de températures, implantés en fonction des caractéristiques pédologiques, topographiques et phénologiques de la zone d'intérêt.

L'ensemble des données est intégré au sein d'une base de données au format PostgreSQL/PostGIS qui permet d'administrer les informations collectées et d'y associer des couches d'information géographique. Cette configuration offre la possibilité de lier des variables de terrain à des objets spatiaux et de réaliser des requêtes thématiques.

2.2 Construction du modèle multi-agents

Le modèle SEVE (Simulating Environmental impacts on Viticultural Ecosystems) a été conçu dans l'optique de décrire la dynamique de la vigne et le déroulement des activités viticoles, à partir d'un réseau d'agents réactifs contraints par des variables exogènes (contraintes bio-physiques, socio-économiques et réglementaires). Chaque élément du modèle est représenté par un agent autonome capable de réagir à des changements de son environnement intervenant à différentes échelles de temps et d'espace.

2.2.1 Caractéristiques des agents

Trois types d'agents permettent de hiérarchiser les relations entre les activités viticoles et leur environnement (FIGURE 1) :

- l'Agent "INAO" a un rôle de superviseur au sein du modèle (classe "agent superviseur". Il fixe les cahiers des charges des différentes appellations de vins et impose des modes de production spécifiques. Il est en relation directe avec les Agents "Viticulteurs" qui lui transmettent des informations sur la qualité du raisin produit par leurs vignes. En fonction de ces informations l'agent "INAO" peut modifier ce cahier des charges ;

- les Agents "Viticulteurs" ont pour objectif de produire du vin conforme à un cahier des charges spécifiques en fonction de l'appellation visée (classe "agent exploitant"). Cette action implique de cultiver la vigne dans des conditions optimales compte tenu des spécificités agronomiques des parcelles du viticulteur et de sa stratégie d'entreprise ;

- les Agents "Vignes" sont des entités de production du raisin (classe "agent production"). Elles correspondent généralement à une parcelle ou une entité jugée

Modélisation des activités viticoles sous contraintes climatiques

comme homogène au plan des caractéristiques agronomiques (délimitation à partir d'unités de terroir par exemple (Bodin et Morlat, 2006)). Ces agents ont pour rôle de restituer la dynamique de croissance de la vigne en fonction des conditions climatiques. Leur dynamiques peuvent être perturbées par des pathogènes modélisés sous la forme d'agents agent autonomes (par exemple un agent typé "mildiou" réagira à la variabilité de l'humidité et de la température).

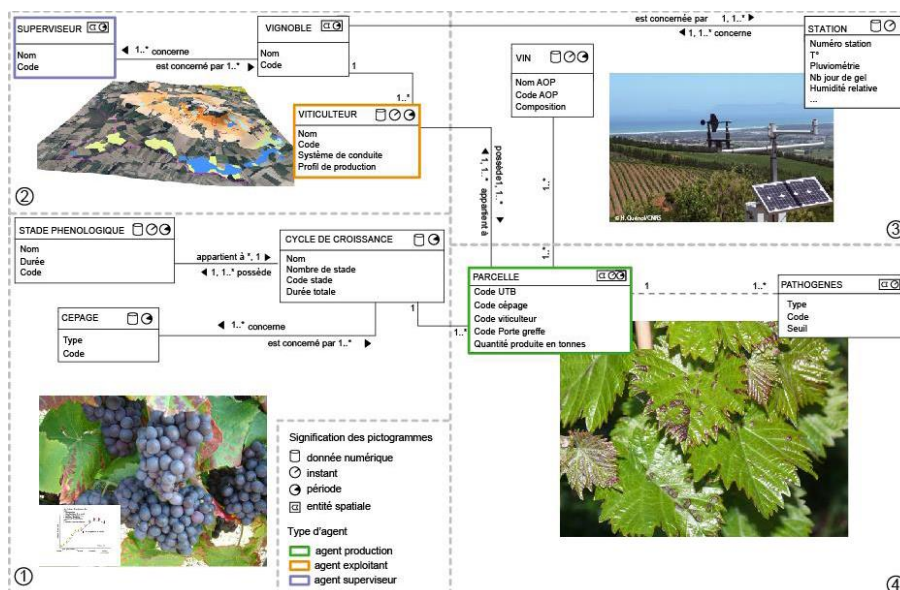


FIGURE 1 : Architecture du modèle SEVE

Les relations entre ces trois types d'agents déterminent les stratégies de production adoptées par les viticulteurs. La classe Agent "Viticulteurs" a donc été enrichie de règles de décisions construites à partir d'une base de règles sous contraintes climatiques. Dans l'optique de disposer d'un prototype permettant d'interagir avec les viticulteurs et de tester des scénarios d'adaptation un premier modèle a été développé sur deux zones d'appellation du val de Loire : les Coteaux du Layon et l'AOC Quart de Chaume (FIGURE 2). Ces deux échelles d'intérêt correspondent à deux niveaux d'agrégation spatiale au sein du modèle :

- un niveau micro (Quart de Chaume) où la modélisation s'effectue à l'échelle de la parcelle. L'objectif est de restituer l'influence de la variabilité climatique à échelle fine sur la phénologie et de calculer des itinéraires agronomiques adaptés à cette variabilité.

- un niveau macro (Layon) où l'entité de référence correspond à un agrégat spatial considéré comme homogène au plan des caractéristiques agroclimatiques. Il peut s'agir d'un terroir spécifique ou d'un ensemble de parcelles ayant des caractéristiques proches (même cépage, même type de sol...). L'objectif est d'évaluer les conditions d'adaptation de zones de production en fonction des scénarios de changements climatiques et des spécificités des zones d'appellation actuelles.

SAGEO'2016 – Nice, 6-9 décembre 2016

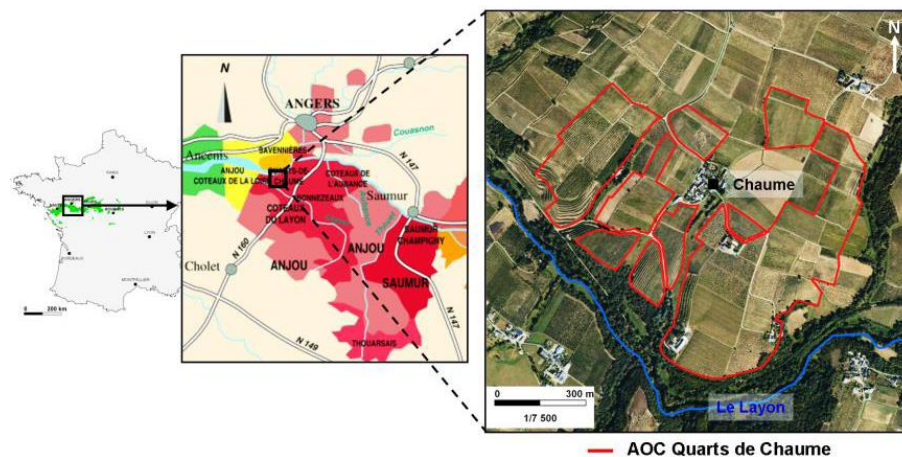


FIGURE 2 : Localisation des AOP Quart de Chaume et Coteaux du Layon.

L'objectif du prototype SEVE est d'intégrer ces deux niveaux au sein d'une même instance et de modéliser les interactions entre ces niveaux. Dans cette perspective, l'implémentation réalisée mobilise les différentes structures d'agents proposées par la plateforme GAMA. La FIGURE 3 illustre la répartition des classes d'agents spatiaux en fonction de leur hiérarchies et des méthodes qui leurs sont associées dans le modèle SEVE. L'espèce "parcelle" possède ainsi une espèce miroir de type "vigne" (association cépage/porte-greffe) qui possède une espèce miroir de type "pathogène" (qui peut rester dormant ou s'activer en fonction de conditions spécifiques). L'espèce "parcelle" est une sous-espèce de l'agrégat "parcelles en zone agroclimatique" qui est une micro-espèce de "zone agroclimatique". Cette structure permet de mettre en relation des entités élémentaires (parcelles) et des groupes (parcelles en zone agroclimatique ou zones agroclimatiques) et de simuler des stratégies individuelles (à l'échelle du viticulteur) ou collectives (à l'échelle d'une coopérative ou de l'INAO par exemple). Elle offre également la possibilité d'intégrer des données à différents niveaux d'agrégation comme par exemple des données climatiques à échelles fines issues des réseaux de mesures actuels et des données régionalisées à 8km de résolution dérivées des modèles du GIEC.

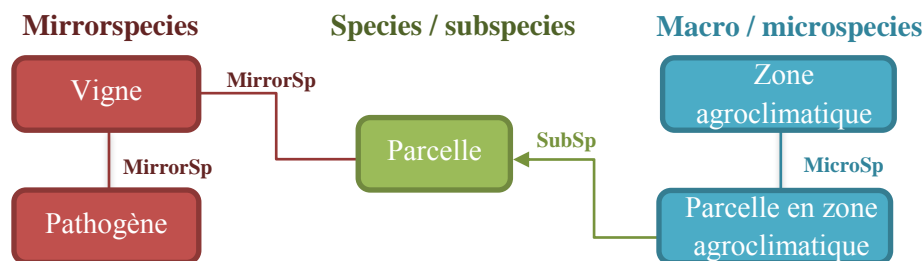


FIGURE 3 : Classes d'agents spatiaux du modèle SEVE sous GAMA

Modélisation des activités viticoles sous contraintes climatiques

Cette structure permet de réaliser des simulations à haute et basse résolution en fonction des données disponibles et de la finalité recherchée (changements à l'échelle de la parcelle ou d'une région de production). L'obtention de données à la parcelle reste une démarche coûteuse qui ne peut s'envisager que sur des espaces restreints. L'expérimentation d'une architecture multi-niveaux s'inscrit également dans la perspective de produire des agrégats à partir de quelques parcelles expérimentales reflétant les différentes configurations agronomiques d'un vignoble.

2.2.2 Méthodes et procédures

La dynamique de croissance de la vigne fait appel à des algorithmes simples permettant de restituer le cycle végétatif à partir d'indices bioclimatiques. Le recours à ces indices vise à étudier l'évolution du climat en relation avec la viticulture notamment en caractérisant les potentialités climatiques d'une région viticole en fonction des types de cépages (Morlat, 2010 ; Tonietto et Carbonneau, 2004 ; Huglin, 1986). Le découpage en stades phénologiques suivant la classification de Baggiolini (Baggiolini M., 1952) est utilisé pour matérialiser cette croissance (FIGURE 4).

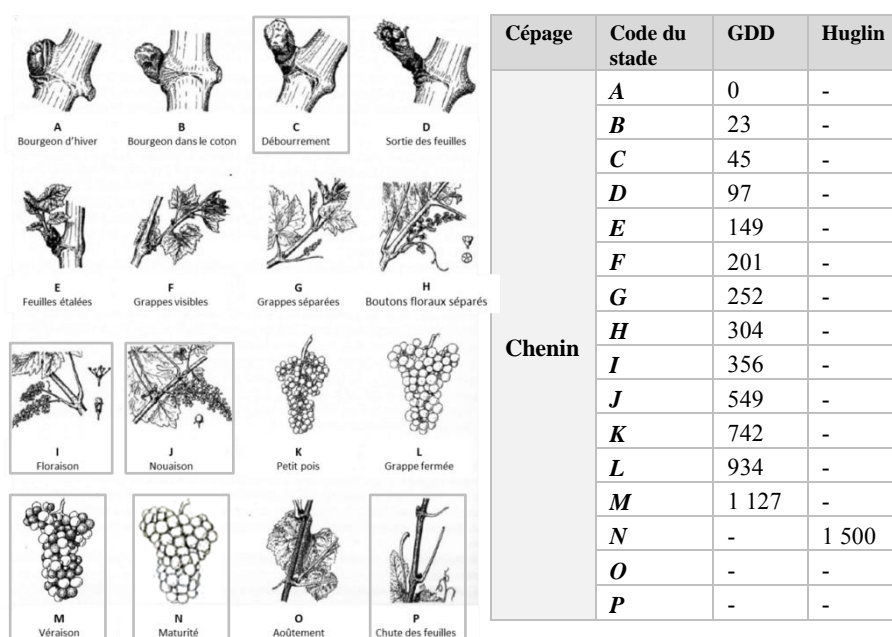


FIGURE 4 : Classification Baggiolini et relations entre stades et indices bioclimatiques en fonction d'un cépage exemple, le Chenin, le plus répandu en Anjou

SAGEO'2016 – Nice, 6-9 décembre 2016

Le passage d'un stade à l'autre est calculé à partir de l'indice de degré-jours (GDD), tel que :

$$GDD = \sum [(T_x + T_n) / 2] - 10^\circ C \quad (1)$$

où 10 °C correspond à la température de base à partir du 1er Janvier (Van Leeuwen *et al.*, 2004), T_x = Température maximale et T_n = Température minimale.

La valeur de cet indice est calculée quotidiennement par les agents-vignes afin de savoir si la vigne est apte, ou non, à passer au stade suivant (FIGURE 5). Les données de températures proviennent de 26 capteurs installés dans la cadre du projet TERVICLIM. Ces capteurs ont été positionnés afin de tenir compte des spécificités de la zone d'appellation Coteaux du Layon et Quart de Chaume (topographie, distance des parcelles aux cours d'eau, morphologie de la végétation...) et de restituer la variabilité des températures à une échelle fine.

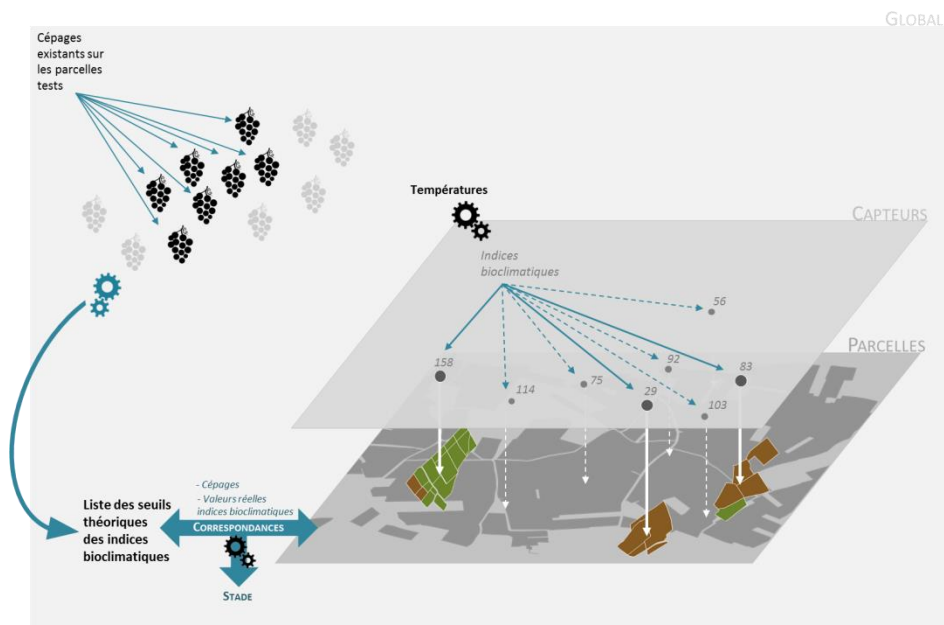


FIGURE 5 : séquence de simulation du cycle phénologique à l'échelle des parcelles

Les données de température permettent également de calculer l'indice héliothermique (IH) de Huglin (Huglin, 1978) qui est utilisé par le modèle pour déterminer la date de vendange potentielle (le lancement de la vendange dépend de la prise de décision du viticulteur).

La valeur de l'Indice de Huglin pour une station correspond à la valeur cumulée de l'indice à la date du 30 septembre, le cumul se faisant depuis le 1er avril (pour

Modélisation des activités viticoles sous contraintes climatiques

l'hémisphère Nord, pour l'hémisphère sud, l'indice est calculé sur la période du 30 septembre au 1er avril) à l'aide de la formule suivante :

$$IH = \sum [(T_m - 10) + (T_x - 10)/2] * k \quad (2)$$

où T_m = Température moyenne, T_x = Température maximale et k le coefficient de longueur du jour variant de 1,02 à 1,06 (en fonction de la latitude de la zone d'intérêt)

L'architecture du modèle vise à formaliser une relation entre le cycle de la vigne, le type de cépage produit et le mode de production utilisé. La variabilité des conditions climatiques conditionne le choix des actions agronomiques qui dépendent également de la stratégie de production du viticulteur. L'ensemble de ces éléments détermine des itinéraires de conduites qui sont ajustés en temps réel dans le modèle. Autrement dit, les actions réalisées à un pas de temps t peuvent avoir une influence sur les choix opérés à $t+n$.

Cette chaîne de relation est difficile à modéliser car elle se réfère à des expériences individuelles. L'enjeu principal du modèle est donc de parvenir à modéliser les chaînes décisionnelles aboutissant aux différents modes de conduite de la vigne observés sur le terrain. Pour déterminer le choix d'une action une vérification des conditions d'environnement (stade de la vigne, conditions météorologiques) est réalisée par l'agent viticulteur. Cette première phase permet de déterminer si le jour considéré est « agronomiquement disponible » pour mener une action sur les parcelles concernées. Par exemple, si la moyenne journalière du vent est supérieure ou égale à 19km/h, alors les viticulteurs ont l'interdiction d'utiliser des pesticides, et doivent donc attendre ou pratiquer des actions de substitution.

Parmi les actions que les viticulteurs vont pouvoir mener, certaines vont jouer directement un rôle sur la conduite de la vigne, d'autres vont faciliter son développement (pratique de l'enherbement, travail du sol). La communication entre les agents vignes et viticulteurs est donc permanente. Afin de tenir compte des différents modes de conduite agronomique, des profils de production ont été associés aux agents viticulteurs (mode conventionnel, raisonné ou biologique). Ils permettent de prendre en compte des stratégies plus spécifiques associées à des orientations de production. Enfin, dans la perspective d'intégrer des contraintes socio-économiques dans le modèle, des profils d'entreprises ont également été implémentés. L'objectif est de pouvoir catégoriser les exploitations viticoles en fonction de leur structure économique (entreprise familiale, grand domaine, multinationale). Cette information n'est que partiellement exploitée à ce stade de développement du modèle et ne se répercute que sur la main d'œuvre potentiellement mobilisable pour mener à bien des actions.

3. Simulations

L'objectif des simulations est de mettre en relation les stades phénologiques de la vigne avec les conditions climatiques et les modes de conduites agronomiques dans un contexte de changement global. Les premiers tests réalisés dans les coteaux du

SAGEO'2016 – Nice, 6-9 décembre 2016

layon visent à évaluer les capacités d'adaptation des viticulteurs pour la production d'un vin liquoreux dans un contexte de réchauffement important des températures (Bonnefoy *et al.*, 2009).

3.1 Déroulement des simulations

La procédure de simulation débute par une phase d'initialisation qui crée l'ensemble de agents manipulés par le modèle à partir des bases de données stockées sur le serveur PostgreSQL/PostGIS. L'agent INAO est initialisé avec les caractéristiques du cahier des charges encadrant l'appellation d'origine contrôlée "Quart de Chaume". Ces éléments de contrôle seront utilisés pour vérifier la conformité de la production fixée par cette AOC. Les agents vignes sont affectés de variables agro-environnementales concernant le type de cépage, le type de sol et ses caractéristiques physiques ainsi que les données météorologiques de la période considérée. Les agents viticulteurs sont créés à partir de trois profils de production : conventionnel (viticulture traditionnelle), raisonnée (limitation des produits phytosanitaires et des engrais) et biologique (cahier des charges spécifique limitant fortement l'usage de produits phytosanitaires). Les enquêtes, menées auprès des viticulteurs permettent d'associer des calendriers de pratique à chaque profil de production. Ces calendriers permettent de construire des arbres de décisions sous contraintes agroclimatiques. Les données de températures et de précipitations sont utilisées pour calculer des jours agronomiquement disponibles pour la réalisation des tâches de travail de la vigne (Barbeau *et al.*, 2014). En fonction de ces données, de son profil de production et des caractéristiques agroclimatiques de ses parcelles (nature du sol, pente, exposition, bilan hydrique...), chaque Agent viticulteur réalise des choix de conduite agronomique (FIGURE 6).

	Nom action	Code stade	Locomotion	Contraintes
1	Broyage des sarments	A	Tracteur	-
2	Fertilisation du sol	B	Tracteur	-
3	Ebourgeonnage et épamprage	F	A pied	Pluie / Température
4	Palissage	H	A pied	Pluie / Température
5	Rognage	J / L	Tracteur	-
6	Effeuilage	J / M	A pied/tracteur	Pluie / Température

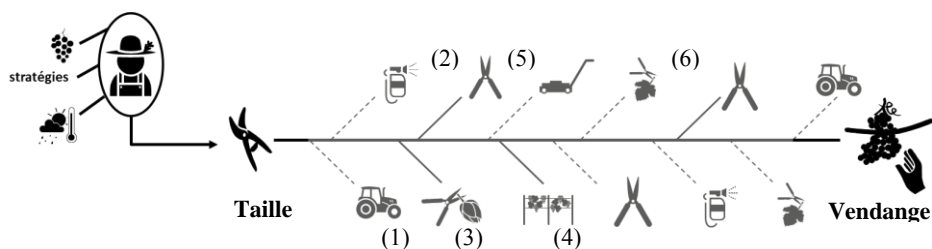


FIGURE 6 : Diagramme d'action associé à l'agent viticulteur

Modélisation des activités viticoles sous contraintes climatiques

Ces choix permettent de construire des itinéraires de conduites agronomiques qui peuvent intégrer des actions classiques du travail de la vigne (taille, rognage, effeuillage...) comme des outils d'adaptations à des contraintes spécifiques (travail du sol, entretien d'un enherbement concurrent...). Chaque itinéraire peut ensuite être analysé pour évaluer son coût (ratio heure/hectare par action) ou le comparer à d'autres itinéraires conduits sur des parcelles ayant des caractéristiques proches mais une stratégie de production différente. L'objectif est d'identifier des panels d'itinéraires agronomiques possibles selon différentes configurations climatiques (années chaudes, fraîches, humides...) et dans une dimension prospective (simulation sur les bases des scénarios du GIEC). Cette étape de modélisation, actuellement en test, nécessite d'être validée en raison de complexité de la chaîne de relations aboutissant à la prise de décision. Dans ce contexte des enquêtes régulières sont menées auprès des viticulteurs afin de connaître les mécanismes orientant leurs choix d'actions agronomiques. Cette phase est cruciale car elle permet de valider une part importante du modèle et de crédibiliser la démarche auprès de la profession viticole ce qui constitue un préalable indispensable à la conduite de simulations participatives et prospectives.

3.2 Simulation à partir d'agrégats spatiaux

Le modèle SEVE a également pour objectif de déterminer des zones agroclimatiques en fonction de leur niveau d'adaptation potentiel au changement climatiques. Les conditions d'adaptation dépendent à la fois de l'aptitude des parcelles à supporter une variation importante des températures au regard des actions agronomiques possibles et des contraintes liées aux cahiers des charges régissant les zones d'appellation.

Les critères de classifications des zones agroclimatiques sont déterminés à l'initialisation du modèle. Ils dépendent de choix utilisateurs en fonction d'objectifs spécifiques (minimisation du nombre d'actions et donc du coût de l'itinéraire agronomiques, conservation des critères d'AOC...) ou de scénarios prédéterminé

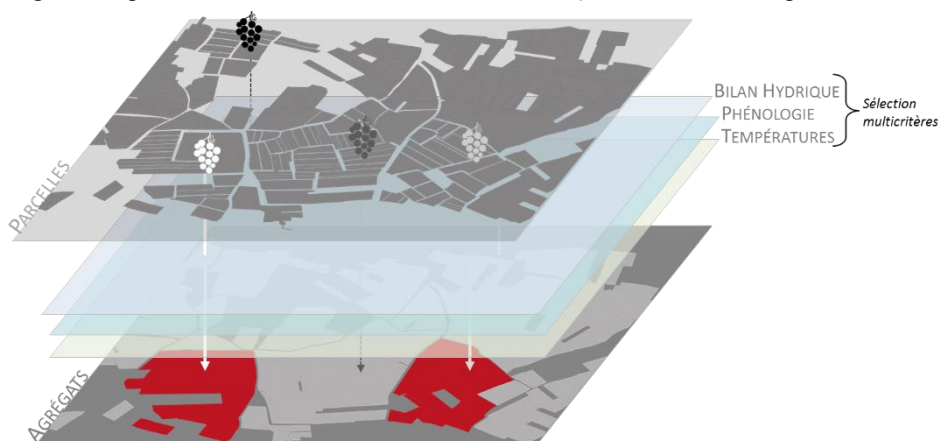


FIGURE 7 : procédure d'agrégation des agents dans le modèle SEVE

SAGEO'2016 – Nice, 6-9 décembre 2016

(conservation d'au moins 50% des parcelles en AOC par exemple). Les filtres d'agrégation sont représentés comme différentes couches par lesquelles transitent les informations générales sur la parcelle testée. Les critères d'agrégation sont combinés et comparés à des valeurs théoriques ou connues. Une fois identifiées, les parcelles sont agrégées pour ne former qu'une entité spatiale de type macrospecies, à un niveau d'échelle supérieure (FIGURE 7).

La temporalité de la fonction d'agrégation est totalement paramétrable, à des dates fixes (dernier jour de l'année agronomique, horizon temporel + 20 ans etc.) ou en fonction d'évènements particuliers (maturité, créneau possible de vendange, etc.).

Par exemple, des parcelles ayant des cycles phénologiques proches et une bonne résistance au stress hydrique peuvent être regroupées au sein d'un agrégat typé comme ayant un potentiel d'adaptation élevé à une augmentation des températures. Cet agrégat étant un agent de type macrospecies il a accès aux caractéristiques des species unitaires qui le composent et peut par exemple produire un itinéraire agronomique moyen ou une série d'itinéraires probables en fonction de la diversité des pratiques menées au niveau unitaire.

Les agents macrospecies obtenus peuvent également interagir avec les agents superviseurs (INAO, Coopératives viticoles) via des fonctions de contrôle de la conformité des zones (vérification des indices bioclimatique du respect des modes de conduites agronomiques...). Les expérimentations en cours visent notamment à déterminer des seuils à partir desquels les cahiers des charges sont trop ou plus assez restrictifs au regard des implications du changement climatique sur la phénologie et les actions agronomiques.

4. Conclusion et perspectives

La démarche de modélisation présentée propose une approche globale visant à simuler le déroulement des activités viticoles sous contraintes d'environnement. Abordée via le développement d'un modèle multi-agents chargé de modéliser les interactions entre les différentes composantes de la viticulture, la problématique de l'impact de la variabilité des conditions climatiques sur les stratégies de production viticole a été plus spécifiquement ciblée. La finalité recherchée était double : il s'agissait d'une part de restituer la dynamique de la vigne à partir d'indices bioclimatiques et d'autre part de simuler les itinéraires de conduites agronomiques dans un contexte de changement climatique.

Sur le plan méthodologique, les développements réalisés ont contribué à formaliser une relation cohérente entre un réseau d'agents réactifs et un territoire viticole modélisée à deux niveaux d'échelles à partir d'une série de contraintes techniques, socio-économiques, réglementaires et environnementales. Le couplage entre base de données spatio-temporelles et modèle multi-agents s'inscrit dans cette perspective et vise notamment à améliorer l'intégration de données spatio-temporelles composites comme variables de forçage afin de restituer la variabilité du déroulement des activités viticoles. Sur le plan thématique, l'analyse des implications du

Modélisation des activités viticoles sous contraintes climatiques

changement climatiques sur les itinéraires de conduites agronomiques et les stratégies de production représente un enjeu important pour la profession viticole qui doit s'adapter aux modifications des conditions d'environnement tout en conservant la typicité des terroirs et les caractéristiques des vins d'appellation.

Les premiers résultats obtenus même s'ils restent partiels montre que le modèle SEVE est en mesure de restituer la dynamique de la vigne et de simuler des itinéraires de conduites agronomiques en lien avec la variabilité des conditions climatiques. Cette étape est cruciale car elle constitue un préalable indispensable à l'analyse des stratégies de conduite de la vigne dans un contexte perturbé. Ces outils sont une aide au contrôle du bon déroulement de la simulation et un support de communication vers la profession viticole. Ils restent néanmoins perfectibles car ils proposent une vision instantanée des sorties de calcul du modèle et nécessitent d'être complétés de nombreuses analyses à postériori pour expliciter les résultats obtenus.

Bibliographie

- Asselin C., Barbeau G., Morlat R., (2003). Les terroirs viticoles : du concept au produit. *Terroir, Zonazione, Viticoltura. Trattado internazionale de Mario Fregoni, Danny Schuster et Andrea Paoletti*. Phytoline Editore, p. 159-186.
- Baggiolini M. (1952). Les stades repères dans le développement annuel de la vigne et leur utilisation pratique. *Revue romande d'Agriculture et d'Arboriculture* 8 vol. 8, n°1, p. 4-6.
- Barbeau G., Cadot Y., Stevez L., Bouvet M H., Cosneau M., Asselin C., Mege A., (2001). Role of soil physical properties, climate and harvest period on must composition, wine type and flavour (*Vitis vinifera* L, cv chenin), Coteaux du Layon, France. *Proceedings of the 26th world congress of the OIV, 11-17 october 2001, Adelaide, Australia, section viticulture* p. 105-118.
- Barbeau C., Barbeau G., Joannon, A., (2014). Analyzing the sensitivity of viticultural practices to weather variability in a climate change perspective: An application to workable-day modelling. *J. Int. Sci. Vigne Vin*, vol. 48, n° 2, p. 141-152
- Bodin, F., Morlat R., (2006). Characterization of viticultural terroirs using a simple field model based on soil depth. I. Validation of the water supply regime, phenology and vine vigour, in the Anjou vineyard (France). *Plant and Soil*, vol. 281, n° 37, p. 37-54.
- Bonnefoy C., Quénot H., Barbeau G., Madelin M., (2009). Analyse multi-scalaire des températures dans le vignoble du Val de Loire. *Géographia Technica*, Vol. 4, p. 85-90.
- Carey V., Saayman D., Archer E., Barbeau G., Wallace M., (2008). Viticultural terroirs in Stellenbosch, South Africa. I. The identification of natural terroirs units. *J. Int. Sci. Vigne Vin*, vol. 42, n° 4, p. 169-183.
- Duboz R. (2004). *Intégration de modèles hétérogènes pour la modélisation et la simulation de systèmes complexes. Application à la modélisation multi-échelles en écologie marine*. Thèse en informatique, Université de Calais.
- Gotway C.A., Young L.J., (2002). Combining incompatible spatial data, *Journal of the American Statistical Association*, vol. 97, n° 458, p. 632-648.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Grignard, A., Taillandier, P., Gaudou, B., Vo, D. A., Huynh, N. Q., & Drogoul, A., (2013). GAMA 1.6: Advancing the art of complex agent-based modeling and simulation. *Proceedings of the 16th international conference on Principles and Practice of Multi-Agent Systems, PRIMA 2013*, p. 117-131. Springer Berlin Heidelberg, Dunedin, New-Zeland.
- Huglin P. (1978). Nouveau mode d'évaluation des possibilités héliothermiques d'un milieu viticole, *C. R. Aca. Agric.*, p. 1117-1126.
- Huglin P. (1986), *Biologie et écologie de la vigne*, Éditions Payot - Tec et Doc, Lausanne - Paris.
- Jackson D.I., Lombard, P.B., (1993). Environmental and management Practices affecting grape composition and wine quality, a review. *Am J Enol Vitic.*, vol. 44, p. 409-430.
- Morlat R. (2010). *Traité de Viticulture de Terroir*, Lavoisier, Paris
- Neethling E., Barbeau G., Quenol H., Bonnefoy C., (2011). Évolution du climat et de la composition des raisins des principaux cépages cultivés dans le val de Loire, *Climatologie*, vol. 8, p 79-90.
- Pelekis N., Theodoulidis B, Kopanakis I., Theodoridis Y., (2005). Literature Review of Spatio-Temporal Database Models, *The Knowledge Engineering Review Journal*, vol. 19, n° 3, p. 235-274.
- Pereira A., Duarte P., Reis L., (2004). Agent-Based Simulation of Ecological Models. *Proceedings 5th Workshop on Agent-Based Simulation*, Lisbon, Portugal.
- Pugliese M, Gullino ML, Garibaldi A., (2011). Effect of climate change on infection of grapevine by downy and powdery mildew under controlled environment. *Commun Agric Appl Biol Sci.*,76:579-82.
- Schindler J. (2010). A Multi-Agent System for Simulating Land-Use and Land-Cover Change in the Atankwidi Catchment in Upper East Ghana, Dissertation. *Ecology and Development Series*, n° 68.
- Shanmuganathan S., Narayanan A., Robinson NJ L., (2011). A multi-agent cellular automaton for grapevine growth and crop simulation, *International Journal of Machine Learning and Computing*, vol. 1, n° 3, p. 291-296.
- Spaccapietra S., Parent C., and Zimányi E., (2007). Spatio-Temporal and Multi-Representation Modeling: A Contribution to Active Conceptual Modeling, *Active Conceptual Modeling for Learning*, vol. 4512, p. 194-205.
- Taillandier P., Grignard A., Gaudou B., Drogoul A. (2014). Des données géographiques à la simulation à base d'agents: application de la plate-forme GAMA. *Cybergeo: European Journal of Geography*.
- Tesic, D., Woolley, D.J., Hewett, E.W., Martin, D.J., (2002). Environmental effect on cv Cabernet Sauvignon (*Vitis Vinifera* L.) grown in Hawkes Bay, New Zealand, 1. Phenology and characterization of viticultural environments. *Aust. J. Grape Wine Res.*, vol.8, n° 1, p. 15-26.
- Tissot C., Rouan M., Neethling E., Quenol H., Brosset D., (2014). Modeling of vine agronomic practices in the context of climate change, *Proceedings of the 37th World Congress of Vine and Wine, "Southern Vitiviniculture, a Confluence of Knowledge and Nature 2014*, Mendoza, Argentina.

Modélisation des activités viticoles sous contraintes climatiques

- Tonietto J., Carbonneau A., (2004). A multicriteria climatic classification system for grape-growing regions worldwide. *Agricultural and Forest Meteorology*, vol. 124, n° 1-2, p. 81-97.
- Van Leeuwen C., Friant P., Chone X., Tregoat O., Koundouras S., Dubourdieu D., (2004). The influence of climate, soil and cultivar on terroir. *Am. J. Enol. Vitic.*, vol. 55, n°3, p. 207-217.

Un cadre d'analyse des Infrastructures de Données Géographiques pour interroger la mise en réseaux des acteurs et des outils

**M. Noucher¹, F. Gourmelon², A. Laumond³, G. Melançon³, B. Pinault³,
A. Maulpoix², J. Pierson¹, O. Pissot¹, M. Rouan²**

1. CNRS, UMR 5319 PASSAGES, site Bordeaux Universités
12 esplanade des Antilles – F-33607 Pessac Cedex
2. CNRS, UMR 6554 LETG-Brest
IUEM, rue Dumont d'Urville – 29280 Plouzané Cedex
3. Université de Bordeaux, UMR 5800 LaBRI
351 Cours de la Libération – 33405 Talence Cedex

Auteur correspondant : matthieu.noucher@cnrs.fr

RESUME. L'apparition de multiples portails cartographiques engendre une démultiplication de l'offre en données géographiques sur Internet. Notre étude propose une méthodologie mixte appliquée à 45 IDG françaises afin de comparer les objectifs affichés par leurs promoteurs et leurs contenus et services effectifs. Fondée sur l'analyse d'entretiens, de questionnaires, des sites web et des métadonnées, la méthodologie produit des résultats variés permettant de revisiter l'idée d'une mise en réseaux généralisée des outils et des acteurs. Il apparaît en effet qu'en dépit d'un discours volontariste, la mise en réseaux des outils reste confinée à des thématiques spécifiques. Par ailleurs, si la géocollaboration organisée par les IDG est très active, la mise en réseau des acteurs ne concerne que le secteur public.

ABSTRACT. The development of multiple webmapping portals generates a gear in the supply of spatial data on the Internet. Our study offers a mixed methodology applied to 45 French SDIs to compare the targets set by their promoters and their actual contents and services. Based on the analysis of interviews, survey, web sites and metadatas, methodology produces various results to visit in a different way the idea of a widespread networking tools and actors. Despite a proactive speech, the networking tools remains confined to specific themes. If geocollaboration organized by SDIs is very active, communities of practice only affects the public sector.

MOTS-CLES : Infrastructure de Données Géographiques (IDG) ; métadonnées ; interopérabilité ; géocollaboration ; analyse réseau ; France.

KEYWORDS: Spatial data infrastructure (SDI) ; metadata ; interoperability ; geocollaboration ; network analysis ; France.

1. Introduction

Par le double effet de l'interopérabilité des systèmes et des évolutions du cadre réglementaire, les données géographiques institutionnelles sont de plus en plus accessibles à travers le développement croissant d'Infrastructures de Données Géographiques (Crompvoets *et al.*, 2004). Les IDG sont ainsi devenues des plateformes web essentielles pour la diffusion des données institutionnelles (Rajabifard *et al.*, 2003). Elles visent à faciliter, à tous les échelons territoriaux, la découverte, l'accès, l'échange et le partage d'information et de services géographiques (Masser, 2010). Souvent réduites à de simples sites web permettant de consulter les données cartographiques d'un territoire à travers un géoportail et de les télécharger via un géocatalogue, les IDG sont, en fait, de véritables dispositifs sociotechniques où se joue une partie des évolutions majeures affectant le régime actuel de production et de circulation de l'information géographique.

Bien que les publications sur les IDG soient en augmentation constante depuis une décennie, Hendricks *et al.* (2012) ont montré que la plupart traitent de leurs composantes techniques, ignorant souvent les dimensions organisationnelles. Coetzee et Wolff-Piggott soulignent eux-aussi dans une revue de littérature récente (2015) et significative¹ que la plupart des recherches sont focalisées sur des développements technologiques et des preuves de concept de leurs implémentations. Considérées comme des outils, les IDG font l'objet d'analyses techniques, sur l'interopérabilité des systèmes ou la normalisation des données. Les questions organisationnelles sont parfois abordées, notamment pour évaluer la reproductibilité des systèmes de gouvernance d'un pays à l'autre (Georgiadou *et al.*, 2006).

Au niveau institutionnel, les IDG sont évaluées sur la base de critères essentiellement quantitatifs (nombre de métadonnées, nombre de services Web de consultation, nombre de téléchargements et de sources de données, etc.). Aucune analyse sur le contenu du patrimoine de données n'est généralement opérée. Par exemple, la décision de la Commission Européenne (2009/442/CE) du 5 juin 2009 inscrit le rapportage de la Directive INSPIRE avant tout dans une logique de *benchmarking*² visant à créer un référentiel cognitif commun destiné à rendre les données comparables d'un pays à l'autre, sans porter de réflexion de fond sur les types de contenus à créer ni sur leur pertinence.

Plusieurs verrous scientifiques expliquent pourquoi les études en cours sont si spécialisées. En effet, il est difficile (a) de disposer d'un panorama exhaustif de l'ensemble des données et des flux associés aux différentes plateformes décentralisées, (b) d'identifier les usagers des IDG, qu'ils soient producteurs ou consommateurs de données géographiques et (c) de décrire et qualifier les usages

¹ Coetzee & Wolff-Piggott ont analysé 237 articles dans leur revue de littérature.

² Le *benchmarking* est une technique issue du marketing pour comparer les performances des entreprises. Un *benchmark* est un indicateur chiffré de performance dans un domaine donné tiré de l'observation des résultats de l'entreprise qui a réussi le mieux dans ce domaine. Cet indicateur peut servir à définir les objectifs de l'entreprise qui cherche à rivaliser avec elle.

SAGEO'2016 – Nice, 6-9 décembre 2016

dont elles font l'objet. Aujourd'hui le développement des plateformes de mise à disposition d'informations géographiques sur le web contribue à la complexité de l'« univers de l'information géographique » (Ballatore, 2014). Or, la compréhension de son fonctionnement et de son évolution est un enjeu fort pour la société dans la mesure où les acteurs publics et privés à différents échelons, s'interrogent sur l'impact de ces dispositifs en termes de gouvernance et de gestion territoriale. Ces réflexions s'inscrivent pleinement dans les débats actuels menés au sein des *critical data studies* (Kitchin and Lauriault, 2014) qui concernent les contextes de production, d'analyse, de diffusion et d'usages des données, en particulier géographiques, qui circulent entre les infrastructures web.

Notre étude propose un cadre méthodologique d'analyse critique des infrastructures de données géographiques qui entend contribuer au nécessaire « travail de conceptualisation sur la base de travaux empiriques » qu'évoquent Kitchin et Lauriault (2014 : 14) pour étayer et étoffer ce champ de recherche. Une méthode mixte pour capturer des jeux de données variés et volumineux (métadonnées, pages éditoriales des sites web, résultats d'enquête en ligne et entretiens) a été mise en oeuvre pour analyser les infrastructures régionales et nationales de données géographiques françaises. L'objectif est de fournir une vue d'ensemble du partage de données géographiques en l'explorant sous quatre dimensions : *l'accessibilité des données*, *la mise en réseau des outils*, *la mise en réseau des acteurs* et *l'équité informationnelle des territoires* en distinguant la théorie (le discours des promoteurs) de la pratique (la réalité pour les usagers). L'accessibilité des données a pour but d'évaluer la capacité effective des IDG à donner accès à des contenus réutilisables en identifiant les données téléchargeables et correctement documentées permettant leur mise en exploitation dans un SIG. La mise en réseau des outils fait référence aux décloisonnements des systèmes d'information, c'est-à-dire à la capacité des IDG à favoriser les échanges entre systèmes experts pour rompre avec les logiques en silos des années 90 (démultiplication de SIG isolés par métier et par territoire sans interconnexion). La mise en réseau des acteurs sera évaluée par l'identification de dynamiques de géocollaboration. Enfin, nous questionnerons l'un des arguments récurrents de la justification de la mise en oeuvre de ces plateformes, i.e. l'argument d'égalité informationnelle des territoires. Ainsi, en abordant la déconstruction des IDG par le prisme des données, des systèmes d'information, des acteurs et des territoires, nous proposons une analyse critique d'un contexte national pour confronter les discours théoriques aux contenus effectifs.

Cet article est organisé en trois sections. La section 2 détaille le cadre exploratoire, le corpus mobilisé et la méthodologie alors que la section 3 présente les résultats de deux des quatre composantes précédemment citées : la mise en réseau des outils et la mise en réseau des acteurs. Ces résultats sont ensuite discutés et les perspectives de recherche sont présentées en section 4.

2. Corpus et méthode

2.1. Corpus

En Europe, la directive INSPIRE (2007/2/CE)³ vise au déploiement d'une infrastructure de données géographiques dans la Communauté européenne en s'appuyant sur les IDG des États membres. En France, depuis 2005 le Conseil national de l'information géographique (CNIG) assure la coordination entre l'infrastructure nationale dont le déploiement est assuré par deux acteurs historiques de la cartographie française (IGN et BRGM) et les nombreuses infrastructures thématiques (sur le sol, sur l'eau, sur le littoral, etc.) ou infranationales (départementales, régionales, transfrontalières...).

En 2014, l'AFIGEO recensait 65 dispositifs se réclamant des IDG. Ce chiffre est en progression constante depuis 2008. L'échelon régional compte à lui seul la moitié des infrastructures françaises. Relais entre les données locales et les standards nationaux, les IDG régionales ont un rôle d'animation (fédération, formation, coordination) important. Nous avons donc choisi de nous focaliser sur les infrastructures nationales et régionales qui concentrent, à elles-seules, les deux tiers des IDG recensées par l'AFIGEO. Un travail d'actualisation⁴ de l'inventaire de 2014 nous a permis de retenir 45 plateformes : 16 nationales et 29 régionales.

2.2. Méthode

Pour analyser ces 45 IDG sous l'angle de l'accessibilité des données, de la mise en réseau des outils et des acteurs et de l'égalité informationnelle des territoires, nous avons identifié onze questions de recherche (tableau 1) qui s'inscrivent dans un cadre méthodologique mixte (Johnson et Onwuegbuzie, 2004) (figure 1). L'approche qualitative est fondée sur l'analyse des composantes éditoriales des sites web et des *verbatim* d'entretiens menés avec quelques promoteurs d'IDG pour préciser leurs objectifs et stratégies. Les données d'une enquête à l'attention des promoteurs des 45 IDG et le contenu des géocatalogues en métadonnées font l'objet d'une approche quantitative à base de statistiques et d'analyses de réseaux. L'analyse spatiale des données cataloguées complète l'ensemble. Ces approches diversifiées doivent ainsi permettre de confronter les dires d'experts aux données effectivement diffusées et de s'interroger sur l'opérationnalité des objectifs assignés aux IDG.

³ <http://inspire.ec.europa.eu/>

⁴ L'actualisation de l'inventaire s'est faite à partir de l'exploration des sites web (vérification de leur maintien) et de plusieurs entretiens auprès des animateurs du réseau national des IDG. Depuis l'inventaire de 2014, certaines IDG ont disparu, fusionné ou ont été créées ce qui témoigne d'un contexte institutionnel encore instable.

SAGEO'2016 – Nice, 6-9 décembre 2016

TABLEAU 1. Quatre composantes déclinées en onze questions de recherche

Composante du partage de données	Question de recherche	Source
A. Accessibilité des données	A1. Quel est le discours des promoteurs sur la question de l'accessibilité des données ?	Site web et entretiens
	A2. Les données indexées dans les IDG sont-elles téléchargeables ?	Géocatalogue
	A3. Les données diffusées dans les IDG sont-elles facilement réutilisables ?	Géocatalogue
B. Mise en réseau des outils	B1. Quel est le discours des promoteurs sur la question du décloisonnement des SI ?	Site web et entretiens
	B2. Les IDG sont-elles interconnectées pour accélérer les flux de données ?	Enquête
	B3. Les IDG facilitent-elles l'hybridation thématique des données ?	Géocatalogue
C. Mise en réseau des acteurs	C1. Quel est le discours des promoteurs sur la question des dynamiques de géocollaboration ?	Site web et entretiens
	C2. Les IDG favorisent-elles des dynamiques de géocollaboration ?	Enquête
	C3. Les IDG permettent-elles la mise en réseau d'acteurs diversifiés ?	Géocatalogue
D. Équité informationnelle	D1. Quel est le discours des promoteurs sur la question de l'égalité informationnelle ?	Site web et entretiens
	D2. Les IDG couvrent-elles de manière homogène leurs territoires ?	Géocatalogue

2.2.1. L'acquisition de quatre types de données.

Les sites web

Les parties éditoriales des sites web des 45 IDG du corpus ont été explorées. Les pages relatives à la présentation de la démarche ou à ses objectifs ont été étudiées pour relever les argumentaires qui justifient le déploiement de ces plateformes.

L'enquête web

L'enquête a été lancée en mai 2015 pour une durée de cinq mois. Réalisée en ligne à partir du questionnaire d'enquête LimeSurvey, elle était destinée aux promoteurs des IDG françaises. Une série de questions fermées a permis de dresser un état des lieux des services proposés (géoportail, géocatalogue, services web, etc.) et des actions d'animation (formation, communication, coproduction, etc.) associées à chaque plateforme. Des questions ouvertes ont également invité les participants à

Cadre d'analyse des IDG

préciser leurs objectifs et leurs cibles. Les promoteurs de 44 des 45 IDG de notre corpus ont répondu à cette enquête.

Les entretiens.

Une série d'entretiens a été menée auprès de 12 promoteurs d'IDG nationales et régionales⁵ afin de préciser leur stratégie en matière d'objectifs, de cibles et de perspectives. A partir d'une trame semi-directive, des questions relatives aux modes de gouvernance ou aux actions d'animation et de coproduction ont également été posées pour obtenir des informations inaccessibles par l'analyse automatisée des métadonnées. Les données récoltées ont permis d'approfondir l'analyse des sites web et de l'enquête en allant plus loin dans la prise en compte de l'argumentaire des promoteurs des IDG par rapport aux quatre dimensions étudiées.

Les métadonnées.

L'extraction des métadonnées des géocatalogues est réalisée à partir des services web de catalogage CSW-2. Le contenu des balises de métadonnées jugées utiles pour répondre à nos questions de recherche est stocké dans une base de données PostGreSQL. 31 des 45 IDG du corpus disposent de services web de catalogage opérationnel qui nous ont permis d'analyser plus de 160.000 fiches de métadonnées. La base de données est exploitée par différents outils d'analyse en fonction des besoins.

2.2.2. L'analyse des données.

Des citations issues des *verbatim* de sites web et d'entretiens

Les argumentaires relevés lors de l'analyse des sites web ou lors des entretiens ont permis de récolter une série de *verbatim* qui sont mobilisés, par le biais de citations, sans traitement statistique pour conserver la richesse des points de vue.

Des statistiques pour traiter l'enquête et les métadonnées

Des traitements statistiques univariés, réalisés avec le logiciel R, ont permis de produire des éléments de synthèse facilitant l'exploitation des résultats de l'enquête et l'exploration des fiches de métadonnées.

Des analyses de réseaux pour traiter l'enquête et les métadonnées

Les traitements statistiques sur l'enquête et la base de données intégrant le contenu des balises de métadonnées sont complétés par des analyses réseaux. Elles permettent en particulier d'explorer les connexions entre IDG par des graphes sur le moissonnage des géocatalogues. D'autres analyses réseaux, réalisées avec le logiciel TULIP, permettent de lier les métadonnées en fonction de leur proximité thématique grâce au calcul d'un indicateur de similarité. Celui-ci peut se définir comme suit :

⁵ Trois entretiens dans une région métropolitaine issue de territoires nouvellement fusionnés (Aquitaine-Limousin-Poitou-Charentes) et trois entretiens dans une région ultra-marine (la Guyane) qui comptent chacune trois IDG ont été complétés par six entretiens réalisés au niveau national auprès des porteurs d'IDG qui couvrent l'ensemble du territoire.

SAGEO'2016 – Nice, 6-9 décembre 2016

Soit deux fiches de métadonnées **f1** et **f2** :

- On détermine un vecteur **vref** de mots-clés correspondant à l'union des mots-clés de **f1** et **f2**.
- Pour **f1** et **f2**, et pour chaque mot-clé **m** de **vref** si **m** est utilisé dans la fiche, on crée pour chacune des fiches un vecteur de valeurs binaires de la taille de **vref** déterminant si le mot-clé est présent ou non.

Ces vecteurs sont nommés **v1** et **v2**. Le score **S** de similarité est déterminé par :

$$S = \frac{\sum_{i \in [0, \text{taille}(v_{ref})[} (v_1[i] * v_2[i])}{\sqrt{\sum_{a \in v_1} a^2} * \sqrt{\sum_{b \in v_2} b^2}} \quad (1)$$

Ce score correspond au cosinus de l'angle entre les vecteurs **v1** et **v2**. Variant entre 0 et 1, il permet de déterminer l'ampleur de la proximité thématique entre deux jeux de données. On détermine ensuite un seuil arbitrairement compris dans [0,1] et si le score entre deux fiches est supérieur au seuil, on crée un lien entre les fiches.

De l'analyse spatiale appliquée aux métadonnées

La couverture territoriale des données géographiques est décrite dans les métadonnées par les quatre coordonnées de leur rectangle d'emprise via la balise <EX_GeographicBoundingBox>. Ces données sont stockées dans PostgreSQL. Un traitement par PostGIS/QGIS permet de générer des cartes de densité des données.

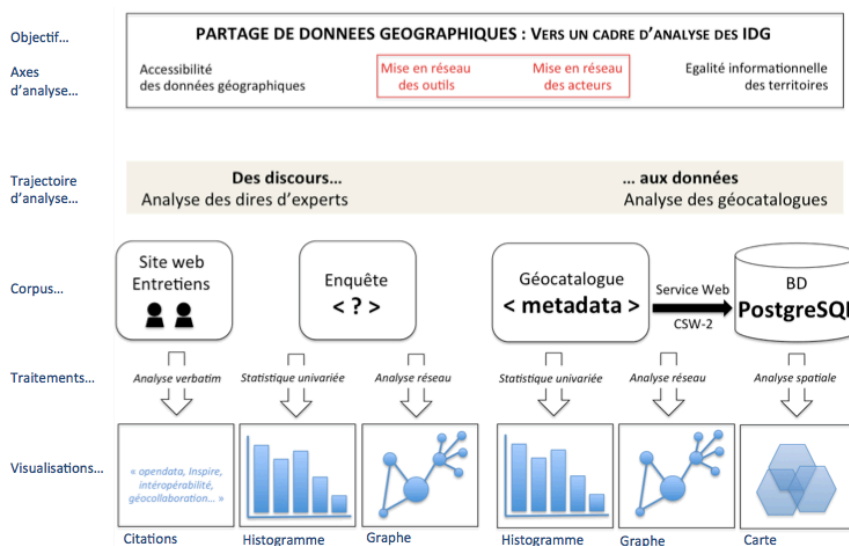


FIGURE 1. Synoptique du cadre d'analyse

Cadre d'analyse des IDG

Enquêtes et entretiens auprès des coordinateurs et exploration des sites web et des géocatalogues forment donc l'ossature du cadre méthodologique mis en œuvre pour évaluer l'opérationnalité des objectifs assignés aux IDG (Noucher et al., soumis). Dans la présente contribution, nous nous focalisons sur deux des quatre axes d'analyse pour nous intéresser à la question du décloisonnement des systèmes d'information (mise en réseau des SIG) et à celle des dynamiques collaboratives associées aux IDG (mise en réseau des acteurs). Ces deux axes nous permettent de mobiliser tout le spectre des corpus et outils d'analyse précédemment présentés.

3. Résultats⁶

3.1. Mise en réseau des outils

3.1.1. *Quel est le discours des acteurs sur la question du décloisonnement des SI ?*

L'un des arguments récurrents pour le déploiement d'IDG est lié aux capacités techniques dont ces plateformes disposent pour mettre en relation des bases de données stockées sur des serveurs distribués. Ainsi, 40 des 45 promoteurs d'IDG enquêtés déclarent mettre en œuvre des services web permettant d'accéder à distance à des métadonnées (CSW) ou des données (WMS, WFS) potentiellement distribuées sur différents serveurs. L'infrastructure nationale de données géographiques place ces questions d'interconnexion au cœur de sa stratégie de déploiement. La refonte en cours (2015/18) du géoportail/géocatalogue (refonte dite « geoportail v4 ») se fait, d'après les propos recueillis lors de notre enquête : « *dans un contexte d'émergence de l'État-plateforme* » et vise notamment à renforcer « *les mécanismes d'agrégation nationale des données produites régionalement ou localement* », à développer « *des géotraitements chaînés sur des sources hétérogènes* » ou encore à « *améliorer des mécanismes de moissonnage / mise en cache* ». L'État-plateforme est une stratégie nationale⁷ qui vise à favoriser l'émergence de nouveaux services publics numériques en s'appuyant sur une meilleure circulation des données. On retrouve ces préoccupations d'amélioration de la circulation de l'information au niveau régional avec des IDG qui se fixent pour objectif le « *développement de nouvelles fonctions (services web) pour faciliter les relais dans les territoires* » ou plus généralement de « *favoriser l'interopérabilité des systèmes pour réduire les dépenses publiques et faire gagner du temps* ». En fédérant les acteurs d'un même territoire ou d'une même thématique, les IDG sont censées décloisonner les SIG qui se sont déployés dans tous les métiers et à toutes les échelles depuis les années 80. Néanmoins une analyse de leurs géocatalogues permet de nuancer ces affirmations.

⁶ Des résultats détaillés sont aussi diffusés tout au long du projet sur le site www.geobs.cnrs.fr qui permet d'accéder aux publications, aux corpus de données et des visualisations interactives.

⁷ Stratégie définie par la DINSIC (Direction interministérielle du numérique, des systèmes d'information et de communication).

SAGEO'2016 – Nice, 6-9 décembre 2016

3.2.2. Les IDG sont-elles interconnectées pour accélérer les flux de données ?

Pour évaluer l'interconnexion des IDG, nous analysons le moissonnage des géocatalogues. Le moissonnage est un mécanisme permettant de collecter des métadonnées sur un catalogue distant et de les stocker sur le nœud local pour un accès plus rapide. Ce moissonnage est plus qu'un import de données simple car les métadonnées locales et celles du catalogue distant sont synchronisées. Par exemple, depuis un catalogue régional il est possible d'accéder à l'ensemble des métadonnées des catalogues départementaux et à une sélection de métadonnées issues d'un catalogue national. Analyser la mise en œuvre du moissonnage permet donc de mettre en lumière les relations entre IDG. Cette étude du processus de moissonnage permet en outre d'identifier par une analyse réseau les nœuds centraux (ou isolés) et d'éventuels *clusters* (à l'aide de la visualisation de sous-réseaux). En plus des connexions entre IDG du corpus, l'enquête a permis un premier recensement (sur une base uniquement déclarative) des connexions entre les géocatalogues des IDG et les géocatalogues d'observatoires, de portails *open data*, d'infrastructures infra-régionales... La figure 2 permet ainsi d'intégrer tous les liens de 2^e niveau, c'est-à-dire l'ensemble des systèmes directement en contact avec les IDG de notre corpus.

La densité⁸ de ce graphe est relativement faible ($d=0,02$). Les IDG régionales s'ouvrent vers des IDG départementales ou infra-départementales qui sont faiblement connectées au reste du réseau. La composante principale ne contient que 34 des 45 IDG du corpus initial auxquels sont associés 45 autres portails de données (d'échelon local pour la plupart). La portée de l'interopérabilité des infrastructures reste donc, encore aujourd'hui relativement limitée. Le géocatalogue du BRGM, point de relais national pour le rapportage de la directive INSPIRE, se singularise avec une connexion à 21 IDG, soit 4 fois la moyenne générale.

La figure 2 permet également de visualiser quelques circuits spécifiques d'accès à l'information :

- Deux composantes secondaires, c'est-à-dire déconnectées du réseau principal, dans le domaine de l'imagerie satellitaire (GeoSud et Theia) sont associées mais restent indépendantes du reste du réseau.
- Des composantes géographiques du réseau sont mises en évidence à travers la relation des IDG régionales aux plateformes infrarégionales. En l'occurrence, les jeux d'échelle sont très marqués à travers la structure centre-périphérie du graphe. La plupart des IDG nationales occupe une place centrale. Une première couronne périphérique donne un panorama des IDG régionales qui renvoient elles-mêmes vers une deuxième couronne périphérique composée majoritairement d'IDG initiées par des collectivités territoriales, des associations ou des syndicats départementaux ou intercommunaux.
- Des composantes thématiques du réseau apparaissent à travers les mises en relation d'IDG travaillant dans le domaine de la mer ou du côtier et

⁸ Densité d'un graphe = nombre de liens existants / nombre de liens possibles

Cadre d'analyse des IDG

sur des territoires ultramarins (lien Sextant – GéoMartinique ou PEIGEO) par exemple.

- Certaines infrastructures utilisant peu les techniques de moissonnage (nœuds de petite taille) occupent pourtant une place centrale car elles sont très moissonnées. C'est le cas d'EauFrance et de Géo-IDE.
- Enfin, le géocatalogue national occupe, sans surprise, le nœud central du réseau. Cependant, le portail « data.gouv.fr », bien que ne constituant pas une IDG *stricto sensu*, est lui-aussi fortement connecté au réseau.

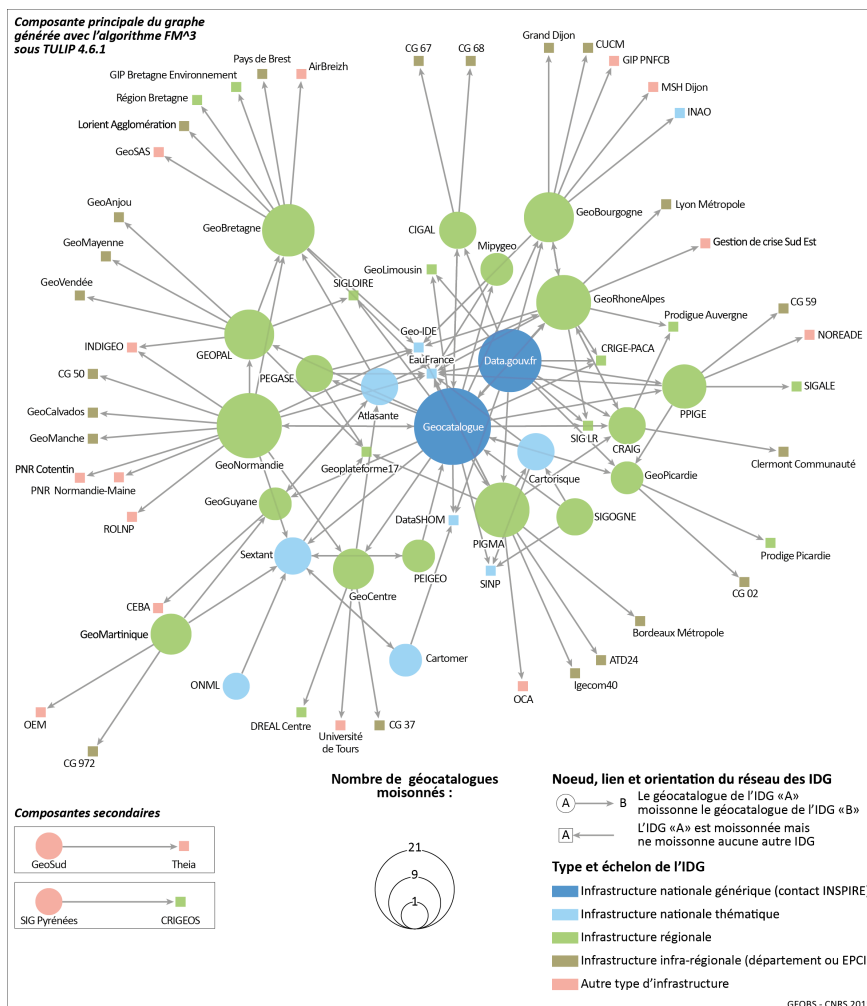


FIGURE 2. Réseau des relations inter-IDG établi par l'analyse des moissonnages entre les plateformes recensées.

SAGEO'2016 – Nice, 6-9 décembre 2016

3.2.4. Les IDG facilitent-elles l'hybridation de données issues de sources variées ?

Un focus sur une étude de cas régionale offre une autre lecture de la mise en réseaux des outils. Ainsi, après avoir regroupé les 2120 fiches de métadonnées issues de cinq portails de données géographiques déployés en Guyane française, nous calculons leur indice de similarité à partir des mots-clés renseignés pour chaque fiche de métadonnées. Cet indice permet la production d'un graphe (figure 3) pour visualiser les proximités thématiques de l'ensemble des différents éléments (métadonnées) du corpus. On voit alors assez nettement se reformer le découpage par IDG. Seulement trois composantes (A, B, C) mixent des métadonnées issues d'IDG différentes. Les « jonctions » observées sont thématiques (la composante A est exclusivement liée aux habitats et biotopes, la composante B ne concerne que des données relatives à l'imagerie satellite) et organisationnelle (au sein de la composante C, le pont entre les métadonnées de GéoGuyane et celles de Guyane SIG est assurée en C1 par les données du Parc Amazonien de Guyane, établissement public partenaire des deux infrastructures). De plus, aucune hybridation entre sphère scientifique et sphère institutionnelle n'est observée.

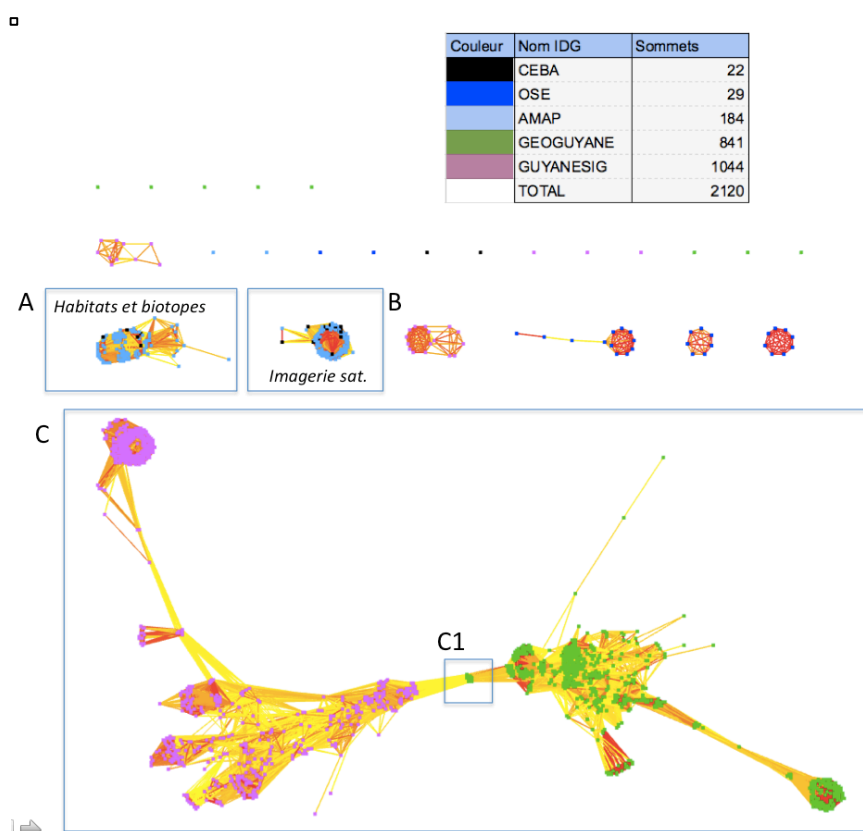


FIGURE 3. Graphe de similarité thématique des fiches de métadonnées de Guyane.

3.3. Mise en réseaux des acteurs

3.3.1. Quel est le discours des acteurs sur les dynamiques de géocollaboration ?

La plupart des IDG accompagne le déploiement de leur dispositif technique d'une série d'actions d'animation visant à favoriser l'appropriation des outils ou des données par un large spectre d'acteurs, et à impulser une dynamique autour d'un thème ou d'un territoire donné. Les formations aux outils, aux données, aux aspects réglementaires ou thématiques côtoient des actions de communication et l'animation de groupes de travail. Ce faisant, ces plateformes cherchent à initier de véritables communautés de pratique pour favoriser des dynamiques de géocollaboration et enrôler une grande diversité d'acteurs, même s'il existe des différences notables entre les échelons national (plus homogène) et régional (figure 4).

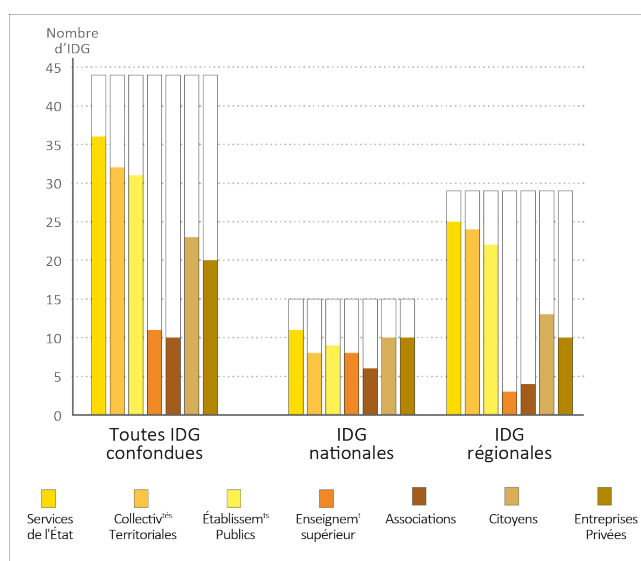


FIGURE 4. Répartition des types d'acteurs visés par les IDG régionales et nationales.

3.3.2. Les IDG favorisent-elles des dynamiques de géocollaboration ?

Environ la moitié des IDG du corpus anime un pôle métier, assimilé à un lieu d'échange dédié à une thématique et/ou à des outils pour favoriser la mise en cohérence des données et leur partage. Si en moyenne chaque IDG anime 5 à 6 pôles métier, des différences importantes sont à souligner entre les plateformes nationales (dont un tiers anime un pôle) et les plateformes régionales (dont les deux tiers animent un pôle). Les 132 pôles métier fonctionnels au sein des 45 IDG du corpus concernent un panel très diversifié de thématiques qui permet de comprendre les principaux chantiers sur lesquels travaillent actuellement les géomaticiens : occupation des sols, urbanisme, forêt, voirie, télécommunication, foncier, etc.

SAGEO'2016 – Nice, 6-9 décembre 2016

3.3.3. Les IDG permettent-elles la mise en réseau d'acteurs diversifiés ?

L'analyse des métadonnées révèle la faible diversité des contributeurs des IDG. Deux exemples peuvent illustrer ce constat. Au niveau national, le géocatalogue du BRGM accueillait, à l'été 2016, une très large majorité des données de l'État (49%) dont une part importante provient des services déconcentrés tels que les directions départementales (37%) qui gèrent le plan cadastral. La très faible présence des entreprises privées (3%) et des ONG (3%) est à souligner. Chaque IDG reste donc compartimentée dans des sphères d'acteurs spécifiques. Au niveau régional, le constat est identique. Ainsi, l'infrastructure régionale « GéoLimousin » propose essentiellement des données de l'État (93% en 2015) bien qu'elle soit la seule plateforme de la région.

4. Discussion et perspectives

4.1. Un fonctionnement en silos qui perdure

Si l'interopérabilité des systèmes d'information a incontestablement permis des avancées majeures pour la normalisation des services web et l'interconnexion potentielle des infrastructures de données, nos analyses concernant le moissonnage des géocatalogues tendent cependant à en limiter leur portée effective. Les IDG étudiées restent compartimentées dans des sphères d'acteurs et des thématiques spécifiques. L'analyse de l'ensemble du corpus montre des possibilités de moissonnage encore faiblement exploitées et l'étude de cas guyanais illustre une forme de fragmentation de l'information dans laquelle chaque plateforme reste dans son périmètre géographique et dans son silo thématique même si le potentiel de croisement existe.

4.2. Des communautés actives mais peu diversifiées

La mise en réseau des acteurs via des dynamiques de collaboration associées au déploiement des IDG semble effective. Les 132 pôles métier identifiés et l'activité importante d'animation des IDG, en particulier celles de l'échelon régional, témoignent de communautés de pratique très actives. Elles jouent un rôle clé dans la fédération des acteurs du territoire pour co-concevoir des modèles, co-produire des jeux de données et/ou les co-contrôler. Cependant, ces dynamiques de géocollaboration s'observent aujourd'hui dans un entre-soi peu diversifié : les administrations d'un côté, la recherche de l'autre, le secteur privé non représenté et le milieu associatif encore marginal.

4.3. Perspectives

Cette étude illustre la nature complexe, dynamique et multiforme des IDG. Nous favorisons une approche exploratoire multi-disciplinaire pour aborder leur

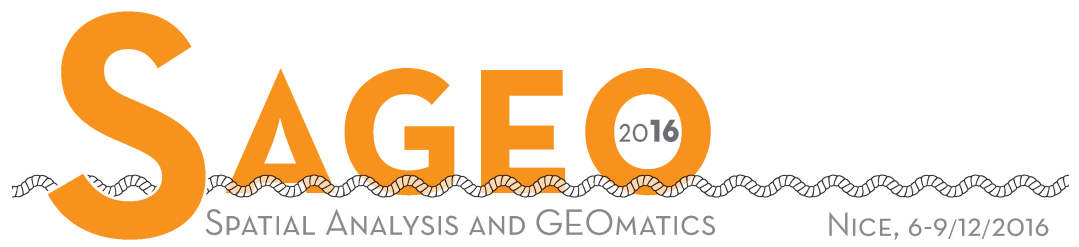
Cadre d'analyse des IDG

complexité et comparer la théorie et la pratique. Si les IDG apparaissent comme un « label » porteur pour les autorités publiques, on constate différentes formes d'acceptation du terme qui témoignent de réalités (technique et organisationnelle) variées. Le cadre méthodologique présenté est une première étape pour comparer les discours aux données effectivement partagées et aux actions de géocollaboration mises en oeuvre. Une deuxième étape consistera à mettre en parallèle les discours des promoteurs, les données et services des IDG et les usages effectifs qui y sont associés (Georis-Creuseveau et *al.*, 2016). L'objectif sera d'évaluer l'impact du partage de données géographiques sur les pratiques de gestion territoriale en analysant finement les usages effectifs et en vérifiant leur cohérence par rapport aux objectifs initiaux des IDG étudiées.

Bibliographie

- Ballatore A. (2014). Exploring the geographic information universe : the role of search technologies. CEUR Workshop Proceedings 1273, Workshop on Geographic Information Observatories at *GIScience*, Vienna, October 2014, pp. 23–28
- Coetzee S., Wolff-Piggott B. (2015) A Review of SDI Literature: Searching for Signs of Inverse Infrastructures. *Cartography - Maps Connecting the World*. Part of the series Lecture Notes in *Geoinformation and Cartography* pp 113-127.
- Crompvoets J., Bregt A., Rajabifard A., Williamson I. (2004). Assessing the world wide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science*, 18, 665-689.
- Georgiadou Y., Rodriguez-Pabon O., Lance K.T. (2006). SDI and e- Governance: A quest for appropriate evaluation approaches, *URISA Journal: Journal of the Urban and Regional Information Systems Association*, vol. 18(2).
- Georis-Creuseveau J., Claramunt C., Gourmelon F. (2016). A modeling framework for the study of spatial data infrastructures applied to coastal management and planning. *IJGIS*, <http://dx.doi.org/10.1080/13658816.2016.1188929>
- Hendriks PHJ, Dessers E., Van Hootegen G. (2012). Reconsidering the definition of spatial data infrastructure. *IJGIS*, 26(8):1479-1494
- Johnson R.B. & Onwuegbuzie AJ. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*. 33:14-26.
- Kitchin R. & Lauriault T. (2014). Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work, in Eckert, J., Shears, A. and Thatcher, J. (eds) *Geoweb and Big Data*. University of Nebraska Press. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Masser I. (2010). *Building European spatial data infrastructures*, ESRI Press, Redlands.
- Noucher M., Gourmelon F., Gautreau P., Maulpoix A., Pierson J., Pissoat O., Rouan M. (soumis). Spatial Data Sharing: A Pilot Study of the French SDIs, *IJDE*.
- Rajabifard A., Feeney M-E., Williamson I., Masser I. (2003). «Chapter 6, National SDI Initiatives», In *Development of Spatial Data Infrastructures: from Concept to Reality*, Williamson I, Rajabifard A., Feeney M.-E.F, London: Taylor & Francis, pp. 95-109.

[ISBN : 2-910545-11-3]



Session 3

Dynamiques urbaines, accessibilité

Analyser la dynamique intra-urbaine dans le temps long : une approche par la localisation des activités

Lucie NAHASSIA¹

1. UMR 8504 Géographie-cités, Université Paris 1 Panthéon-Sorbonne
13, rue du Four, 75006, Paris, France
lucie.nahassia@parisgeo.cnrs.fr

RESUME. La dynamique urbaine, à l'échelle du système de villes comme à l'échelle intra-urbaine, gagne à être étudiée dans le temps long (Pumain, 1997 ; Galliéni, 2000). Nous proposons de l'aborder en analysant les processus spatiaux de la localisation des activités dans l'espace intra-urbain et en mobilisant le cas de la ville de Tours sur toute sa durée d'existence, soit plus de 2000 ans. Cette amplitude temporelle représente un défi à la fois conceptuel et méthodologique et nécessite de se placer dans une approche interdisciplinaire, ici entre la géographie et l'archéologie urbaine. Nous présentons un travail de traitement sémantique et statistique d'un jeu de données archéologiques, analysé dans ses dimensions fonctionnelle, spatiale et temporelle pour caractériser la dynamique des activités à Tours dans le temps long.

ABSTRACT. The dynamic of an urban system is better understood when studied in a long term perspective (Pumain, 1997; Galinié, 2000). We propose to explore this long term dynamic at the scale of the intra-urban space by studying the location of activities in the city of Tours and over its whole existence (over 2000 years). Such a long period of time is both a methodological and conceptual challenge which is addressed by an interdisciplinary approach set between geography and urban archeology. We present here the semantic and statistical processing of the archeological data used to characterize the dynamic of activities in the city of Tours along three dimensions: function, space and time.

MOTS-CLES : dynamique urbaine, temps long, activités urbaines, analyses spatio-temporelles

KEYWORDS: urban dynamic, long term, urban activity, spatio-temporel analysis

Analyser la dynamique intra-urbaine dans le temps long

1. Introduction

Si la forme de la ville peut être analysée à partir des éléments qui la compose aujourd'hui, la structure urbaine actuelle gagne également à être lue comme le résultat d'une longue évolution (Pumain, 1997). Cet ancrage dans le temps long est valable tant pour l'échelle du système urbain, où les réseaux sont anciens et durables (Fletcher, 1986), que pour l'échelle intra-urbaine, où la matérialité se fabrique en héritant de son propre passé (Galinié, 2000).

Cette communication se situe dans le cadre de la dynamique urbaine dans le temps long. D'un point de vue thématique, il s'agit d'étudier la localisation des activités dans la ville à l'échelle intra-urbaine. L'originalité de ce travail consiste à prendre des bornes temporelles les plus étendues possibles – autrement dit à étudier un objet urbain dans toute la durée de son existence. Le cas de la ville de Tours est ainsi mobilisé, depuis sa fondation au premier siècle de notre ère jusqu'à aujourd'hui, et étudié dans ses limites urbaines telles qu'elles évoluent au cours du temps

En géographie, les durées étudiées pour saisir la dynamique urbaine ne dépassent généralement pas quelques décennies. Ce positionnement dans la longue durée engage donc des enjeux importants en termes de démarche. Du point de vue empirique, le parti pris de ce travail consiste à croiser l'analyse de sources multiples et complémentaires. Dans cette communication, on présentera l'exploitation d'une base de données archéologiques et historique dans le cadre d'analyses spatio-temporelles dans la longue durée.

2. Les activités, entre pratique et matérialité urbaine, comme marqueur du changement urbain dans le temps long

Le travail présenté ici porte d'un point de vue thématique sur la localisation des activités en milieu urbain : activités marchandes, économiques, religieuses, etc. La notion d'activité constitue en effet un objet intéressant pour l'étude de la dynamique de l'espace intra-urbain dans le temps long.

D'une part, les activités s'inscrivent très concrètement dans la matérialité de la ville : dans son bâti, sur ses places, etc. L'étude de leur localisation, et de l'évolution de leur localisation, permet donc de dégager une structure de l'espace intra-urbain qui change au cours du temps. On peut ainsi observer des logiques de localisations relatives au contexte urbain (centre/périphérie, attractivités de certains types de voies de communication...) ou mettant en jeu les activités les unes par rapport aux autres (jeu d'attraction ou de répulsion entre les activités).

D'autre part, les activités en tant que telles et leur rôle vis-à-vis de la société changent selon les époques : la religion par exemple n'a plus le même rôle d'encadrement de la population aujourd'hui que celui qu'elle assurait du moyen-âge jusqu'au XIXe siècle. Ces modifications au niveau de la société ont un impact fort sur l'organisation spatiale de la ville : si la religion n'est plus si importante, les bâtiments qui la portent ne sont plus aussi polarisateurs pour les autres activités, et

SAGEO'2016 – Nice, 6-9 décembre 2016

d'autres logiques spatiales se mettent alors en place. L'activité permet donc de lier les usages de la société en ville avec la matérialité urbaine (Roncayolo, 2002).

Par ailleurs, en étudiant toutes les activités observables en ville, on s'intéresse tout autant aux services et commerces banals du quotidien de la vie urbaine – la boulangerie, l'école primaire – qu'aux activités exceptionnelles qui participent à définir la place d'une ville dans un système urbain – la préfecture, l'université. En ce sens, l'activité permet de relier les deux échelles de compréhension de la ville comme « système dans un système de ville » (Berry, 1964).

L'objectif de ce travail consiste donc à étudier les dynamiques spatio-temporelles de la localisation des activités dans l'espace intra-urbain, en mobilisant la ville de Tours comme cas d'étude, du premier siècle de notre ère à aujourd'hui. Il s'articule autour de la description des structures intra-urbaines définies par la localisation des activités et de l'identification des logiques de ces localisations, de leurs processus spatio-temporels.

Les questionnements induits par cette thématique sont donc d'ordres spatiaux et temporels. Du point de vue spatial, en prenant par exemple le cas des activités industrielles, on peut se demander où se situent les bâtiments industriels dans l'espace urbain. Peut-on établir des critères de cette localisation et sont-ils constants dans le temps ? Quels sont les effets de cette localisation industrielle sur la localisation d'autres types d'activités ? Du point de vue temporel, se pose la question des rythmes du changement : les modifications des logiques de localisations des différents types d'activités sont-elles synchrones ? Si une première activité est un attracteur spatial pour l'implantation d'un deuxième type d'activité, quel est le décalage temporel entre leur implantation respective ? Au sein des logiques spatiales propres à une activité, peut-on identifier des moments d'accélération et des moments de permanences ?

Ces questionnements temporels et les difficultés posées par leur inscription dans le temps très long nécessitent d'adopter une posture interdisciplinaire. Ainsi ce travail, s'il est principalement géographique, s'articule à une approche archéologique de l'espace urbain, tant sur le plan théorique que méthodologique.

3. Des données aux activités urbaines

3.1. Le corpus de données

Pour étudier la ville sur le temps très long, l'observation empirique s'appuie sur des sources, historiques et archéologiques, qui sont par nature incomplètes. Il est donc nécessaire de mettre en place un corpus de sources complémentaires, tant du point de vue de leur nature (textuelles, iconographiques...) que des périodes qu'elles renseignent. Dans ce travail, la localisation des activités à Tours est donc identifiée à partir :

- de la littérature historique et archéologique disponible sur la ville,

Analyser la dynamique intra-urbaine dans le temps long

- des documents cartographiques où apparaissent des informations sur les activités urbaines (carte d'état-major, carte topographique de l'IGN...),
- d'entretiens effectués avec des experts archéologues et historiens spécialistes de l'urbain ou de la ville de Tours,
- et de données géographiques, archéologiques et historiques géolocalisées renseignant plusieurs dimensions de la topographie de la ville de Tours sur plus de 2000 ans, formalisées et centralisées dans le système d'information géographique « Topographie de Tours Pré-Industrielle » (ToToPI) (cf. infra).

Les analyses présentées dans cette communication s'appuient sur des données extraites de ce SIG et enrichies par les entretiens réalisés auprès des experts archéologues et historiens.

3.2. Les données de ToToPI

L'analyse empirique de la ville de Tours s'appuie en partie sur le SIG ToToPI, développé depuis 1996 au Laboratoire Archéologie et Territoire de l'Université François Rabelais de Tours (LAT)¹ (Rodier, 2000), qui assemble plusieurs couches d'informations centrées sur la ville de Tours : cadastres, modèle numérique de terrain, sites de fouilles, etc. Nous exploitons en particulier une base de données de ce SIG qui porte sur des éléments constitutifs de la topographie urbaine (un bâtiment, un rempart, un verger, etc.), issus d'études archéologiques et historiques sur la ville de Tours (résultats de fouille, étude du bâti, etc.) (Galinié *et al.*, 2004). Pour plus de simplicité, nous désignerons ici cette BDD sous le nom de « BDD ToToPI ».

Cette BDD est construite selon le principe de la triade de Peuquet (Peuquet, 1994) : chaque individu dans la base est un « Objet Historique » (OH) qui est caractérisé par une localisation (« where ») sous forme de polygone, de ligne et/ou de point géolocalisé, une durée (« when ») établie par une date de début et de fin, et une nature (« what ») (figure 1) définie par un thésaurus archéologique fonctionnel². Des métadonnées complètent ces trois informations centrales (site de fouille et sources de référence par exemple).

Cette BDD nous fournit donc des éléments de base en termes de datation, de localisation et de caractérisation sémantique pour l'étude de la dynamique des localisations des activités dans l'espace intra-urbain.

¹ Equipe Laboratoire et Archéologie Urbaine, UMR 6173 CITERES

² Cette conception des données en Objet Historique / Fonction / Espace / Temps a donné lieu au modèle de donnée « OH_FET » qui déconstruit le temps et l'espace en plus petites unités définies par la superposition des durées et des géométries existantes (Rodier et Saligny, 2010; Perret *et al.*, 2015). Ce modèle n'étant pas utilisé dans sa totalité ici, il n'est pas présenté en détail.

SAGEO'2016 – Nice, 6-9 décembre 2016

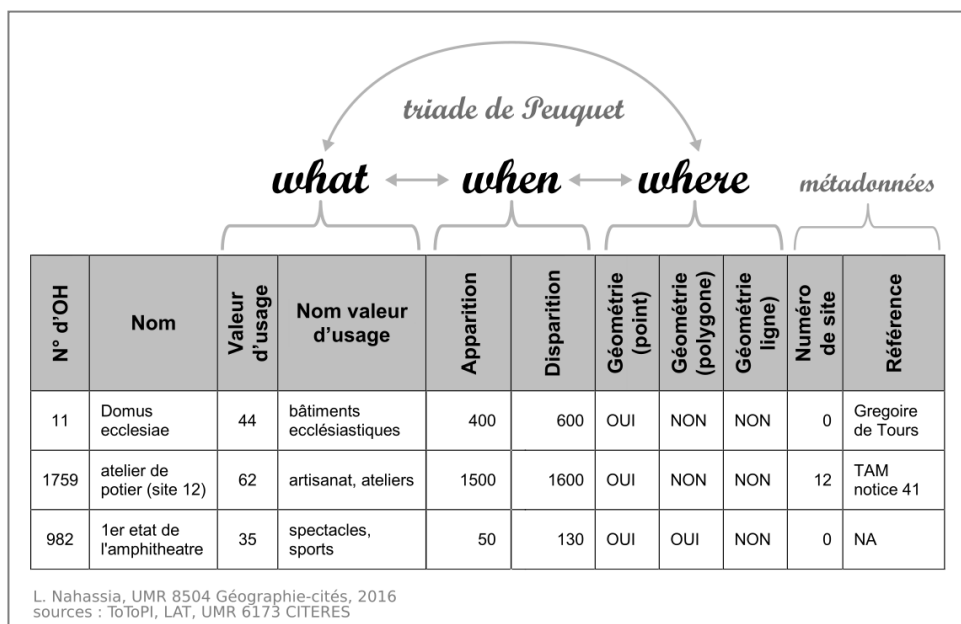


FIGURE 1. Extrait de la BDD ToToPI au regard de la triade de Peuquet.

Les données de la BDD ToToPI sont initialement caractérisées sémantiquement selon un thésaurus (extrait présenté en figure 2) établi par le Centre National d'Archéologie Urbain (Cerruti *et al.*, 2007) destiné à l'indexation et l'analyse et la comparaison des données urbaines issues des fouilles archéologiques urbaines. Il classe les OH selon une identification fonctionnelle des vestiges en deux niveaux hiérarchisés : les « valeurs d'usage », caractérisant un type d'occupation à l'échelle locale, regroupées en « valeurs urbaines », plus générales.

1. Voirie, aménagements	2. Structures défensives et militaires	3. Constructions civiles
11. voies, rues	21. système défensif urbain	31. espaces publics
12. espaces libres	22. structures fortifiées	32. pouvoir civil, justice
13. aménagement des berges	23. garnisons, casemements	33. éducation, culture
14. aménagements du relief		34. santé
15. franchissements		35. spectacles, sports
16. adductions d'eau		36. bains, thermes
17. collecteurs, évacuations		37. habitat privé
18. monuments, vestiges		
19. édifice monumental indéterminé		

Source : ToToPI, LAT, UMR 6173 CITERES

FIGURE 2 : Exemples de catégories du thésaurus de ToToPI. Premier niveau en gras : valeurs urbaines, deuxième niveau en régulier : valeurs d'usage.

Analyser la dynamique intra-urbaine dans le temps long

Cette typologie, si elle permet d'identifier la nature des usages des OH, est insuffisante pour obtenir un objet « activité urbaine » satisfaisant pour notre problématique. En effet les catégories utilisées se révèlent parfois trop englobantes : une valeur d'usage « bâtiment culturel catholique » recouvre par exemple tant les chapelles que les églises paroissiales et les cathédrales, qui ne relèvent pas du tout de la même importance en termes d'encadrement de la population, et dont le rôle dans l'organisation de la structure urbaine est potentiellement différent.

3.2. Compléter la caractérisation sémantique des données pour aborder les activités dans l'espace urbain

Afin de catégoriser de manière plus fine les objets historiques, nous proposons d'enrichir la caractérisation sémantique d'origine de la BDD ToToPI selon une deuxième grille fondée sur les portées des différentes activités au cours du temps. Par exemple, quelle est la portée d'un amphithéâtre pendant l'antiquité ? La population le fréquentant provenait-elle seulement de la ville considérée ou cette activité drainait-elle une population plus lointaine ? Quatre niveaux de portée ont été établis :

- 1 : portée courte, la population est desservie à l'échelle du quartier (ex : l'atelier de potier)
- 2 : portée moyenne, la population est desservie à l'échelle de la ville, de l'espace intra-urbain (ex : les thermes romains)
- 3 : grande portée, la population est desservie à l'échelle régionale (ex : la cathédrale)
- 4 : portée exceptionnelle, la population est desservie à l'échelle nationale et au-delà (ex : la basilique de Saint Martin, lieu d'un des pèlerinages les plus importants d'Europe au moyen-âge).

Cette nouvelle catégorisation a donc l'avantage de distinguer des objets qui ont la même valeur d'usage, mais dont le rayonnement, et donc la place dans l'organisation sociale et spatiale de la ville, est très différent. Pour reprendre l'exemple des bâtiments religieux, l'église paroissiale a pour fonction d'encadrer une population dans un périmètre restreint dans la ville, on peut donc considérer que sa portée est locale (portée 1), tandis que la cathédrale est le bâtiment de référence pour l'évêché et de ce fait a une portée régionale (portée 3).

Pour attribuer ces nouvelles catégories aux objets historiques, un ensemble d'enquêtes a été mené auprès d'experts archéologues, spécialisés en archéologie urbaine et sur le cas de Tours. Des entretiens semi-directifs ont été réalisés à ce jour auprès de quatre archéologues rattachés au LAT, et d'autres entretiens sont prévus d'une part avec des historiens et d'autre part avec d'autres archéologues de l'urbain.

La figure 3 présente la structure générale des données dans leur dimension fonctionnelle (histogrammes des OH par valeur urbaine et par portée) et spatiale (carte) et qui sont croisées avec la dimension temporelle dans les analyses suivantes.

SAGEO'2016 – Nice, 6-9 décembre 2016

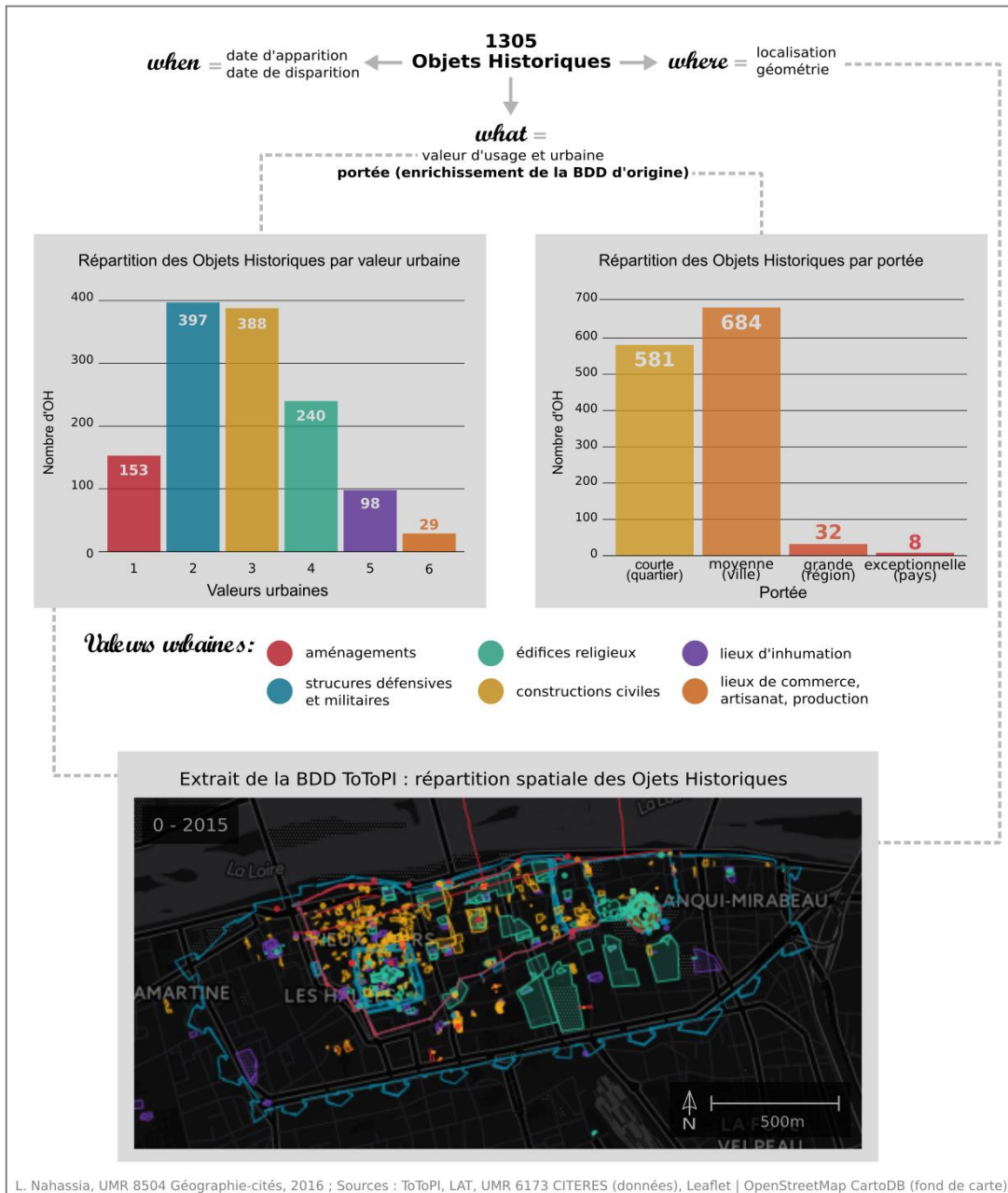


FIGURE 3 : ToToPI et portée : présentation des données.

Analyser la dynamique intra-urbaine dans le temps long

4. Des activités à la caractérisation de la structure urbaine dans le temps long

4.1. Mise en place d'une chaîne de traitement répliquable

La BDD ToToPI est en soi un objet en constante évolution. D'une part, de nouveaux objets historiques y sont régulièrement enregistrés à partir d'études archéologiques anciennes ou récentes sur la ville de Tours. D'autre part, comme on l'a vu, la dimension sémantique du modèle de données a pu être enrichie avec la construction de la typologie relative aux portées des activités, et d'autres caractérisations sont toujours susceptibles d'être ajoutées.

Afin de pouvoir s'adapter à cette versatilité de la BDD, le choix a été fait de mettre en place une chaîne de traitement qui soit la plus automatisée possible, aboutissant *in fine* à des analyses spatio-temporelles permettant d'identifier les changements dans la structure urbaine. Les outils choisis doivent donc être autant que possible interopérables et dans l'idéal permettre les traitements sous forme de script. Ce format de script permet également une reproduction des traitements par d'autres chercheurs en minimisant le temps de réappropriation des outils. Enfin, ce travail s'inscrit plus largement dans une perspective d'*open science* : les outils sélectionnés sont, dans la mesure du possible, libres et *open source*.

Les trois principaux outils choisis pour la mise en place du traitement des données sont (figure 4) :

- SQLite³ et son extension pour la gestion des données spatiales spatialite⁴ comme système de gestion de base de données, où se font toutes les tâches de complétion et de modification des données existantes dans ToToPI, comme par exemple l'ajout des portées pour chaque OH ;
- QGIS⁵, comme système d'information géographique qui nous permet d'une part d'enregistrer de nouveaux individus dans ToToPI à partir de leur localisation et d'autre part de visualiser les données dans leur implantation spatiale directement depuis la BDD SQLite ;
- R⁶, en lien direct avec SQLite, comme support des analyses statistiques et spatiales comme celles présentées ci-dessous.

L'usage conjoint de SQLite et de R permet de scripter toutes les opérations de traitements des données et d'analyses statistiques et spatiales, ce qui remplit les différents objectifs énoncés. En complément, un ensemble d'outils et de sources libres et interopérables est mobilisé pour les tâches de dessin vectoriel (Inkscape), de webmapping (leaflet, OpenStreetMap), etc.

³ <https://www.sqlite.org/>

⁴ <http://www.gaia-gis.it/gaia-sins/>

⁵ <http://www.qgis.org/>

⁶ <https://www.r-project.org/>

SAGEO'2016 – Nice, 6-9 décembre 2016

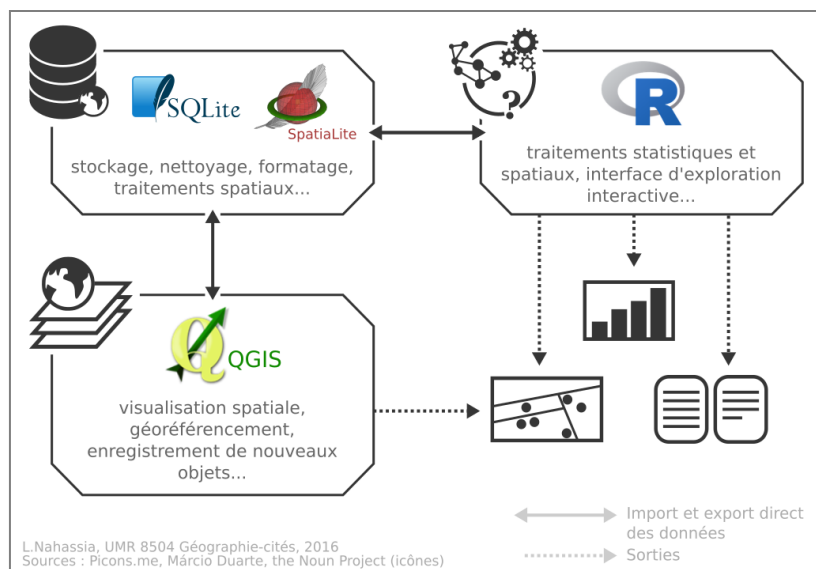


FIGURE 4 : la chaîne de traitement entre les trois principaux outils

4.2. Premiers résultats : la répartition des OH dans le temps selon leur activité

La communication présente une première série d'analyses exploratoires de la base de données à partir de deux dimensions : le temps et les activités selon leur valeur dans le thésaurus et selon leur portée (figure 5 et 6).

Une première série de sorties concerne par exemple la distribution des OH caractérisés par leur valeur urbaine dans le temps en effectif (figure 5.A) et en part du corpus total (figure 5.B).

Ces graphiques nous donnent des informations sur la structure élémentaire de nos observables et, de ce fait, mettent en évidence certains biais liés aux sources de la BDD. En 5.A, par exemple, la faible quantité d'OH entre 0 et 300 est le reflet d'une faible connaissance de la période du Haut Empire, tandis qu'au Xe siècle, une période mieux renseignée, le nombre d'entités présentes dans la BDD est multiplié par deux.

La répartition des valeurs urbaines dans le temps renvoie également à des structures urbaines bien établies. Que ce soit en effectif ou en pourcentage, l'importance que prennent les structures défensives et militaire avec des début de phase très net en 350, en 900 ou en 1350 correspondent aux moments de construction des enceintes de la ville : en 350 l'établissement d'une ville close à l'est de l'actuel centre-ville, puis l'édification d'un *castrum* à l'ouest autour d'un nouveau secteur autour de la communauté de Saint-Martin, et enfin une grande enceinte réunissant les différents ensembles urbains en 1350.

Analyser la dynamique intra-urbaine dans le temps long

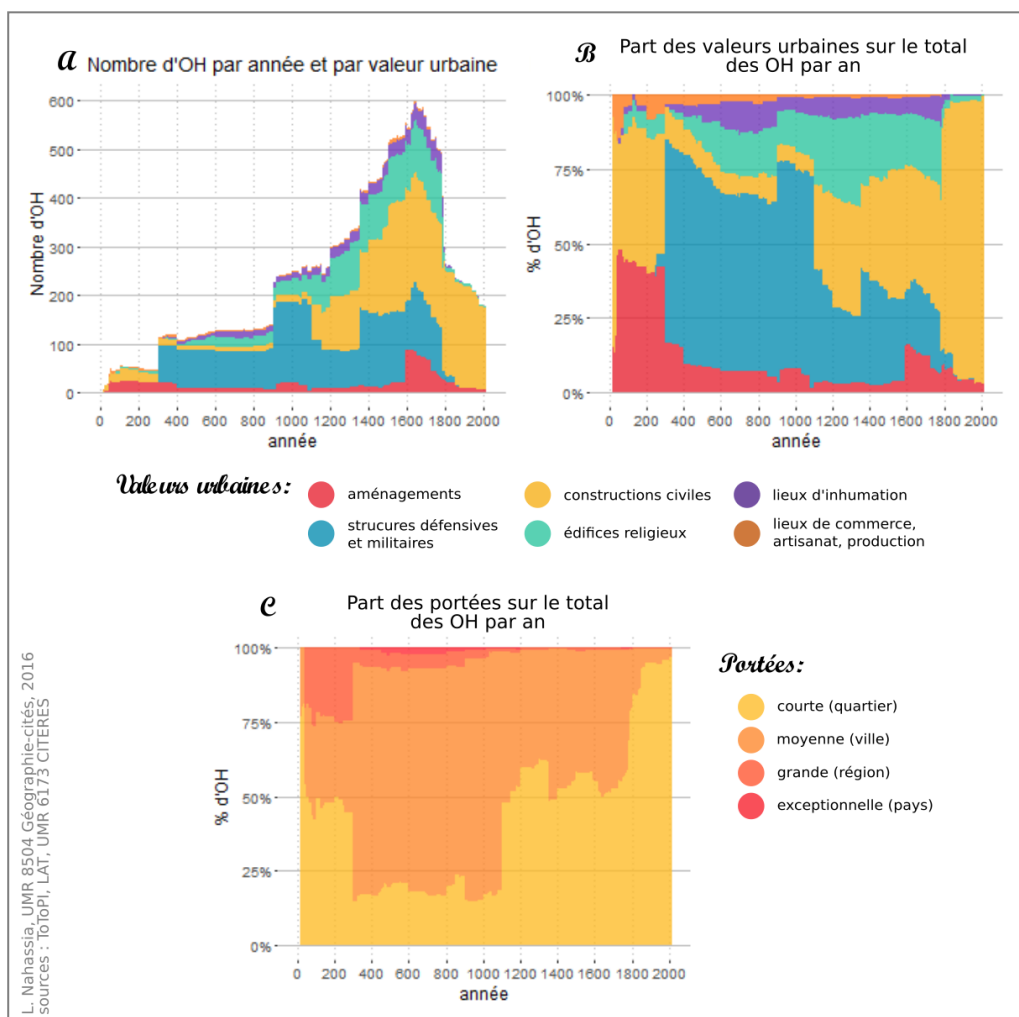


FIGURE 5 : Distribution dans le temps des Objets Historiques par valeur urbaine et par portée

L'analyse factorielle des correspondances présentée sur la figure 6 permet de préciser cette vision d'ensemble. On y croise des intervalles temporels (classes de 50 ans) et l'importance de chaque type de valeur urbaine (comptabilisée en nombre d'OH dans chaque catégorie). Les deux premières composantes de l'AFC, qui expliquent respectivement 64 et 22% de la variance totale, mettent très nettement en évidence des ruptures temporelles entre des ensembles de périodes proches les unes des autres.

Certaines de ces ruptures correspondent à de grandes périodes structurantes de l'histoire de la ville de Tours. Les historiens et archéologues définissent ainsi une

SAGEO'2016 – Nice, 6-9 décembre 2016

« ville ouverte » à partir la fondation de Tours par les romains et jusqu'au milieu du III^e siècle, étape bien individualisée ici par le groupe de périodes comprises entre 50 et 250, caractérisées surtout par les constructions civiles (valeur urbaine 3) et les aménagements (des berges, du relief, pour la circulation de l'eau en ville, etc., valeur urbaine 1).

A l'inverse, alors que certaines étendues temporelles sont très fragmentées par les spécialistes de l'histoire urbaine de Tours, elles apparaissent dans cette AFC d'un seul bloc, ici par exemple l'ensemble 550-1050, complètement agrégé à droite du premier axe. Entre ces deux dates, les études historiques et archéologiques distinguent au moins trois moments dans le développement de la ville : la construction de la basilique Saint-Martin sur la tombe du saint du même nom à l'ouest à l'extérieur du centre fortifié et l'attraction que cela confère à cet espace ; le développement et l'émancipation de ce secteur vis-à-vis de la cité à l'est ; la bipartition effective de la ville avec la construction d'une deuxième enceinte qui enclos le secteur martinien. Dans l'AFC, ces changements n'apparaissent pas de manière claire, les périodes considérées sont très proches car elles comptent toutes de très nombreux objets historiques de type « structure militaire et défensive » (valeur d'usage 2) qui constituent les enceintes des espaces ouest et est (chaque enceinte est décomposée en courtines, murs, portes et tours). Cette similarité est ainsi plus forte que la diversité introduite par la présence des activités religieuses et civiles.

Dans ces analyses, les dimensions temporelle et fonctionnelle des données permettent de mettre en évidence une première caractérisation de l'espace intra-urbain en évolution au cours du temps et confirment certaines connaissances sur la ville de Tours. Néanmoins, il est indispensable d'introduire la dimension spatiale des objets historiques pour affiner l'analyse. L'émergence de la bipartition de la ville pendant le haut moyen-âge n'est pas différenciée au niveau des données par un changement dans les types d'activités présents, mais bien par une structure spatiale particulière, comme l'illustrent les cartes en *snapshot* de la figure 6.

Ces quelques extraits d'analyses font ressortir deux manières complémentaires d'aborder le temps. Pour une partie d'entre elles, il est considéré comme continu : on observe les fonctions année par année sans interruption, ce qui permet de dégager des plages de temps cohérentes *a posteriori* (cf. figure 5). Pour d'autres, en revanche, le découpage *a priori* du temps est nécessaire : il s'agit de définir des périodes bornées. On peut alors choisir de découper le temps de manière régulière (comme on l'a fait pour l'AFC), à l'instar d'un découpage de l'espace en carroyage régulier, sans a priori sur les périodes historiques classique ou sur l'histoire propre de la ville de Tours, ou au contraire considérer ces spécificités historiques pour établir des bornes temporelles – ce qui reviendrait pour l'espace à prendre comme découpage des limites de territoires par exemple administratifs.

Analyser la dynamique intra-urbaine dans le temps long

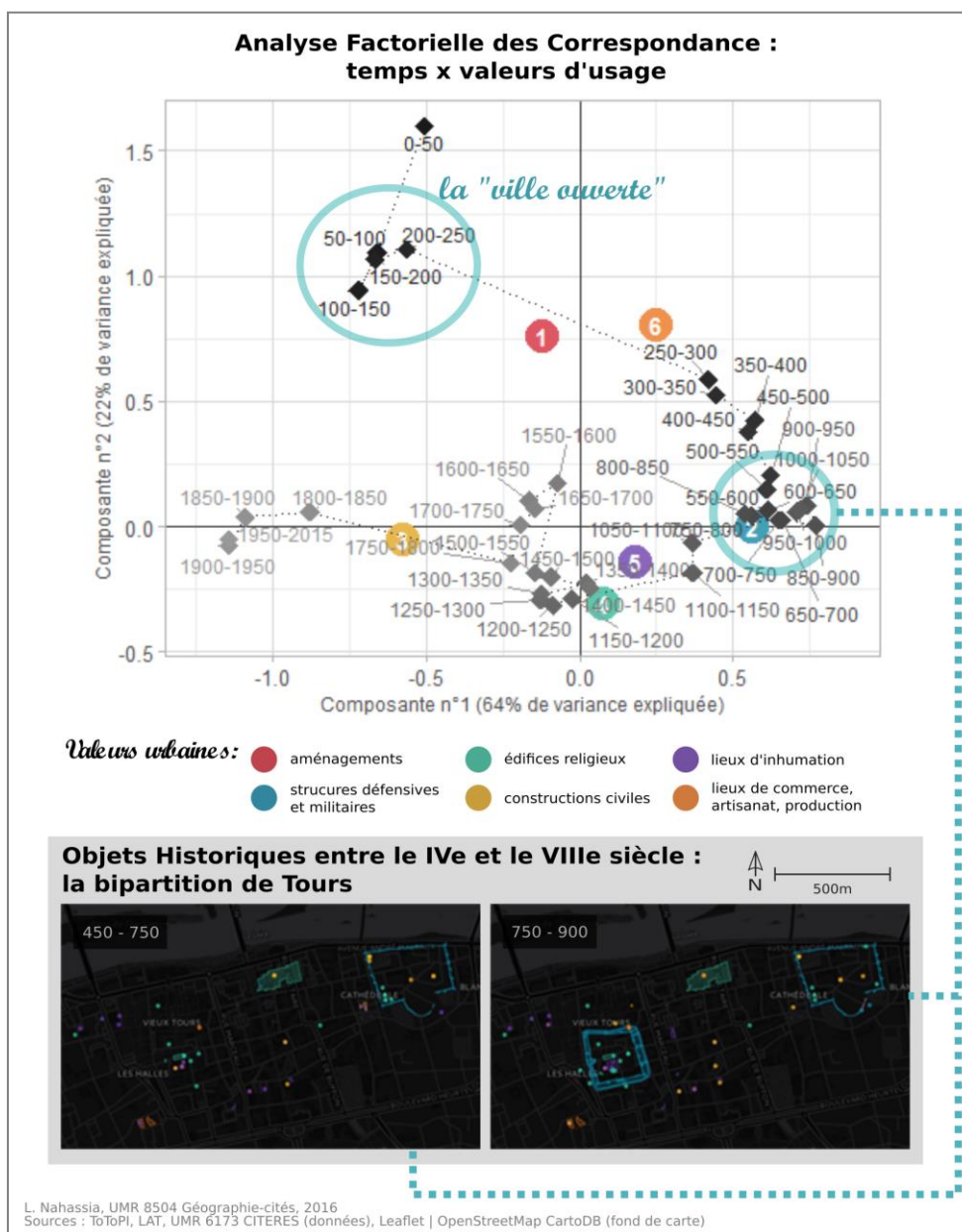


FIGURE 6 : Analyse Factorielle des Correspondances

SAGEO'2016 – Nice, 6-9 décembre 2016

4.4. Traitements en cours de développement : ajouter la dimension spatiale

Une troisième série d'analyses statistiques croisant les dimensions temporelles et fonctionnelles des OH avec leur localisation sont en cours de développement et seront présentées au cours du colloque.

On proposera une analyse du semis des OH selon leur valeur d'usage et leur portée, en termes de concentration, d'autocorrélation spatiale, ou encore au regard des limites urbaines évaluées pour chacune des époques considérées. Il s'agira également par exemple d'étudier la répartition de ces activités en termes de voisinage les unes par rapport aux autres, toujours selon les deux catégories usage/portée, et selon l'évolution de leurs proximités et de leurs éloignements au cours du temps (figure 7).

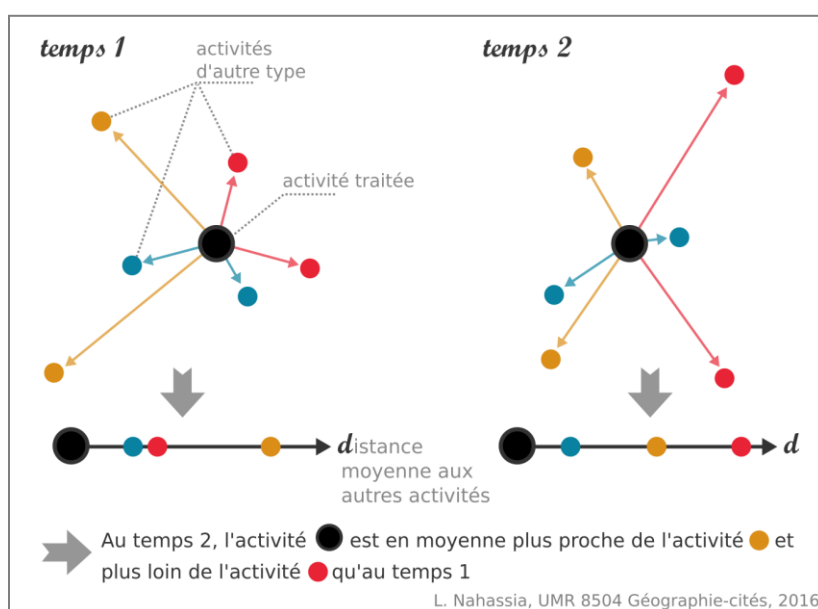


FIGURE 7 : Principe de traitement pour l'analyse du voisinage des activités au cours du temps.

5. Perspectives : l'analyse empirique comme étape d'une démarche modélisatrice

La richesse et l'amplitude temporelle des données issues du SIG ToToPI permettent d'étudier dans le temps long les changements de la structure urbaine par le prisme de la localisation des activités, à travers des analyses statistiques et spatiales encore en cours de développement. Ces analyses, associées à l'étude du reste du corpus présenté, et malgré les biais liés à la nature des sources (elles ne couvrent pas uniformément toutes les périodes ni toute l'étendue de Tours), mettent ainsi en évidence des logiques de localisation spatiales, génériques ou propres à

Analyser la dynamique intra-urbaine dans le temps long

certains types d'activités, permanentes ou changeant dans le temps. Ces logiques constituent des candidats explicatifs pour à une compréhension plus générale de la localisation des activités en ville.

Le travail de recherche présenté ici est une des composantes d'une approche modélisatrice de la dynamique urbaine. Les règles identifiées à partir des différentes sources sont formalisées et combinées dans des modèles conceptuels et graphiques. Les modèles statiques ainsi produits sont ensuite implémentés dans des modèles de simulations afin de modéliser le changement urbain de manière dynamique. Ces derniers constituent des espaces de test de différents scénarios, des laboratoires d'expérimentation des possibles permettant de mieux comprendre l'imbrication des différentes logiques mises en jeu et leurs implications spatiales dans le temps long.

Bibliographie

- Berry B. (1964). Cities as systems within a system of cities. *Papers in Regional Science*, 13(1).
- Cerruti M., Desachy B., Guilloteau C. (2007). *Les données du Centre national d'archéologie urbaine, aperçus statistiques*.
- Fletche, R. (1986). Settlement in Archaeology : World-Wide Comparisons. *World Archaeology*, 18(1), p. 59-53.
- Galinié H. (2000). *Ville, espace urbain et archéologie*.
- Galinié H., Rodier X., Saligny L. (2004). Entités fonctionnelles, entités spatiales et dynamique urbaine dans la longue durée. *Histoire & mesure*, XIX(3/4), p. 223-242
- Perret J., De Runz C., Rodier X., Varet-Vitu A., Dumenieu B., Saligny L., Cristofoli P., Lefebvre B., Desjardin E. (2015). Études des dynamiques de l'occupation du sol. Questionnements, simplifications et limites. *Revue Internationale de Géomatique*, 25(3/2015), p. 301-330.
- Peuquet D. (1994). It's about time; A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of the American Geographers*, (3), p. 441-461.
- Pumain D. (1997). Pour une théorie évolutive des villes. *l'Espace géographique*, 26(2), p. 119-134.
- Rodier X. (2000). Le système d'information géographique TOTOPI : Topographie de TOurs Pré-Industriel. *Les petits cahiers d'Anatole*, (4).
- Rodier X., Saligny L. (2010). Modélisation des objets historiques selon la fonction, l'espace et le temps pour l'étude des dynamiques urbaines dans la longue durée. *Cybergeo: European Journal of Geographie*.
- Roncayolo M. (2002). *Lectures de villes : formes et temps*. Marseille, France: Éd. Parenthèses.

RELOC : un modèle pour (Re)localiser plutôt que construire des infrastructures de transport en vue d'améliorer l'accessibilité.

Cyrille Genre-Grandpierre¹, Serigne Gueye²,
Cyril Sahuc²

1. UMR Etude des Structures, des Processus d'Adaptation et des
Changements de l'Espace (ESPACE), Avignon, France
cyrille.genre-grandpierre@univ-avignon.fr

2. Laboratoire d'Informatique d'Avignon (LIA), Université d'Avignon, France
{serigne.gueye,cyril.sahuc}@univ-avignon.fr

*RÉSUMÉ. L'évaluation coût- bénéfice des infrastructures de transport reste principa-
lement basée sur l'amélioration de l'accessibilité induite par les gains de temps de
déplacements. Le temps gagné est monétarisé afin d'estimer, au regard du coût de
l'infrastructure, si elle est « rentable ». Ce mode d'évaluation a largement contribué à
favoriser la construction d'infrastructures rapides participant ainsi au développement
de la dépendance automobile. Dans un contexte de pénurie financière et d'aspiration à
plus de durabilité, ce travail vise à savoir s'il est possible d'améliorer l'accessibilité en
relocalisant optimalement les aménités urbaines (ici l'emploi) plutôt qu'en construi-
sant des routes. L'outil de simulation Ac-Rel a été développé dans cette perspective et
est utilisé à titre heuristique pour le cas de l'Anneau des Sciences de Lyon.*

*ABSTRACT. The cost-benefits evaluation of new road infrastructures remains mainly
based on the improvement of the accessibility due to the decrease in travel times. The
time saved is monetized and allows to say if the infrastructure is cost effective or not
for a given temporal term. It often leads to new speed infrastructures with automobile
dependency as a consequence. In a context of less funding, and a more effective as-
piration to sustainability, our work aims to evaluate in what extent it is possible to
improve accessibility by relocating optimally urban facilities (here jobs) rather than by
providing new road infrastructures. The simulation tool Ac-Rel has been developed in
this perspective and is tested for the case of the Ring of science in Lyon.*

MOTS-CLÉS : accessibilité, relocalisation, optimisation, évaluation d'infrastructures

KEYWORDS: Accessibility; Relocation; Optimization; Infrastructure evaluation

Introduction

Pendant longtemps, l'évaluation coûts-bénéfices des nouvelles infrastructures routières de transport s'est très largement basée sur l'estimation des gains de temps générés par ces infrastructures (Boiteux M., 2001, « les gains de temps représentent 80 à 90 % des avantages mesurés par les évaluations socioéconomiques des grands projets »). Caricaturalement, ces gains de temps, qui concernent très essentiellement les déplacements en voiture, sont monétarisés, pour estimer, à échéance donnée, si l'infrastructure est, ou pas, « rentable ». Aujourd'hui, les processus d'évaluation se veulent beaucoup plus complets et cherchent à analyser les effets de l'infrastructure sur la carte de l'accessibilité et sur les dynamiques territoriales induites¹. C'est particulièrement vrai, pour l'accessibilité à l'emploi, qui est une des principales qualités de la ville et conditions de sa compétitivité (Prud'homme R., Lee C., 1999), eu égard à l'émergence des économies d'agglomération et d'urbanisation qu'elle permet. Dans cette approche de l'évaluation basée sur la mesure de la possible amélioration de l'accessibilité, les localisations des points de mesure (ex les domiciles) et de la ressource (ex les emplois) sont données, de même que le réseau de transport considéré comme fixe, à l'exception de la nouvelle infrastructure (Crozet, Y., Mercier, A., Ovtracht, N., 2013). Une fois la mesure des changements d'accessibilité automobile réalisée, à l'aide de systèmes d'information géographique et de bases de données décrivant les routes, on cherche à en modéliser les effets à terme sur l'utilisation du sol, et plus précisément sur la localisation des ménages et activités, comme le font les modèles LUTI land-use transport integrated (Simmonds D., Echenique M. et al. Bates J., 1999), tel que Pirandello (Piron V., Delons J., 2008) pour, in fine, estimer si l'infrastructure « vaut le coût ».

Ce type d'évaluation revient à analyser les effets de la nouvelle infrastructure routière soit sur l'accessibilité, soit sur le développement urbain à terme, plus qu'à réellement en tester la pertinence, en se posant par exemple la question de savoir à quelles conditions, en termes de localisations des aménités urbaines ou encore de développement de modes de transport alternatifs à la voiture, il serait possible de se passer de cette infrastructure, tout en obtenant des résultats du même ordre en termes d'accessibilité. Ainsi, depuis des décennies, la recherche d'une bonne accessibilité à l'emploi (gage d'une bonne adéquation entre l'offre et la demande qui est aussi profitable aux entreprises qu'aux salariés, (Mignot D., Rosales-Montano S., 2006, l'accessibilité à l'emploi est très directement reliée au taux de chômage)), est essentiellement passée par l'augmentation de la vitesse des déplacements automobiles grâce à l'amélioration des routes.

Or, aujourd'hui, avec des ressources publiques plus réduites et dans un

1. L'accessibilité se définit comme la facilité depuis un lieu pour atteindre un (des) autre(s) lieu(x) ou une ressource localisée.

SAGEO'2016 – Nice, 6-9 décembre 2016

contexte où la problématique du développement durable est plus prégnante, et va notamment de pair avec des aspirations à plus de lenteur pour des raisons de sécurité, de limitation des émissions de polluants, mais aussi pour la recherche de fonctionnements territoriaux davantage basés sur la proximité plus que sur la vitesse (cf. par exemple le concept de ville des courtes distances), la construction de nouvelles infrastructures rapides paraît de plus en plus difficile, mais sans pour autant que la population veuille dans le même temps renoncer à l'accessibilité. L'accessibilité dépendant certes de la vitesse (accéder vite à des lieux), mais aussi, comme on a trop tendance à l'oublier, de la localisation des aménités (avoir la ressource à proximité), l'objectif de ce travail consiste à voir dans quelle mesure il est possible d'évaluer la pertinence d'une infrastructure routière, non en simulant et évaluant les effets, mais en voyant si cette infrastructure peut être évitée, à objectif d'accessibilité donné, en jouant sur la localisation des aménités. Plus précisément, il s'agit :

- soit de maximiser l'accessibilité moyenne en relocalisant la ressource pour voir si le niveau d'accessibilité fournie par l'infrastructure peut être atteint ou dépassé
- soit de trouver la quantité minimum de ressource à relocaliser pour atteindre un niveau d'accessibilité fixé a priori (ici celui fourni par la nouvelle infrastructure).

Dans les deux cas l'infrastructure sera considérée comme inadéquate si le coût des relocalisations, toutes choses égales par ailleurs, s'avère inférieur au coût de l'infrastructure.

Il est aujourd'hui aisé, avec les systèmes d'information géographique de calculer des accessibilités et d'en simuler les évolutions consécutives à des modifications des réseaux routiers. Il est aussi relativement aisé, avec des méthodes de location-allocation (Thomas I., 2012) de distribuer dans l'espace une quantité donnée de ressource pour maximiser l'accessibilité lorsque les lieux de demande sont connus. En revanche il n'existe pas (à notre connaissance) de méthode ou outil qui permette une relocalisation d'une ressource dans un espace en vue d'en maximiser l'accessibilité en tout lieu sans demande prédéfinie, ou qui permette d'atteindre un niveau d'accessibilité donné en trouvant, à partir d'une distribution originelle, le minimum de ressource à relocaliser. Nous avons donc développé un outil Ac-Rel qui répond à cette problématique de relocalisation optimale d'une ressource (ici les emplois) avec un objectif d'amélioration de l'accessibilité. Outre la nouveauté du regard portée sur l'accessibilité, cet outil se démarque également par le fait qu'il s'appuie sur des méthodes issues de la recherche opérationnelle qui restent très peu mobilisées en analyse spatiale. L'objectif de ce travail est donc à la fois heuristique, puisqu'il s'agit de montrer le poids relatifs des localisations et de la vitesse dans le mesure de l'accessibilité, et méthodologique puisqu'il est nécessaire de développer l'outil conséquent. Le cas d'étude concerne l'Anneau des Sciences (ADS) du Grand Lyon. L'ADS est une infrastructure autoroutière prévue pour 2028 qui vise à relier différentes

RELOC

zones d'activités scientifiques de l'ouest de l'agglomération (Grand Lyon et Conseil Général du Rhône., 2012) en vue d'améliorer l'accessibilité à l'emploi sur le territoire.

Notre approche consiste à calculer l'accessibilité à l'emploi avec et sans l'ADS en se basant sur la distribution des emplois en 2010 et celle projetée par le Grand Lyon en 2030, puis à comparer ces accessibilités à celles obtenues en relocalisant les emplois de façon optimale grâce à Ac-Rel. Les calculs s'effectuent à l'échelle des 744 IRIS du Grand Lyon, pour lesquels nous avons le nombre d'emplois en 2010 et 2030 et les populations totales aux mêmes dates. Les emplois sont indifférenciés c'est-à-dire que nous ne connaissons pas le type d'emplois présents. Cela constitue une limite de ce travail, sachant qu'à l'évidence la difficulté de relocalisation d'un emploi dépend de sa nature, sans pour autant remettre en question la portée heuristique et méthodologique du travail. Par ailleurs, plus thématiquement, s'il est relativement simple de connaître le type d'emplois actuels par IRIS, il n'en va pas de même pour 2030. Le Grand Lyon fait en effet des projections mais qui ne vont pas jusqu'à préciser quel type d'emploi se localisera où.

Ac-rel est donc utilisé pour calculer l'accessibilité à l'emploi pour les scénarios de transport et de localisation des emplois suivants :

- localisation de l'emploi projeté en 2030 sans et avec l'ADS
- localisation de l'emploi de 2030 optimisée de façon à maximiser l'accessibilité moyenne des IRIS sans et avec l'ADS
- Les mêmes scénarios que précédemment mais en se basant sur la localisation de l'emploi en 2010

Enfin, pour encore mieux évaluer le poids relatif de la vitesse et des localisations dans les calculs d'accessibilité, certains de ces scénarios sont repris en simulant des baisses de vitesses sur le réseau routier pour les automobiles : augmentation généralisée de 15 et 30 % des temps de transport.

La matrice des temps d'accès IRIS à IRIS en voiture nécessaires au calcul de l'accessibilité a été obtenue grâce à la plateforme MOSART² du Laboratoire d'Economie des Transports de Lyon (Bonnaïfous, A., Crozet, Y., Mercier, A., Ovtracht, N., Puech, F., Péguy, P-Y., 2010). Cette plateforme intègre une base de données très bien calibrée décrivant le réseau routier et ses attributs fonctionnels. Les mesures sont obtenues par des méthodes classiques de calcul de plus courts chemins que l'on trouve dans les SIG. Précisons que dans ce travail nous ne considérons que l'accessibilité en voiture afin de ne pas considérer trop

2. Nos remerciements à Nicolas Ovtracht et Aurélie Mercier du Laboratoire d'Economie d'Aménagement et des Transports de Lyon pour nous avoir fournis les temps d'accès entre IRIS en voiture issus de la plateforme MOZART, ainsi que les données sur les emplois actuels et projetés par le Grand Lyon.

SAGEO'2016 – Nice, 6-9 décembre 2016

de scénarios et parce que cet article a avant tout une visée méthodologique, plus que d'évaluation de l'ADS stricto-sensu. Néanmoins il est tout à fait possible de réaliser le même travail avec d'autres modes de transport (marche ou transport public), puisqu'il suffit de changer en conséquence la matrice des temps de transport IRIS à IRIS qui est utilisée. Se baser sur une matrice en temps de transport en commun permettrait par exemple de comparer l'accessibilité avec l'ADS sans relocalisation avec celle obtenue en transport public sans l'ADS mais en relocalisant les emplois. Une autre possibilité consisterait à utiliser une matrice IRIS à IRIS donnant le plus court chemin en temps par comparaison des différents modes et en retenant le plus rapide. Après un rappel sur le calcul de l'accessibilité gravitaire, qui est la mesure d'accessibilité la plus classiquement utilisée, et son utilisation dans une approche classique qui compare l'accessibilité avant et après la mise en place de l'infrastructure routière, nous présenterons la méthode de relocalisation optimale développée, puis les résultats obtenus pour les différents scénarios. Nous discutons enfin dans une dernière partie des améliorations de l'outil Ac-Rel.

Après un rappel sur le calcul de l'accessibilité gravitaire, qui est la mesure d'accessibilité la plus classiquement utilisée, et son utilisation dans une approche classique qui compare l'accessibilité avant et après la mise en place de l'infrastructure routière, nous présenterons la méthode de relocalisation optimale développée, puis les résultats obtenus pour les différents scénarios. Nous discutons enfin dans une dernière partie des améliorations de l'outil Ac-Rel.

1. L'accessibilité gravitaire à l'emploi dans le grand Lyon avec et sans l'Anneau des Sciences

1.1. l'accessibilité gravitaire

L'accessibilité locale exprime la facilité avec laquelle une ressource (emplois) ou des lieux peuvent être atteints au départ d'un lieu i à l'aide d'un système de transport donné (Morris et al., 1978). Elle dépend donc à la fois de la distribution de la ressource et de la qualité du système de transport (Geurs, K.T., Van Wee, B., 2004). Dans ce travail nous utiliserons la mesure classique d'accessibilité gravitaire qui permet au départ d'un lieu i de relativiser l'intérêt de la ressource présente en j en fonction de sa quantité et de la distance entre i et j . Au départ de i , plus la ressource en j est importante et proche, plus elle contribuera fortement à la qualité de l'accessibilité mesurée en i .

L'accessibilité gravitaire mesurée en i s'écrit ainsi :

$$A_i = D_j \exp(-\beta C_{ij}) \quad (1)$$

Avec A_i l'accessibilité en i (ici le centre des IRIS), D_j la ressource présente en j (ici les emplois),

β paramètre de sensibilité au coût généralisé du déplacement (0.18)

C_{ij} le coût généralisé du transport entre i et j qui est donné par :

$$C_{ij} = Cm_{ij} + T_{ij} * VdT$$

Avec Cm_{ij} le coût monétaire qui dépend de la distance en kilomètres entre i et j et d'un coût moyen au kilomètre (incluant le carburant, la maintenance du véhicule et les frais d'assurance) ici fixé à 0.49€, T_{ij} le temps de trajet entre i et j aux heures de pointe et VdT la valeur du temps par heure telle qu'elle est définie par la gouvernement français pour les déplacements urbains (ici 11.4).

1.2. Un faible gain d'accessibilité pour l'anneau des sciences

Cette accessibilité gravitaire est tout d'abord utilisée pour voir dans quelle mesure l'Anneau des Sciences est susceptible d'améliorer l'accessibilité à l'emploi à l'échelle des 744 IRIS du Grand Lyon. Ceux-ci comptent 1647722 habitants pour 803479 emplois en 2010 et les projections sont de 1982833 habitants pour 982775 emplois en 2030. L'Anneau des Sciences est une infrastructure de 14.8 km, prévue pour 2028, qui correspond à la partie ouest de la rocade autoroutière qui relie différentes zones d'activités scientifiques autour de Lyon. L'ADS vise à améliorer l'accessibilité des principaux sites d'activités du Grand Lyon, et à développer les activités dans l'ouest.

Que ce soit avec la distribution des emplois en 2010 ou celle prévue en 2030 par le Grand Lyon, l'ADS améliore logiquement l'accessibilité à l'emploi, mais de façon réduite seulement puisque le gain n'est que de l'ordre de 2% à l'échelle de la métropole. Ce gain potentiel est même en réalité plus réduit puisqu'il ne prend pas compte le trafic induit par l'infrastructure (jusque-là contenu en raison de temps de transport trop importants, mais qui peut se réaliser une fois l'infrastructure réalisée, Mogridge and Holden, 1987).

Spatialement, les gains les plus importants se font logiquement à l'ouest de Lyon, à proximité de l'ADS, mais aussi dans les zones plus lointaines qui y sont bien connectées (figure 1), ce qui permet d'éviter le tunnel sous Fourvière aussi bien pour accéder aux emplois au sud qu'à l'ouest. Les faibles diminutions d'accessibilité à l'Est sont imputables aux changements des plans de circulation consécutifs à l'ADS. Cette première analyse est utilisée comme point de repère, notre objectif étant de savoir dans quelle mesure cet apport de l'ADS en termes d'accessibilité (limitée à 2%) peut être obtenu en jouant sur les localisations des emplois. À cette fin, nous présentons tout d'abord la méthode de relocalisation utilisée avant de la tester pour différents scénarios.

2. Relocaliser les emplois pour maximiser l'accessibilité : méthode

L'objectif consiste à savoir comment relocaliser les emplois afin de maximiser l'accessibilité moyenne des 744 IRIS ou trouver le nombre de relocalisations

SAGEO'2016 – Nice, 6-9 décembre 2016

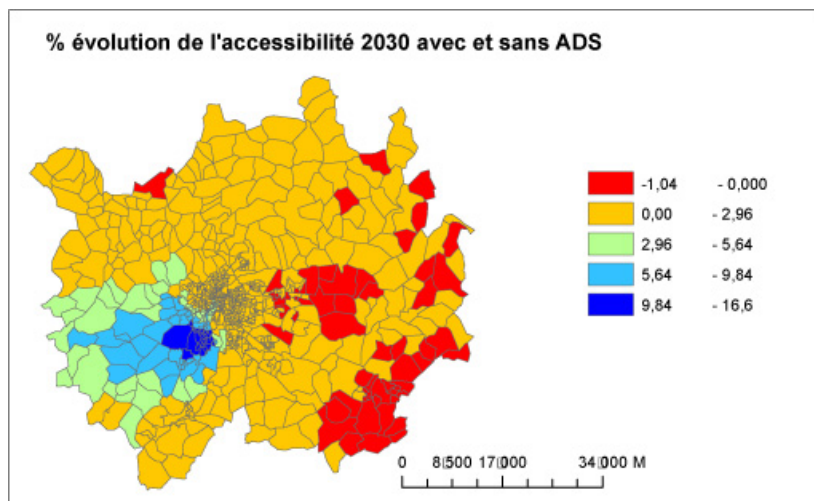


FIGURE 1. *Evolution de l'accessibilité gravitaire à l'IRIS avec l'Anneau des Sciences*

minimum à effectuer pour obtenir le même niveau d'accessibilité que celui donné par l'ADS en 2030. Dans notre approche, la ville est vue ici comme un potentiel, une offre (ici d'emplois) dont il s'agit de faciliter l'accès en tout lieu. Cette vision de la ville comme un potentiel, plutôt par exemple que d'analyser les navettes domicile-travail actuelles des travailleurs puis de chercher à en minimiser la somme des distances, s'inscrit dans une visée prospective et de résilience. On ne sait pas qui va où (ce qui peut changer rapidement) mais on cherche à faciliter l'accès au marché de l'emploi en tout lieu, ce qui permet une adaptation plus aisée de la ville (la nature des emplois, la distribution des travailleurs et leur type respectif peuvent changer, il n'en reste pas moins que le potentiel d'accès est là et qu'il est une des principales qualités urbaines (Glaeser E., 2011).

Dans le programme d'optimisation de l'accessibilité Ac-REI, la quantité à faire varier est donc D_j qui donne le nombre d'emplois par Iris.

Notons I l'ensemble des Iris, avec $q = |I|$ le nombre total d'Iris. La quantité à faire varier est donc D_j , où $j \in I$, qui donne le nombre d'emplois par Iris.

On suppose par ailleurs, pour avoir un problème bien posé, qu'il y a un nombre d'emplois total T possible, et que chaque variable D_j est bornée supérieurement par une quantité connue \bar{D}_j .

$$i.e \sum_{j \in I} D_j = T, \text{ et } D_j \leq \bar{D}_j \forall j \in I$$

L'absence de bornes de ce type aurait en effet pour corollaire de concentrer tous les emplois sur l'Iris le mieux positionné. Par la suite, quand $D_j = \bar{D}_j$ nous dirons que l'Iris j est **saturé** en emplois (à contrario non saturé). Pour chaque Iris $i \in I$, nous savons que l'accessibilité est définie par la quantité suivante :

$$A_i = \sum_{j \in I} D_j \frac{1}{e^{\beta C_{ij}}}$$

où $C_{ij} = Cm_{ij} + T_{ij}VdT$. L'accessibilité moyenne se définit alors par la quantité :

$$\frac{1}{q} \sum_{i \in I} A_i$$

Comme q est une constante, il est aisé de voir qu'en terme d'optimisation mathématique, déterminer la localisation maximisant l'expression ci-dessus revient à maximiser uniquement :

$$f(D) = \sum_{i \in I} A_i = \sum_{i \in I} \sum_{j \in I} D_j \frac{1}{e^{\beta C_{ij}}}$$

En permutant les deux sommes ci-dessus, on obtient l'expression équivalente suivante :

$$f(D) = \sum_{j \in I} D_j \left[\sum_{i \in I} \frac{1}{e^{\beta C_{ij}}} \right]$$

Soit $\alpha_j = \sum_{i \in I} \frac{1}{e^{\beta C_{ij}}}$. α_j dépend uniquement des valeurs C_{ij} donc implicitement des temps de transport entre Iris car à aucun moment n'intervient dans cette expression le nombre d'emplois ou les flots Origine-Destination entre Iris. Cette observation est à l'origine de la relative simplicité de la méthode de calcul des localisations optimales que l'on peut résumer par la proposition suivante.

PROPOSITION. — *Les valeurs D_j ($j \in I$) maximisent la fonction $f(D)$ (i.e accessibilité moyenne) si et seulement si elles vérifient la propriété suivante :*

(P) : *Pour tout Iris j et k , si $\alpha_j > \alpha_k$ alors soit j est saturé (i.e $D_j = \bar{D}_j$) ou $D_k = 0$.*

PREUVE. — La propriété signifie, en d'autres termes, que si la valeur α_j est supérieure à α_k alors la localisation optimale (i.e celle maximisant l'accessibilité

SAGEO'2016 – Nice, 6-9 décembre 2016

moyenne) des emplois saturera l'Iris j ou sinon ne placera aucun emploi dans k . Pour démontrer cette propriété, raisonnons par l'absurde en supposons le contraire. Admettons donc que dans la localisation optimale des emplois, ils existent deux Iris j et k pour lesquels $\alpha_j > \alpha_k$, $D_j < \bar{D}_j$ (i.e j n'est pas saturé) et $D_k > 0$ (i.e l'Iris k contient des emplois). Dans ce cas, il suffirait de déplacer un emploi de l'Iris k vers j pour augmenter l'accessibilité moyenne de $(\alpha_j - \alpha_k) > 0$. Ce qui est absurde puisque la localisation est supposée optimale. \square ■

La propriété énoncée dans la proposition ci-dessus, et la preuve qui l'accompagne, conduit alors à une méthode simple de calcul de la relocalisation optimale. Elle consiste à saturer d'emplois les Iris dans l'ordre décroissant des valeurs α_j . En procédant de la sorte, la propriété (P) est nécessairement vérifiée et les localisations ainsi trouvées optimales.

Le programme reposant sur ce modèle permet deux types d'optimisation :

- de maximiser l'accessibilité moyenne des IRIS.
- de trouver le nombre d'emplois minimum à ajouter à une situation donnée (ici celle de 2010) et leur distribution spatiale pour atteindre un niveau d'accessibilité donné (ici celui de 2030 avec l'ADS). Il s'agit donc ici en quelque sorte de minimiser « l'effort », le coût de changement, pour atteindre un niveau fixé d'accessibilité.

Pour chacun de ces deux types d'optimisation, plusieurs possibilités de contraintes de relocalisation ont été testées :

- relocaliser librement tous les emplois. Un seul Iris pourrait alors théoriquement accueillir tous les emplois. Les résultats de cette option peu crédibles ne sont pas présentés.
- donner une capacité d'accueil maximale à chaque IRIS. Puisque ces capacités n'existent pas théoriquement ou dans les plans d'aménagement, nous avons utilisé deux possibilités à titre heuristique. Dans la première le nombre d'emplois pouvant être localisés dans un IRIS correspond à sa population en 2010 * 1.5. Dans la seconde, elle est égale à sa population en 2030. Ces seuils ont été ici fixés à titre d'exemple. Ils ne reposent pas sur des observations empiriques (certains quartiers ont un nombre d'emplois qui dépasse 1.5 fois leur population). Ils sont paramétrables à souhait par l'utilisateur qui peut en outre choisir des seuils différents pour chaque IRIS.
- donner une capacité maximale d'accueil à chaque IRIS (comme ci-dessus), mais aussi un nombre minimum d'emplois afin d'éviter que le processus de relocalisation ne « vide » entièrement certains Iris de leurs emplois. Ce nombre minimum a été fixé à la moitié du nombre d'emplois de l'IRIS en 2010 (là encore à titre d'exemple).

RELOC

L'accessibilité a ensuite été calculée pour ces différentes logiques d'optimisation et pour différents scénarios de transport : temps de transport inter et intra-iris actuels, temps avec l'ADS, temps actuels *1.15 et temps actuels *1.3. Ces deux derniers cas ont pour objectif de voir si de bons niveaux d'accessibilité peuvent être obtenus non seulement en n'augmentant pas la vitesse, mais au contraire en la diminuant. Les temps de transport Iris à Iris sont calculés de centroïde à centroïde, ceux intra-iris ont été estimés en calculant les distances moyennes séparant 50 points tirés aléatoirement par IRIS (soit 2500 valeurs).

Le tableau ci-dessous synthétise les résultats pour les simulations assurant que chaque IRIS aura suite au processus de relocalisation au minimum 50% des emplois prévus en 2030, ce qui constitue le cas le plus défavorable en termes d'optimisation de l'accessibilité. Précisons que les résultats suivent la même logique si la liberté est laissée à la méthode d'optimisation de vider totalement certains IRIS de leurs emplois, à la différence près que les gains d'accessibilité augmentent encore en moyenne d'une dizaine de pour cent pour les scénarios 21 à 32. (Figure 2).

N° Scénario	21	24	27	30	33	34	35	36
Opération	Relocalisation emplois de 2030				Ajout emplois situation 2010			
ADS	Non		Oui		Non			
Limite de capacité d'accueil des Iris	Nombre emplois 2010 x 1,5	Population 2030	Nombre emplois 2010 x 1,5	Population 2030	Nombre emplois 2010 x 1,5	Population 2030	Nombre emplois 2010 x 1,5	Population 2030
Type Optimisation	Maximiser accessibilité						Minimiser nombre emplois à ajouter optimalement pour atteindre accessibilité 2030 avec ADS	
% emplois relocalisés	24%	34,35%	24%	34,35%	∅			
Nbr emplois ajoutés	∅				179296 (2030 - 2010)		97738	97360
Accessibilité avant (*10 ⁶)	122,55		124,92		106,25			
Accessibilité après (*10 ⁶)	148,31	152,63	151,01	155,32	139,14	140,3	124,92	
Gain accessibilité	21,03%	24,54%	20,88%	24,33%	31,15%	32,24%	17,74%	

FIGURE 2. Accessibilité à l'emploi pour différents scénarios

3. Relocaliser pour assurer un bon potentiel d'accessibilité sans vitesse

3.1. Maximiser l'accessibilité moyenne

La comparaison des accessibilités obtenues pour les différents scénarios envisagés montre qu'il est possible de se passer de l'ADS pour fournir un bon niveau d'accessibilité moyen. Par exemple, en relocalisant optimalement 24%

SAGEO'2016 – Nice, 6-9 décembre 2016

des emplois de 2030, l'accessibilité obtenue en 2030 avec rocade sans relocalisation augmente de 19% (124.92) avec comme contrainte de limiter la capacité d'accueil des IRIS à 1.5 fois le nombre d'emplois de 2010 et en assurant que chaque Iris aura en 2030 au minimum 50% des emplois qu'il avait en 2030 (sans cette dernière contrainte le gain est de près de 30%). Spatialement le couplage de l'outil de relocalisation avec un système d'information géographique permet de cartographier les résultats. Le processus de relocalisation des emplois qui maximise l'accessibilité moyenne sur la base des emplois de 2030 (scenario 21) conduit à améliorer sensiblement la situation des IRIS les centraux où se concentre la majorité des emplois relocalisés (Figure 3).

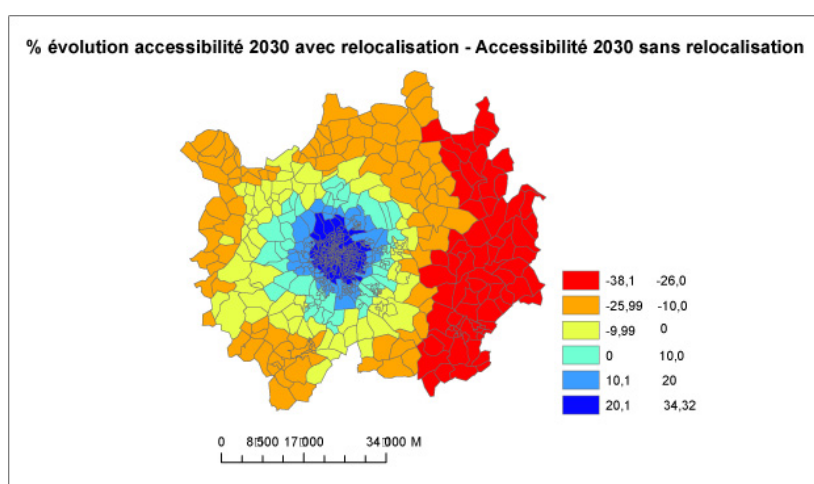


FIGURE 3. Evolution de l'accessibilité à l'IRIS consécutive au processus de relocalisation des emplois

En moyenne la relocalisation des emplois conduit à une augmentation d'une vingtaine de pour cent de l'accessibilité (scenarios de 21 à 30) et même davantage dans les cas où la marge de liberté est plus grande en ce qui concerne le nombre d'emplois relocalisables et la capacité d'accueil des IRIS. Les plus fortes valeurs d'accessibilité sont ainsi obtenues en limitant la capacité d'accueil des IRIS à la population 2030 ce qui conduit à pouvoir relocaliser 35% des emplois.

Les tests effectués en modifiant les temps d'accès montrent, que ce soit avec ou sans relocalisation, que la baisse d'accessibilité n'est que de 6% pour des temps d'accès multipliés par 1.15 et de 12% pour une multiplication par 1.3. Par ailleurs, le processus de relocalisation s'avère d'autant plus performant en termes de comparaison de l'accessibilité avant/après que les vitesses sont basses, même si les différences sont faibles. Cela s'explique par le fait que la géographie (i.e. les localisations) a d'autant plus d'importance que les vitesses sont faibles,

RELOC

car elles ne parviennent pas à « lisser » les situations comme peuvent le faire les hautes vitesses. Enfin, le processus de relocalisation permet de sensiblement améliorer l'accessibilité même en présence de la rocade (scenario 27). Relocaliser optimalement 24% des emplois avec l'ADS augmente ainsi de 20% l'accessibilité obtenue avec l'ADS seulement. C'est d'ailleurs logiquement dans cette situation (avec rocade et relocalisation) que les plus fortes valeurs d'accessibilité sont obtenues (151.01).

Si ces résultats qui montrent qu'il est possible d'obtenir le même niveau d'accessibilité qu'avec l'ADS en jouant « seulement » sur la localisation des emplois, ils peuvent toutefois paraître difficilement atteignables concrètement puisqu'ils conduisent à relocaliser entre un quart et un tiers des emplois selon la marge de liberté définie quant aux capacités d'accueil des IRIS. Si cette ampleur de relocalisation à horizon 2030 n'est concrètement pas impossible, elle reste toutefois difficile à réaliser, ou du moins le coût du changement risque d'être élevé surtout pour les emplois non tertiaires.

3.2. Minimiser les relocalisations à effectuer pour atteindre le niveau d'accessibilité de 2030 avec l'ADS

Aussi, nous intéresserons nous à présent à la deuxième logique d'optimisation, qui cherche à minimiser le nombre d'emploi à relocaliser pour obtenir un niveau d'accessibilité donné (ici celui de 2030 avec ADS) en partant d'une situation donnée (celle de 2010). Dans ce cas que les conclusions précédentes restent valables (scenario 35 à 40), à savoir que de bons niveaux de potentiel d'accessibilité peuvent être atteints sans ADS. Ainsi, en prenant comme situation l'emploi de 2010, il « suffit » par exemple de relocaliser optimalement 97738 emplois pour obtenir le niveau d'accessibilité de 2030 avec rocade, toujours dans la situation où les capacités d'accueil des IRIS sont limitées à 1.5 fois l'emplois de 2010 et où chaque iris doit garder 50% des emplois de 2030 (pour mémoire les projections du grand Lyon prévoient une augmentation totale du nombre d'emplois par rapport à 2010 de 179296) (Figure 4).

Si à présent on relocalise optimalement l'ensemble des 179296 emplois prévus d'ici 2030 à partir de la situation 2010 alors on améliore l'accessibilité de 2030 avec ADS de 11.5%. Si à présent l'ensemble des 179296 emplois prévus d'ici 2030 sont relocalisés optimalement à partir de la situation 2010, alors l'accessibilité de 2030 avec ADS augmente de 11.5%

Spatialement, les emplois relocalisés de façon à atteindre l'accessibilité de 2030 avec ADS, le sont dans les petits IRIS du centre de l'agglomération qui sont ceux pour qui chaque ajout d'emplois augmente le plus l'accessibilité moyenne. Ceci s'explique tout d'abord par leur localisation (ils sont en moyenne les plus accessibles), et par leur taille et leur nombre. Comme ils sont petits et concentrés l'algorithme de relocalisation les privilégie car c'est là que chaque emploi supplémentaire contribue le plus à l'augmentation de l'accessibilité moyenne

SAGEO'2016 – Nice, 6-9 décembre 2016

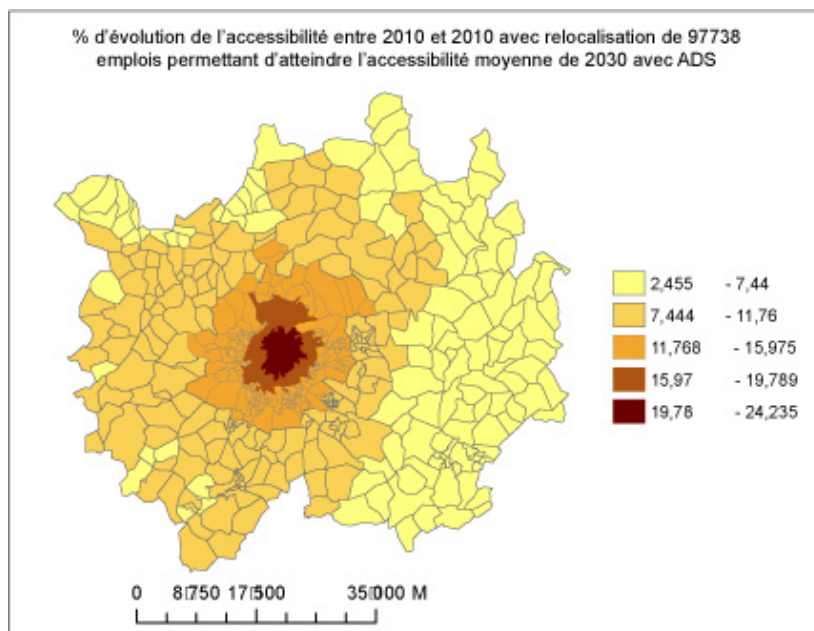


FIGURE 4. *Atteindre un niveau d'accessibilité donné en minimisant le nombre d'emplois à relocaliser*

étant donné que ces IRIS centraux ont de nombreux voisins accessibles. Cette relocalisation au centre s'explique aussi par un effet de bord. En effet, plus on va en périphérie moins les IRIS ont de voisins à une distance donnée (on sort de la zone d'étude) ce qui les rend moins attractif pour le processus de relocalisation. Pour limiter ce problème classique, il serait possible de collecter des données à une échelle plus large pour replacer chaque IRIS dans son contexte spatial. Précisons que ce problème n'est pas seulement technique, mais qu'il est aussi un problème en aménagement du territoire, où très souvent ce qui est à côté du territoire à aménager est ignoré, secteur sur lequel les aménageurs n'ont pas la main (les PLU voisins, ou encore les SCOT s'élaborent par exemple en grande majorité indépendamment les uns des autres, les PLUi et l'interscot étant encore embryonnaires).

4. Évolutions du modèle pour éviter les artéfacts liés aux découpages sur les relocalisations

Outre l'élargissement de la collecte des données pour limiter « l'effet centre », trois évolutions du modèle ont été réalisées.

RELOC

La première vise à prendre en compte les populations présentes dans chaque IRIS. Il ne s'agit plus de maximiser l'accessibilité moyenne où chaque IRIS est équivalent, mais de faire en sorte que la population qui profite de bons niveaux d'accessibilité soit la plus nombreuse possible. On maximise alors l'accessibilité moyenne pondérée par la population de chaque IRIS. Il ne s'agit plus comme précédemment de maximiser un potentiel théorique d'accès à l'emploi en tout lieu, mais de tenir compte de la structure actuelle du territoire et plus spécifiquement de la distribution de la population, pour maximiser la population ayant un bon niveau d'accessibilité.

La deuxième cherche à tenir compte de la surface des IRIS pour limiter l'effet lié à l'usage d'un découpage hétérogène comme peut l'être celui des IRIS. L'algorithme cherche alors à maximiser le fait que n'importe quelle portion du territoire, n'importe quel mètre carré, bénéficie du meilleur niveau d'accessibilité possible. La maximisation porte ainsi sur le niveau « d'accessibilité surfacique » c'est-à-dire l'accessibilité pondérée par la surface de l'IRIS.

La troisième évolution est plus thématique. Il s'agit en effet de favoriser les relocalisations dans les unités spatiales où les prix du foncier sont les moins élevés, qui sont globalement les plus périphériques.

Pour cette troisième évolution deux approches sont possibles :

- dans la première, les emplois sont relocaliser en priorité dans les secteurs les moins chers, quitte parfois à perdre un peu en accessibilité gravitaire standard. La maximisation s'effectue alors en utilisant le ratio " $\alpha_j / (\text{prixFoncier}_j)^a$ ". Plus l'exposant a au dénominateur est important plus le prix du foncier prime dans la logique de localisation au détriment de l'accessibilité générée par les IRIS.

- dans la seconde, il est tenu compte des prix du foncier, mais en s'assurant que cela ne puisse pas conduire à une diminution de l'accessibilité. La démarche consiste à réaliser une relocalisation standard (sans les prix) afin de connaître les IRIS destinés respectivement à gagner ou perdre de l'emploi. Puis, x emplois sont prélevés dans les IRIS destinés à en perdre, par ordre de priorité d'abord dans les IRIS pour lesquels la ponction va le moins possible faire diminuer l'accessibilité. Enfin, ces emplois sont relocalisés dans les IRIS destinés à en gagner en maximisant, comme dans la première méthode, le ratio " $\alpha_j / (\text{prixFoncier}_j)^a$ ".

Les tests réalisés pour une relocalisation des emplois en 2030 sans l'ADS et pour une capacité maximum des IRIS limitée à 1.5 * le nombre d'emplois présent en 2010 et un seuil minimum de 0.5 * le nombre d'emplois prévu en 2030 montrent que le processus de relocalisation augmente dans chaque cas l'accessibilité considérée (pondérée par la population ou la surface) d'une vingtaine de pour cent pour une relocalisation de moins d'un quart des emplois.

Toutefois, le calcul systématique de l'évolution de l'accessibilité gravitaire correspondante montre que si la maximisation de l'accessibilité pondérée par

SAGEO'2016 – Nice, 6-9 décembre 2016

la population va de pair avec un accroissement de l'accessibilité gravitaire, il n'en va pas de même pour la maximisation de l'accessibilité surfacique ou la prise en compte du foncier. En effet, dans ces deux derniers cas pour compenser les prix fonciers plus importants au centre de l'agglomération et la grande taille et la périphéricité des IRIS aux marges de l'agglomération, le processus de relocalisation y localise de nombreux emplois, diminuant alors d'autant l'accessibilité gravitaire moyenne standard que l'on a vu être « portée » par les IRIS centraux.

5. Conclusion

Les simulations réalisées dans le cas de l'ADS de Lyon ont permis de montrer que l'accessibilité est faite de vitesse, mais aussi est surtout de localisations, puisqu'il est apparu que les gains d'accessibilité dus à l'ADS (2%) sont très faibles au regard de ceux possibles en jouant sur la localisation des emplois (20%). Certes, et c'est logique, la vitesse améliore l'accessibilité, mais les gains obtenus peuvent l'être sans infrastructure avec des relocalisations relativement limitées (97000 emplois). Plus que la construction d'infrastructure, la relocalisation des emplois paraît susceptible de redistribuer en profondeur les niveaux d'accessibilité des lieux et d'en augmenter très sensiblement le niveau. Logiquement plus la marge de liberté sur l'ampleur des relocalisations possibles est grande, plus les niveaux d'accessibilité s'accroissent. Conséquemment à sa définition et à son mode de calcul, la maximisation de l'accessibilité moyenne tend à privilégier les zones les plus centrales géométriquement en termes de relocalisation de l'emploi. Pour éviter ce processus, il est possible de modifier la sémantique de la logique de relocalisation pour maximiser la population ayant un bon niveau d'accessibilité, maximiser le niveau d'accessibilité de chaque mètre carré de la zone d'étude ou encore tenir compte des prix du foncier. L'intégration des prix du foncier actuels dans la simulation rend le modèle plus « opérationnel », même si alors la logique est un peu différente en termes de prospective, puisque cela revient à se baser sur la structure actuelle des prix pour définir les relocalisations, sachant qu'à terme ces relocalisations vont elles-mêmes rétroagir sur les prix.

Si les résultats empiriques sont discutables, puisque de nombreux paramètres d'analyse comme les capacités d'accueil des IRIS, ont été fixés a priori sans base scientifique (elles n'existent pas) ou empirique, ces simulations ont néanmoins permis d'illustrer toutes les capacités exploratoires et perspectives d'utilisation du modèle. Ac-Rel peut ainsi être mobilisé dans la phase d'évaluation de la pertinence de la construction d'infrastructures, qu'elles soient routières ou de transport en commun, mais aussi pour de l'accessibility planning, c'est-à-dire de la planification qui s'appuie sur des modifications de la carte de l'accessibilité. Dans les deux cas, la principale difficulté n'est pas technique, mais réside dans le fait d'être capable d'évaluer le coût de localisation ou relocalisation d'un emploi qui dépend des prix du foncier, mais aussi de la

nature des emplois et de leur « sécabilité ». En effet, les emplois ne sont pas tous indépendants. Il faut donc dans ce cas être capable d'évaluer le coût de relocalisation d'une usine ou un service en entier. Précisons enfin qu'Ac-Rel peut être utilisé pour travailler sur l'accessibilité à n'importe quelle ressource localisée (services, commerces). L'outil est libre et téléchargeable sur le site de l'UMR ESPACE (www.umrespace.org).

Bibliographie

- Boiteux M. (2001). *Transports : choix des investissements et coût des nuisances*. Commissariat Général du Plan, la Documentation Française.
- Bonnafoous, A., Crozet, Y., Mercier, A., Ovtracht, N., Puech, F., Péguy, P-Y. (2010). *MOSART et le projet PLAINSUDD : une plate-forme de modélisation et de simulation de l'accessibilité pour l'aide à la décision et l'aménagement du territoire*. In *Modéliser la ville*. J-P. Antoni (sous la direction de), Economica.
- Crozet, Y., Mercier, A., Ovtracht, N. (2013). *MOSART : Les politiques de mobilité urbaine à la lumière de l'accessibilité*, *Transports*, 479, 9-16.
- Geurs, K.T., Van Wee, B. (2004). *Accessibility evaluation of land-use and transport strategies: review and research directions*. *Journal of Transport Geography* 12, 127-140.
- Glaeser E. (2011). *Des villes et des hommes. Enquête sur un mode de vie planétaire*. Flammarion, 364 p.
- Grand Lyon et Conseil Général du Rhône. (2012). *Débat public Anneau des Sciences, le dossier, 132p*. Disponible sur <http://www.debatpublic-anneau-top.org/docs/documents-maitre-ouvrage/dossier-mo-ads-top-121102-.pdf>, 25 sept. 2013.
- Mignot D., Rosales-Montano S. (2006). *Vers un droit à la mobilité pour tous*. La Documentation Française.
- Piron V., Delons J. (2008). *Pirandello, un nouvel outil d'aide à la décision*. *Transports*, n 449.
- Prud'homme R., Lee C. (1999). *Size, sprawl, speed and the efficiency of cities*. *Urban Studies*, 36(11), pp.1849-58.
- Simmonds D., Echenique M. et al. Bates J. (1999). *Review of Land-Use/ Transport Interaction Models*. Department of the Environment, Transport and the Regions: London, 72p.
- Thomas I. (2012). *Transportation network and the optimal location of human activities. A numerical geography approach*. Cheltenham, Edward Elgar Publishing ltd.

Une base de données pour modéliser l'urbanisation résidentielle en zone côtière

Iwan Le Berre¹, Adeline Maulpoix¹, Marius Thériault², Françoise Gourmelon¹

1. LETG-Brest, GEOMER, IUEM-UBO

Rue Dumont D'Urville, 29280 Plouzané, France

Auteur correspondant : iwan.leberre@univ-brest.fr

2. Université Laval,

Ecole supérieure d'aménagement du territoire et de développement régional

Pavillon Félix-Antoine-Savard, 2325, rue des bibliothèques, Québec, Canada

Résumé. Cette contribution vise à présenter une base d'information géographique destinée à analyser l'évolution spatio-temporelle de l'urbanisation résidentielle dans le Pays de Brest. Son originalité réside dans sa base cadastrale qui permet d'intégrer, à l'échelle de chaque parcelle foncière, les données décrivant les variables environnementales décrites dans la littérature comme influençant les choix résidentiels des ménages. Ce référentiel est renseigné par des données attributaires multi-sources obtenues à travers différents dispositifs, accessibles pour l'enseignement et la recherche, qui font l'objet de l'article. Grâce à cette base de données, nous étudions l'influence du littoral à la fois en tant que facteur d'attraction et de contraintes pour la construction résidentielle notamment liées aux réglementations spécifiques qui s'y appliquent, en particulier la loi Littoral. Une réflexion critique sur les données disponibles montre un certain nombre d'inconsistance et de lacunes qui limite la portée des résultats.

ABSTRACT. This proposal aims to present a geographic database for analyzing the spatiotemporal evolution of the residential urbanization in the Pays de Brest. Its originality relies in its cadastral base that allows to integrate across each land parcel data describing the environmental variables described in the literature as influencing the residential choices of households. This reference data is filled by multi-source attribute data obtained through various devices, available for teaching and research, which are the subject of the article. Through this database, we study the influence of the coastal zone both as an attraction and constraints for residential construction, especially related to the French specific rules applied to coastal urbanization. A critical reflection of the available data shows a number of inconsistency and deficiency which limits the scope of the results.

MOTS-CLES : acquisition de données spatiales, données de référence, loi littoral, planification

KEYWORDS: spatial data acquisition, reference data, Coastal law, land planning

1. Introduction

Contrainte par l'accès aux données de référence, l'analyse de l'évolution urbaine a longtemps porté sur des analyses communales ou supra-communales, fondées sur des bases de données sociodémographiques issues des recensements de la population (Mattei and Pumain, 2000).

Grâce à la numérisation des référentiels géographiques entreprise depuis le milieu des années 1990, associée au mouvement actuel de libéralisation des données publiques, notamment initié par la législation européenne sur l'accès à l'information environnementale (Convention d'Aarhus, 1998) et sur la mise en place d'infrastructures de données géographiques (Directive Inspire, 2007)¹, il est désormais possible d'accéder à une information géographique de référence au format vectoriel. Ces référentiels (RGE et Plan cadastral informatisé notamment) peuvent offrir la possibilité d'améliorer la résolution spatiale des analyses de l'évolution urbaine et d'en accroître la profondeur historique (Geniaux *et al.*, 2009). En particulier, l'urbanisation résidentielle peut désormais être analysée en détail, en étudiant l'évolution des territoires sous l'effet à la fois des politiques publiques et des décisions individuelles prises par les ménages.

Cet article présente une base de données multi-sources décrivant, à l'échelle de la parcelle cadastrale, les facteurs déterminants de la construction résidentielle mentionnés dans la littérature, à savoir les caractéristiques des parcelles, leur environnement immédiat et l'accessibilité aux équipements urbains (Can, 1992). L'objectif poursuivi est de réaliser une modélisation statistique de ces facteurs explicatifs afin de calculer la probabilité que chaque parcelle soit construite durant une période de temps spécifique (intercensitaire ou annuelle). Nos travaux concernent le Pays de Brest, qui constitue un exemple de région urbaine structurée autour d'une ville moyenne, et formée d'un espace périurbain fortement pourvoyeur d'aménités paysagères et de loisirs, liées notamment au littoral. La période étudiée remonte aux années 1960, amorce des phénomènes massifs de périurbanisation et d'étalement urbain en Bretagne (Lebahy and Le Délézir, 2006), jusqu'à nos jours. Notre intérêt porte plus spécifiquement sur la réglementation particulière aux espaces littoraux, notamment la loi Littoral dont l'application donne lieu à des débats récurrents sans parvenir réellement à en établir l'efficacité dans le contrôle de l'urbanisation (Prieur, 2005 ; Eymery, 2014). A l'aide de notre BD et de la modélisation statistique, nous espérons pouvoir démontrer l'effet de cette réglementation sur la construction résidentielle. Le présent article est consacré à la description de la BD développée dans cette perspective. La méthode et les principaux résultats de la modélisation statistique sont présentés en détail dans une autre publication (Le Berre *et al.*, 2016).

¹ Voir la page consacrée sur le site du CNIG : http://cnig.gouv.fr/?page_id=1177

SAGEO'2016 – Nice, 6-9 décembre 2016

2. Pourquoi étudier le bâti résidentiel ?

La fonction résidentielle constitue une préoccupation importante pour tous les ménages en raison de sa dimension à la fois utilitaire (un abri indispensable), affective (lieu d'intimité familiale et individuelle), sociale (l'adresse contribuant notamment à l'identité sociale) et économique (Mercier, 2006). C'est aussi un bien marchand – en France c'est à la fois le plus important actif patrimonial et le premier poste du budget des ménages – et un objet prioritaire des politiques publiques (Driant, 2009; Lacaze, 1996) qui constitue un facteur majeur de consommation des espaces ruraux et naturels et de la transformation de leurs paysages (Bell and Irwin, 2002; EEA, 2006).

Les formes urbaines qui sont associées à la fonction résidentielle dépendent à la fois de la vaste gamme de motivations qui déterminent les choix des ménages et des dispositifs politiques et réglementaires mis en place pour l'encourager et pour la contrôler. Parmi les critères de choix des ménages figurent la proximité résidentielle avec des proches, avec l'emploi, ou avec différents services, l'offre de transports, les qualités du voisinage, le goût pour la vie urbaine ou au contraire l'attrait de la nature (Thériault *et al.*, 2005). Pour ces raisons, les études urbaines accordent une attention particulière aux facteurs sociaux (composition des ménages, démographie) et économiques (revenus, taxation, etc.) en tant que déterminants des capacités ou des choix des ménages pour leur implantation résidentielle (Kestens *et al.*, 2008).

De plus, les choix individuels sont fortement encadrés par les dispositifs législatifs et réglementaires mis en œuvre pour contrôler l'évolution urbaine (Carrion-Flores and Irwin, 2004; Kestens *et al.*, 2008). L'analyse des facteurs socio-économiques doit donc être complétée par celle la réglementation en vigueur qui s'applique de manière différenciée en fonction des territoires : urbains ou ruraux, métropolitains ou périphériques, littoraux, montagne, etc. (Eymery, 2014).

2.1. L'approche foncière, une échelle pertinente pour analyser l'évolution du bâti

Pour l'étude de l'évolution urbaine, l'approche foncière prend toute sa pertinence (Manson *et al.*, 2009), car elle s'appuie sur un référentiel dont les entités (les parcelles cadastrales) correspondent au découpage élémentaire du territoire. D'une certaine manière, la parcelle est en effet l'entité de base à partir de laquelle les acteurs de l'aménagement du territoire et de l'urbanisme appliquent leurs décisions relatives à l'usage des sols (Geniaux *et al.*, 2009). C'est aussi l'échelle à laquelle les ménages élaborent et concrétisent leur projet résidentiel en y intégrant les critères qui leur sont propres, notamment en termes de capacité d'investissement, d'accessibilité à leurs emplois, services (écoles, commerces,...) et aménités (Bell and Irwin, 2002). Par ailleurs, c'est le niveau de base auquel s'applique la réglementation relative à l'affectation des sols, qu'elle se rapporte à l'urbanisme, à l'économie ou à l'environnement. Elle correspond notamment, en France tout au moins, à l'entité à laquelle s'appliquent les permis de construire avec toutes les contraintes qui y sont

associées (affectation des sols, coefficient d'occupation du sol, règles architecturales et paysagères, etc.). Enfin, c'est un niveau très pertinent en termes de description fine de l'usage des sols, dans la mesure où les espaces résidentiels périurbains présentent souvent de faibles densités qui ne sont pas toujours bien décrites par d'autres sources (Bhatta, 2012 ; Sparfel *et al.*, 2011).

L'échelle de la parcelle permet ainsi de relier les décisions individuelles d'usage du sol avec les formes urbaines qu'elles contribuent à produire (Irwin *et al.*, 2009). De plus, grâce à l'enregistrement des mises en construction, elle permet de reconstituer l'historique de l'étalement urbain et plus globalement de la vaste transformation socio-économique, et par conséquent territoriale, qui va profondément transformer les paysages de la France à partir des années 1950 (Dézert, 1991). Enfin, un autre apport des données foncières réside dans la possibilité d'analyser les territoires à des échelles variées en fonction des problématiques traitées, qu'elles soient locales ou intercommunales notamment en lien avec l'évolution des agglomérations (Rousseaux, 2009), ou même régionales (Vimal *et al.*, 2012).

2.2. Méthode d'analyse : la modélisation statistique

L'enjeu que constitue l'étalement urbain pour l'aménagement du territoire explique l'intérêt que lui portent des disciplines variées (géographie, économie, aménagement, écologie du paysage) dont une partie s'est fédérée depuis près de 20 ans dans le cadre des *Land Change Sciences* (Lambin and Geist, 2006). Ces sciences ont très précocement trouvé dans la modélisation spatiale les méthodes et les outils lui permettant d'observer et d'analyser les dynamiques spatiales liées aux changements observés. En retour, elles ont apporté une contribution très significative au développement des outils et des méthodes de la géomatique (Agarwal *et al.*, 2002).

Au sein des familles de modèles souvent distinguées dans la littérature (Agarwal *et al.*, 2002; Lambin et Geist, 2006), on peut retenir d'une part les méthodes visant à dégager des règles de transition, et d'autre part celle qui visent à évaluer l'impact vraisemblable de scénarios d'évolution (Le Berre *et al.*, 2016) et à explorer de possibles futurs types d'occupation et de paysages (Houet et Gourmelon, 2014). Les premières peuvent d'ailleurs être intégrées dans les simulations conduites avec les secondes. La première famille comprend les modèles empiriques et statistiques dans lesquels les changements observés sont liés à des variables exogènes. Parmi les modèles employés, les méthodes de régression constituent une approche particulièrement pertinente (Kestens *et al.*, 2008; Kirk *et al.*, 2011) dans la mesure où elle permettent de tester la contribution relative de chaque variable au modèle, tout en contrôlant leurs rétroactions du fait qu'elles peuvent être associées les unes aux autres (Le Berre et al., 2016). Concrètement, il s'agit d'établir la relation entre la date de construction d'une parcelle et ses caractéristiques spatiales et environnementales.

SAGEO'2016 – Nice, 6-9 décembre 2016

3. Méthodologie d'élaboration de la BD

L'élaboration de la base de données repose sur trois étapes : 1) l'évaluation de l'état de développement de chaque parcelle (déjà construite ou disponible) année par année ; 2) la description du contexte de l'évolution urbaine et périurbaine des points de vue physique, environnemental et réglementaire ; 3) le filtrage des parcelles disponibles en fonction des restrictions d'usage du sol liées à la réglementation en vigueur et à certains aménagements, pour ne conserver, année après année, que les parcelles aptes à la construction.

3.1. Sources de données exploitées

3.1.1. Référentiel cadastral

Administré par la Direction générale des Finances Publiques (DGFIP), le cadastre fournit « un inventaire exhaustif et permanent, descriptif et évaluatif de la propriété foncière » qui est progressivement numérisé depuis 1993 (CERTU, 2006). Il contient une documentation graphique, le Plan cadastral informatisé (PCI), et une documentation littérale, la matrice cadastrale, appelée base MAJIC (Mise A Jour des Informations Cadastreales). Le PCI comporte différents objets : le parcellaire (parcelles, numéros de parcelles, signes de mitoyenneté, ...), le bâti et certains détails topographiques (cimetières, voies ferrées, ...). La base MAJIC comporte plusieurs tables : propriétaires (identification des personnes redevables ou exonérées des différentes taxes), propriétés bâties (identification et évaluation des locaux), propriétés non bâties (informations relatives aux parcelles non bâties, ...).

Dans le cadre de cette étude, deux jeux de données anonymisés ont été obtenus auprès des services de l'Etat : du PCI ont été extraites l'ensemble des parcelles cadastrales du Pays de Brest, qui ont été renseignées par leur date de mise en construction extraite de la base MAJIC.

3.1.2. Référentiel à grande échelle

Le RGE, produit et entretenu par l'IGN², est composé de bases de données vectorielles et matricielles. Il est exploitable à des échelles de l'ordre du 1/10 000 ou supérieures, donc compatibles avec le référentiel cadastral.

La base de données topographique ou BD TOPO® est constituée de plusieurs couches d'information thématique vectorielles qui fournissent une description détaillée de l'emprise et de certaines caractéristiques des voies de communication, des réseaux de transport, de l'hydrographie, de lignes et limites diverses, des bâtiments, de la végétation et de l'altimétrie. Des couches thématiques de la BD TOPO® ont été extraits les thèmes suivants : limites communales, réseau routier, ainsi que certaines entités utiles à la production des filtres d'inconstructibilité.

Le RGE comprend également la BD PARCELLAIRE® qui constitue en quelque sorte la version géométriquement correcte du PCI, mais dont la richesse thématique

² <http://professionnels.ign.fr/rge>

reste limitée à quelques descripteurs. Par conséquent, elle n'est pas été utilisée dans cette étude.

3.1.3. Localisation des services publics et commerciaux

Le temps d'accès aux écoles ou aux pôles urbains de différents niveaux figurent parmi les facteurs déterminant la construction résidentielle (des Rosiers *et al.*, 2001). A défaut de pouvoir disposer de données décrivant l'évolution de la carte scolaire, nous avons pu bénéficier d'une base de données entretenue par le Ministère de l'Education Nationale et contenant l'ensemble des établissements scolaires français, sur la période 1963-2012, ainsi que leur date d'ouverture ou de fermeture. Pour le Pays de Brest, cette base contient 410 établissements, tous types confondus (primaire, secondaire ou supérieur ; privé et public). Plus de 64% des établissements ont été ouverts au début de notre période d'étude (entre 1962 et la fin de l'année 1974). Le rythme s'est ensuite ralenti avec moins de 10% d'ouverture d'établissements pour chacune des périodes suivantes.

Concernant, les services commerciaux, le rapport de présentation du SCOT du Pays de Brest propose une analyse de l'organisation du territoire par les services marchands et non marchands, effectuée par l'ADEUPA³. Elle définit ainsi une hiérarchie reposant sur des pôles de trois niveaux selon leurs fonctions urbaines. Une couche de points localisés au centre-ville de chacun de ces pôles a été produite.

3.1.4. Zones inconstructibles

Afin de ne prendre en compte que les parcelles potentiellement constructibles, ont été exclues – outre les parcelles déjà bâties – celles qui ne peuvent être construites à cause de trois types de contraintes.

a) L'utilisation du sol incompatible ou restreinte comprend les grands équipements de type cimetières, emprises ferroviaires, portuaires et aéroportuaires. Les emprises correspondantes ont été extraites de la BDTOPO®. Compte tenu de la particularité brestoise dans ce domaine et de l'importance des périmètres concernés, les sites militaires ont fait l'objet d'investigations particulières pour pouvoir en dater la mise en réserve. Nous avons procédé de même pour les carrières.

b) Les zones de protection des risques concernent essentiellement les établissements classés à risque industriel et les sites SEVESO qui se traduisent par une exclusion réelle de la construction. Dans le Pays de Brest, les zonages mis en place pour la prévention des risques naturels sont relativement limités et ont été mis en place postérieurement à l'urbanisation des sites concernés. Les données correspondantes sont produites par la DREAL via l'IDG GéoBretagne⁴.

c) La protection de l'environnement et les zones de conservation regroupent différents outils qui contraignent l'urbanisation : arrêté de biotope, terrains du Conservatoire du littoral, espaces naturels sensibles, forêts, périmètres de captage d'eau, réserves nationales, sites classés. Les couches d'information géographique

³ Agence d'urbanisme de Brest Bretagne, <http://www.adeupa-brest.fr/>

⁴ <http://cms.geobretagne.fr/>

SAGEO'2016 – Nice, 6-9 décembre 2016

correspondantes sont, pour la plupart, produites par les services du Ministère de l'Environnement et mises à la disposition des usagers par Carmen⁵.

3.1.5. *Données socio-économique*

L'Insee produit et diffuse des données statistiques relatives à divers aspects économiques, démographiques, et sociaux sur l'ensemble du territoire français⁶. Une part importante de ces données est produite à l'occasion des recensements pour lesquels des données harmonisées existent depuis 1968. Jusque récemment, ces données étaient diffusées seulement à l'échelle communale et, pour les communes de plus de 10 000 habitants, selon un découpage infra-communal : les Iris (Ilots Regroupés pour l'Information Statistique). Depuis quelques années, des données carroyées peuvent également être obtenues à la résolution de 200 m ou de 1 km. Elles offrent donc la possibilité d'approches détaillées à des échelles infra-communales, mais ne concernent que les résultats des recensements les plus récents. Certaines données de l'INSEE sont accessibles au grand public par internet. En tant qu'établissement de recherche, nous avons bénéficié du Réseau Quetelet, une des composantes de la très grande infrastructure de recherche PROGEDO ayant pour but la PROduction et la Gestion des Données en sciences humaines et sociales⁷.

3.2. *Elaboration de la base de données*

Plusieurs procédures d'analyse spatiale, de simulation de déplacements sur les réseaux routiers, de géotraitements et de filtrages ont été effectuées sur ArcGIS et ses extensions pour calculer et mettre à jour les attributs de chaque parcelle au début de chaque période de temps. La procédure générale est décrite dans la figure 1.

Chaque parcelle est décrite par les caractéristiques suivantes : taille, date de construction, altitude moyenne. Le modèle numérique de terrain de la BD Alti® à la résolution de 25 mètres a été utilisé pour calculer la pente moyenne de chaque parcelle, calcul réalisé à l'aide de l'extension *Spatial Analyst* d'ArcGIS. La pente moyenne des parcelles doit être contrôlée car les coûts de construction augmentent lorsque le terrain est accidenté, ce qui conduit à une plus grande probabilité de construire sur des terrains plats.

L'environnement est caractérisé par trois distances euclidiennes (à la route principale la plus proche, à la zone d'habitation la plus proche au début de la période et à la mer), et par des critères d'accessibilité.

La distance aux routes (calculée en utilisant la BD TOPO®) est incluse dans le modèle car l'accès au réseau routier est une condition préalable à la construction et que sa proximité permet de réduire les coûts d'infrastructure. L'effet de la proximité des zones déjà urbanisées est susceptible d'augmenter avec le temps, par le renforcement de l'influence de politiques d'aménagement du territoire plus strictes

⁵ <http://carmen.naturefrance.fr/>

⁶ <http://www.insee.fr>

⁷ www.progedo.fr

Urbanisation résidentielle en zone côtière

dans les dernières périodes. Il s'agit donc d'un des attributs variables dans le temps qui doivent être calculés au début de chaque période pour tenir compte des évolutions précédentes. La distance à la mer vise à tester l'effet de la proximité de la mer sur la construction résidentielle.

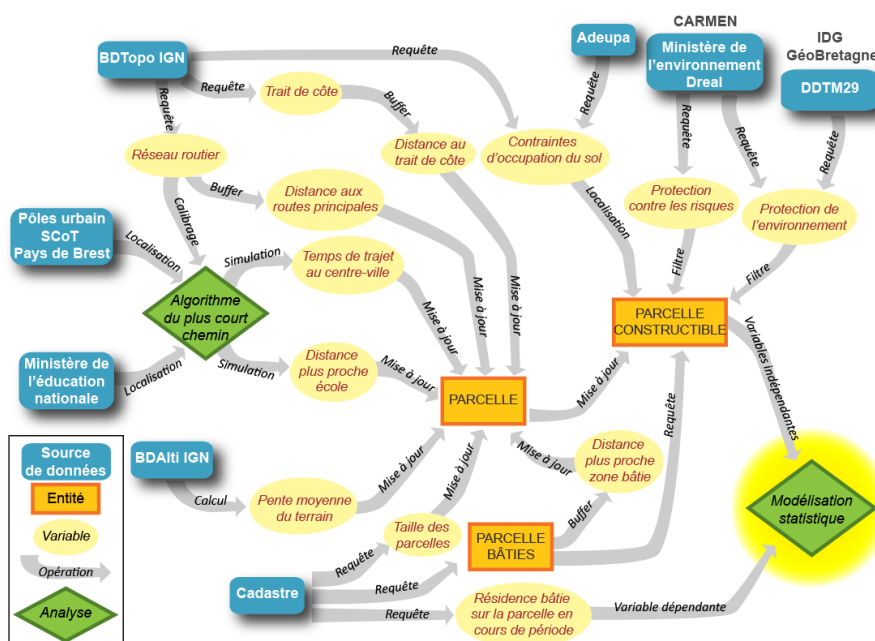


Figure 1 – Procédure générale d'intégration des données sur la construction résidentielle dans le référentiel cadastral

L'accessibilité aux écoles et aux équipements urbains implique la simulation des déplacements sur le réseau routier. Son contrôle est crucial pour refléter les théories sur les rentes de situation et les lieux centraux. L'accès aux écoles a été simulé à pied (algorithme de la plus courte distance à la marche) tandis que l'accès aux centres urbains repose sur le temps du trajet en voiture. Le réseau routier est topologique, avec une impédance et des directions. Les liaisons routières ont été sélectionnées à partir des couches appropriées de la BD TOPO® ; l'impédance a été créée en utilisant les vitesses maximales de flux libre telles que proposées par le CERTU (2007) pondérées par des ralentissements dans les virages. Les restrictions de circulation (sens unique, voies non carrossables) sont paramétrées à partir des variables renseignées dans la BD TOPO®. Les simulations ont été effectuées sous ArcGIS avec l'extension *Network Analyst* (Thériault et al, 1999). Pour chaque période de temps, les écoles actives ont été identifiées à partir des données fournies par le Ministère de l'éducation nationale. Des simulations indépendantes ont été réalisées pour chaque période en tenant compte des fermetures et des ouvertures d'écoles.

SAGEO'2016 – Nice, 6-9 décembre 2016

Enfin, dans le but d'examiner l'effet de la hiérarchie urbaine sur la demande de logements, l'accessibilité automobile aux trois niveaux de centres urbains définis dans le SCOT du Pays de Brest, a été calculée.

Les parcelles disponibles (non bâties) ont été sélectionnées et filtrées (cf. contraintes en 3.1.4.) pour définir celles qui étaient constructibles en début de période. Lorsqu'une restriction est permanente, ou s'applique à partir d'une date donnée, les parcelles concernées sont exclues du modèle statistique. De même, les parcelles déjà construites au début de chaque période de temps sont exclues de l'analyse.

4. Résultats : une BD pour analyser la construction résidentielle

La base de données ainsi constituée est structurée sous forme d'une couche d'information géographique dont la table attributaire est ensuite exploitée par des méthodes de modélisation statistique. Elle contient 415 436 parcelles décrites par trois catégories de variables (tab.1).

TABLE 1. Variables décrivant les parcelles foncières du Pays de Brest (1968-2008)

Catégorie	Variable	Source de la donnée
Caractéristiques des parcelles	Année de mise en construction	PCI / MAJIC
	Superficie	Calcul à partir du PCI
	Pente moyenne	Calcul à partir de la BD ALTI IGN
Environnement et accessibilité	Distance à la mer	Calcul à partir du trait de côte Histolitt
	Distance au bâti préexistant	Calculée à partir des zones bâties (PCI)
	Distance au réseau routier	Calculée à partir de la BD TOPO IGN
	Temps d'accès automobile aux centres urbains	Modèle d'accessibilité (Network Analyst / BD TOPO IGN / Pôle urbains SCOT)
	Temps d'accès piéton aux établissements scolaires	Modèle d'accessibilité (Network Analyst / BD TOPO IGN / BD établissements scolaires Ministère de l'Education Nationale)
Filtres	Terrains inconstructibles exclus de l'analyse	BD TOPO IGN (cimetières, aéroports, ports), DDTM29 (terrains militaires, Sites SEVESO, Plans de prévention des risques technologiques et des risques inondation), MAJIC (carrières), Adeupa (Zones d'activité)
	Réglementation environnementale	CARMEN/DREAL (arrêtés de protection de biotope, réserves naturelles, forêts, sites classés), GEOBRETAGNE (terrains du Conservatoire du littoral, espaces naturels sensibles), DDTM29 (périmètres de captages d'eau)

Parmi les caractéristiques des parcelles, l'année de mise en construction constitue la variable dépendante de nos analyses (celle que l'on cherche à expliquer).

Urbanisation résidentielle en zone côtière

50869 parcelles étaient déjà construites en 1968, pour 124218 en 2008. On constate donc que 60,6% du bâti résidentiel est postérieur à 1968 et que la part de l'occupation résidentielle s'est accrue de 3,3% (5217 ha) à 8,5% (13549 ha) de la surface cadastrée entre 1968 et 2008, soit 2,5 fois plus. Le rythme de construction montre des cycles qui peuvent être corrélés à une combinaison complexe de paramètres conjoncturels locaux – démographiques, restructuration de la Défense nationale (Le Roy, 2009) – et nationaux : crises économiques, taux d'intérêt des prêts immobiliers, coûts de la construction (Friggit, 2014). On voit par-là que le Pays de Brest est arrimé aux dynamiques nationales et régionales, malgré un effet de décrochage perceptible (Lebahy and Le Délézir, 2006) qui pourrait se matérialiser à travers la plus faible croissance de construction après 2000.

Les variables indépendantes (ou explicatives) concernent l'accessibilité aux centres urbains (en voiture) et aux écoles (à pied) ; la distance aux routes principales, aux zones bâties les plus proches et au trait de côte. Une partie de ces variables est utilisée pour contrôler les effets connus sur la probabilité d'une parcelle à se construire. D'après notre modélisation, la distance au bâti apparaît comme le facteur le plus déterminant. Son influence se renforce au fil du temps, ce qui tend à démontrer une rigueur accrue dans l'aménagement résidentiel, conformément au principe de continuité du bâti et d'économie de la consommation de territoire. De même la proximité du littoral apparaît parmi les facteurs les plus structurants. Il exerce en effet une attraction particulière sur la construction résidentielle puisque, au cours de la période considérée, sa part progresse de 3,8% à 9,6% dans les communes littorales, de 6,4 à 15,2% dans la bande de 300 m et de 7,7 à 14,7% dans celle de 100 m.

Enfin, plusieurs variables sont intégrées comme filtres d'inconstructibilité (tab. 1) en tant qu'occupation du sol incompatible avec la construction résidentielle, au titre de la prévention des risques ou de la protection de l'environnement. Lorsque la date d'implantation ou de mise en place de ces équipements ou dispositifs a pu être identifiée, les filtres ne s'appliquent qu'à compter de celle-ci. De 1968 à 2008, ces filtres conduisent à une diminution du nombre de parcelles constructibles de près de 330 000 à environ 240 000. Dans le Pays de Brest, la construction résidentielle n'apparaît donc pas contrainte par une pénurie de foncier. En particulier sur le littoral (bande de 300 m), le stock de parcelles diminue (de 44 668 en 1968 à 29 951 en 1999), mais le rythme de construction s'atténue au fil des périodes, attestant semble-t-il du rôle croissant de la réglementation spécifique qui s'y applique.

5. Discussion et conclusion

La numérisation des données et la transcription dans le droit français de la législation européenne sur l'accès aux données environnementales ont véritablement impulsé un mouvement de libération de l'information géographique. Celui-ci permet d'accéder non seulement à l'information géographique de référence la plus appropriée pour étudier l'évolution urbaine à des échelles pertinentes, mais aussi à des données thématiques variées permettant d'enrichir la connaissance. Néanmoins notre étude révèle des limites liées à l'inconsistance ou l'indisponibilité de certaines données.

SAGEO'2016 – Nice, 6-9 décembre 2016

5.1. Evaluation des données disponibles.

Sur notre terrain d'étude, le PCI n'est disponible que depuis 2008 et il n'existe pas d'historique de son évolution. Au-delà de cette date, cette donnée existe sur un support papier extrêmement difficile à exploiter, d'autant que généralement les modifications foncières y ont été reportées au fur à et à mesure des transactions (CERTU, 2006), ce qui rend de fait illusoire une reconstitution historique de l'évolution du parcellaire. Ainsi les analyses doivent-elles être fondées a posteriori sur le cadastre actuel, alors qu'il serait plus juste de les effectuer a priori. De plus, la cohérence topologique du PCI n'est pas systématique, aussi est-il recommandé d'employer plutôt la BD Parcellaire® de l'IGN pour représenter les parcelles foncières. Cependant, dans les deux cas, la qualité de représentation des parcelles, ainsi que celle des bâtiments (emprise en sol) présente d'importantes variations spatiales, également observées sur la BD Topo® ; aucun des référentiels n'étant significativement meilleur que l'autre... Par ailleurs, un décalage de 5 ans peut être observé entre la déclaration de mise en construction et l'intégration effective du bâtiment dans le PCI. Ce décalage, qui peut être constaté par comparaison entre le PCI, l'orthophotographie et le terrain, pose la question de l'actualité des données.

Sur notre site d'étude, aucun produit numérique décrivant l'évolution du réseau routier n'est disponible. Même si la structure générale du réseau a peu évolué depuis les années 1960, ces évolutions peuvent être localement très importantes, en se surimposant à la structure parcellaire préexistante notamment au sein des zones d'aménagement commerciales, administratives ou résidentielles. Considérant la forte densité du réseau routier en France, et particulièrement dans le Pays de Brest, la reconstitution de l'évolution du réseau routier constitue un travail très chronophage qui n'a pu être entrepris dans le cadre de notre étude. Il n'existe pas non plus d'historique des documents de planification urbaine, permettant de reconstituer l'évolution des zonages d'urbanisme depuis leur adoption progressive par les communes du Pays de Brest. Ces zonages n'ont donc pas été intégrés à la modélisation.

Dans la BD MAJIC, qui est consacrée à l'imposition foncière, seul le bâti résidentiel est décrit. N'ayant pu accéder aux dates de mise en construction du bâti non résidentiel (agricole, commercial, industriel), nous avons été contraint d'employer ce bâti en tant que filtre d'inconstructibilité. De fait, les données cadastrales sont révélatrices des limites liées à la production et à l'entretien de données individuelles. Ainsi, les descripteurs des biens immobiliers sont des données déclaratives rarement contrôlées et mises à jour. Des écarts sont donc possibles entre la valeur renseignée et l'état réel du bâti (par exemple pour le nombre de niveaux, les surfaces habitables, etc.).

5.2. Recensement des données complémentaires.

Afin d'étudier plus précisément les choix résidentiels, il serait nécessaire de décrire certaines caractéristiques socio-économiques pouvant influencer les décisions des ménages, comme l'âge, les catégories socioprofessionnelles et la structure familiale. Grâce aux données de recensement, l'évolution du profil de la population municipale peut être reconstruite dans le cadre d'une analyse des effets fixes. Mais, à une échelle

Urbanisation résidentielle en zone côtière

plus fine, l'analyse des caractéristiques individuelles des ménages implique d'accéder aux données personnelles (telles que celles du cadastre). Les données carroyées de l'INSEE peuvent contribuer à combler cette lacune, tout en respectant les libertés individuelles, mais ces données ne sont disponibles que pour la période récente.

De plus, la discrimination des résidences principales et secondaires serait très pertinente car la motivation des ménages pour la construction de l'une ou l'autre diffère sensiblement (Kirk *et al.*, 2011). En particulier les centralités ne sont pas les mêmes ; le facteur d'accessibilité à l'emploi étant remplacée par l'accès aux aménités, ou à certains types de services (de santé plutôt que scolaires par exemple). Cette donnée est effectivement renseignée pour chaque résidence dans la BD MAJIC, mais seulement pour la période récente.

De la même manière, l'accès aux données concernant les valeurs foncières ne peut, pour le moment, se concevoir qu'à travers des modélisations économétriques établies par exemple à travers les données sur les transactions immobilières, accessibles par des bases de données telles que Perval⁸ alimentée par les notaires, ou DVF (Demande de Valeurs Foncières)⁹ de la DGFIP.

Pour renforcer la connaissance des aspects spatiaux de la périurbanisation, la mobilisation de certains facteurs économiques locaux, tels que les coûts ou les changements dans la fiscalité foncière, serait intéressante (Bell et Irwin, 2002). Ces données ne sont cependant disponibles que pour la période récente (milieu des années 1990) et à l'échelle communale, alors que d'importantes variations locales existent en fonction des quartiers. Il nous faut donc bien admettre qu'une certaine part des facteurs déterminant la construction résidentielle nous échappe, et que ces variations locales ne peuvent être intégrées à notre modélisation statistique qu'à travers l'évolution des variables socio-économiques (en tant qu'effets fixes).

Quoiqu'il en soit, on constate que l'accès aux données historiques reste souvent une gageure. Non pas tant parce qu'elles se situent hors des cadres législatifs actuels de diffusion de la donnée publique, mais plutôt parce qu'elles ont été produites à une époque qui n'intégrait pas les principes de transparence et de diffusion appliqués aujourd'hui. Ces données quand elles existent sont restées en l'état (format papier) et s'avèrent par conséquent délicates d'emploi, ou bien leur mise à jour s'effectue en continu, sans qu'un historique soit conservé.

En conclusion, en dépit de certaines lacunes, de nombreuses données sont toutefois mobilisables dans les études relatives à l'évolution urbaine à l'échelle métropolitaine (Rousseau, 2009) ou régionale (Vimal *et al.*, 2012). Localement, leur exploitation par une démarche de modélisation a permis d'analyser l'influence qu'exerce le littoral sur la construction résidentielle, en tant que territoire attractif en liaison avec les aménités qu'il procure d'un point de vue paysager ou récréatif, mais aussi en tant que territoire contraignant du fait des dispositifs réglementaires spécifiques qui s'y appliquent (Le Berre *et al.*, 2016). A court terme, nos travaux s'attacheront de fait à affiner la restitution spatiotemporelle de ces dispositifs, de

⁸ <http://www.immoprix.com/ipInfo.php>

⁹ <http://guide-dvf.fr/>

SAGEO'2016 – Nice, 6-9 décembre 2016

manière à mieux comprendre la manière dont ils influencent la construction résidentielle.

Remerciements :

Ce travail a bénéficié du soutien de l'Agence Nationale de la Recherche dans le cadre du programme "Investissements d'avenir" (ANR-10-LABX-19-01). Nous souhaitons également remercier les organismes qui nous ont gracieusement fourni certains jeux de données : Brest Métropole Océane, l'ADEUPa, la DDTM29, le réseau Quetelet et le Ministère de l'éducation.

Bibliographie

- Agarwal, C., Green, G.L., Grove, J.M., Evans, T., Schweik, C., 2002. A Review and Assessment of Land Use Change Models dynamics of space, time, and human choice U.S.D.A., Forest Service, Northeastern Research Station, Newton Square, PA.
- Bell, K.P., Irwin, E.G., 2002. Spatially explicit micro-level modelling of land use change at the rural-urban interface. *Agricultural Economics* 27, 217-232.
- Bhatta, B., 2012. Urban Growth Analysis and Remote Sensing - Review of Literature, in: *Urban Growth Analysis and Remote Sensing*. Springer Netherlands, Dordrecht, pp. 9-32.
- Can, A., 1992. Specification and estimation of hedonic housing price models. *Regional science and urban economics* 22, 453-474.
- Carrion-Flores, C., Irwin, E.G., 2004. Determinants of Residential Land-Use Conversion and Sprawl at the Rural-Urban Fringe. *American Journal of Agricultural Economics* 86, 889-904. doi:10.1111/j.0002-9092.2004.00641.x
- CERTU, 2006. Données cadastrales et immobilières : réflexion à partir des retours d'expérience de partenariat DGI et services déconcentrés de l'Équipement. Centre d'Étude sur les Réseaux les Transports, l'Urbanisme et les constructions publiques.
- Des Rosiers, F., Lagana, A., Theriault, M., 2001. Size and proximity effects of primary schools on surrounding house values. *Journal of Property Research* 18, 149-168. doi:10.1080/09599910110039905
- Dézert, B., 1991. La périurbanisation en France. SEDES, Paris.
- Driant, J.-C., 2009. Les politiques du logement en France. *Les Etudes de la Documentation Française* 181.
- European Environment Agency, 2006. Urban sprawl in Europe: The ignored challenge. Office for Official Publications of the European Communities, Luxembourg.
- Eymery, C., 2014. Du texte à la carte : contribution de la géographie à la traduction spatiale de la "loi Littoral". Application en Bretagne. (Thèse de doctorat de Géographie). Université de Bretagne Occidentale, Brest.
- Friggit, J., 2014. Prix de l'immobilier d'habitation sur le long terme. Conseil général de l'environnement et du développement durable, Ministère de l'écologie, du développement durable et de l'énergie. URL <http://www.cgedd.developpement-durable.gouv.fr/>
- Geniaux, G., Podsleski, C., Leroux, B., 2009. Les données majic et leur valorisation au service de l'observation foncière. *Etudes foncières* 28-32.
- Giloppé, M., Jacquot M., Robin F.-X., Tougard C., 2007. Indicateurs d'accessibilité automobile aux équipements structurants des aires urbaines. CERTU.

Urbanisation résidentielle en zone côtière

- Houet, T., Gourmelon F., 2014. La géoprospective – Apport de la dimension spatiale aux démarches prospectives, *Cybergeo*, doi: 10.4000/cybergeo.26194
- Irwin, E.G., Jayaprakash, C., Munroe, D.K., 2009. Towards a comprehensive framework for modeling urban spatial dynamics. *Landscape Ecology* 24, 1223–1236.
- Kestens, Y., Thériault, M., Des Rosiers, F., 2008. Choix résidentiels des ménages lors de l'acquisition d'une maison unifamiliale, in: Thériault, M., Des Rosiers, F. (Eds.), *Information géographique et dynamiques urbaines*. Hermes science : Lavoisier, Paris, pp. 197–226.
- Kirk, R.W., Bolstad, P.V., Manson, S.M., 2011. Spatio-temporal trend analysis of long-term development patterns (1900–2030) in a Southern Appalachian County. *Landscape and Urban Planning*. doi:10.1016/j.landurbplan.2011.09.008
- Lacaze, J.-P., 1996. *Le logement au péril du territoire*. éd. de l'Aube.
- Lambin, E.F., Geist, H.J., 2006. *Land Use and land cover changes*, Springer ed., the IGBP series.
- Le Berre, I., Maulpoix, A., Thériault, M., Gourmelon, F., 2016. A probabilistic model of residential urban development along the French Atlantic coast between 1968 and 2008. *Land Use Policy* 50, 461–478. doi:10.1016/j.landusepol.2015.09.007
- Le Roy, N., 2009. *Brest, de la ville militaire à la métropole occidentale de la Bretagne : constructions politiques de territoires et productions identitaires* (Thèse de doctorat en Sociologie). Université de Bretagne Occidentale, Brest.
- Lebahy, Y., Le Délézir, R., 2006. *Le littoral agressé, pour une politique volontariste de l'aménagement en Bretagne*. éd. Apogée, Paris.
- Manson, S.M., Sander, H.A., Ghosh, D., Oakes, J.M., Orfield, Jr, M.W., Craig, W.J., Luce, Jr, T.F., Myott, E., Sun, S., 2009. Parcel Data for Research and Policy. *Geography Compass* 3, 698–726. doi:10.1111/j.1749-8198.2008.00209.x
- Mattei, M.-F., Pumain, D. (Eds.), 2000. *Données urbaines*. *Anthropos : Diffusion Economica*, pp. 363–372.
- Mercier, G., 2006. *La norme pavillonnaire: mythologie contemporaine, idéal urbain, pacte social, ordre industriel, moralité capitaliste et idéalisme démocratique*. *Cahiers de géographie du Québec* 50, 207. doi:10.7202/014087ar
- Prieur, L., 2005. *La loi Littoral*. Éd. Techni.cités, Voiron.
- Rousseaux, F., 2009. *Une méthode d'analyse pour mesurer l'impact des documents d'urbanisme sur la maîtrise de l'étalement urbain : l'exemple de La Rochelle, France*. Vertigo.
- Sparfel, L., Le Berre, I., Gourmelon, F., 2011. Evaluation des changements d'occupation des sols en zone côtière à partir de données hétérogènes. *Revue internationale de géomatique* 21, 381–403. doi:10.3166/riig.15.381-403
- Thériault, M., Rosiers, F.D., Joerin, F., 2005. Modelling accessibility to urban services using fuzzy logic: A comparative analysis of two methods. *Journal of Property Investment & Finance* 23, 22–54. doi:10.1108/14635780510575085
- Thériault, M., Vandersmissen, M.-H., Lee-Gosselin, M., Leroux, D., 1999. Modelling Travel Route and Time within GIS: Its Use for Planning. *Journal of Geographic Information and Decision Analysis* 3, 41–55.
- Vimal, R., Geniaux, G., Pluvinet, P., Napoleone, C., Lepart, J., 2012. Detecting threatened biodiversity by urbanization at regional and local scales using an urban sprawl simulation approach: Application on the French Mediterranean region. *Landscape and Urban Planning* 104, 343–355. doi:10.1016/j.landurbplan.2011.11.003

Urban Fabric with Bayesian LINCS. Empirical Evidence from the French Riviera.

Giovanni FUSCO¹, Alessandro ARALDI²

1. Université Côte d'Azur, CNRS, ESPACE
98 Bd Herriot, BP 3209, F-06204, Nice, France
giovanni.fusco@unice.fr

2. Université Côte d'Azur, CNRS, ESPACE
98 Bd Herriot, BP 3209, F-06204, Nice, France
alessandro.araldi@unice.fr

RESUME. Un ensemble d'indicateurs morphologiques est proposé pour identifier les tissus urbains par clustering spatial. L'approche LINCS est préférée aux plus classiques LISA pour mieux intégrer le point de vue du piéton dans la ville. Certains indicateurs morphologiques, comme ceux de couvertures du sol, sont des taux et posent un problème de variabilité avec la taille de la population mère. La correction empirique Bayésienne utilisée en épidémiologie, où la population mère serait la surface de l'unité spatiale, semble mal s'appliquer à la morphologie urbaine, car celle-ci influence le découpage spatial. Des nouvelles corrections Bayésiennes sont ainsi proposées et testées sur les paysages urbains de la Côte d'Azur. Une correction utilisant une fonction sublinéaire de la surface se montre plus apte à réduire l'hétéroscédasticité des taux et à identifier les concentrations de tissu pavillonnaire.

ABSTRACT. A set of morphological indicators is proposed to identify urban fabric using spatial clustering. The LINCS approach is preferred to classical LISA in order to better integrate the point of view of pedestrians moving in the city. Some morphological indicators like land coverage are rates and this poses the well-known problem of rate variability with the base population size. Classical empirical Bayesian correction used in epidemiology, with the spatial unit surface area as base population, seems unfit to the analysis of urban morphology, as spatial units depend from morphological phenomena. New empirical Bayesian corrections are thus proposed and tested on the case study of urban landscapes of the French Riviera. A new Bayesian correction which is a sublinear function of the unit surface proves better able to reduce rate heteroscedasticity and to identify hotspots of individual houses.

MOTS-CLES : Morphologie Urbaine, Taux de Couverture, Tissu Urbain, Réseau Viaire, LISA, LINCS, Correction Empirique Bayésienne, Côte d'Azur.

KEYWORDS: Urban Morphology, Coverage Ratio, Urban Fabric, Street Network, LISA, LINCS, Empirical Bayesian Correction, French Riviera

1. Introduction.

The classical schools of urban morphology (Caniggia and Maffei 1979, Castex et al. 1980, Conzen 1960, for an overview Pinon 1991) have developed a rich methodology to measure and evaluate urban morphology, and namely urban fabric, at the scale of a few blocks or a neighbourhood (Levy 2005). The analyses being carried out manually, their application to the whole urban space of a big city or of a conurbation becomes a daunting task. The use of geoprocessing and LISA/LINC (Anselin 1995, Yamada and Thill 2010) has thus been proposed in the analysis of urban fabric within a large urban area (Araldi and Fusco 2016). The goal is to implement computer-aided chains of treatments to carry out analyses of urban fabric previously limited to small areas. The key to geo-processing is the use of morphological indicators well suited to identify specific traits of urban fabric. More particularly, the approach proposed by Araldi and Fusco (2016) is bottom-up: elementary morphological units are aggregated in order to identify through LISA/LINC analysis of morphological indicators coherent patterns corresponding to particular urban fabrics.

However, when indicators of urban form are rates (for example a land coverage rate), the question arises of the influence of the base population size on rate variability among the spatial units of the analysis. If this variability is overlooked, LISA analysis can produce spurious spatial patterns. This problem is well known in epidemiology and has been addressed through the empirical Bayesian correction procedure (Marshall 1991, Assunção and Reis 1999). Straightforward application of solutions conceived in epidemiology seems nevertheless inappropriate for urban morphological analysis. The main specificity of urban morphology is the link between spatial unit definition and morphological characteristics, which makes the identification of a base population influencing rate variability much more complex.

The main aim of this paper is thus to present these specificities of urban morphological spatial analysis and to explore possible solutions of the Bayesian correction problem. The different solutions will be applied and tested on the case study of the French Riviera, allowing an expert-based feedback on the results of the analyses. The rest of the paper is organized as follows. Section 2 will present the main methodological choices of our research: spatial unit definition, morphological indicators and spatial analysis approach. Section 3 will explain why the classical Bayesian correction methods of epidemiology could be unfit for urban morphological analysis and will propose a few alternatives. Section 4 will present the main empirical results on the case of the French Riviera, when coverage rate of single-family houses are considered. Conclusions and future perspectives will put the research results in the broader context of computer-aided analysis of urban morphology.

SAGEO'2016 – Nice, 6-9 décembre 2016

2. Geoprocessing and LISA/LINCS for Urban Morphology Analysis: Methodological Choices.

2.1. Defining Elementary Spatial Units from a Pedestrian Point of View.

In the Italian, French and English traditional schools of urban morphology, the analysis of urban form, and more specifically the analysis of the urban fabric, has focused on three main aspects (Pinon 1991): (i) the identification of urban form components (urban network, buildings and parcels), (ii) their geometrical description and (iii) the analysis of their spatial relationships. The analyses of traditional urban morphology were normally carried out at the scale of a city neighbourhood, with manual calculations and a focus on the historical process behind observable urban forms. Geoprocessing of urban morphology within a GIS environment has become more widespread in the last twenty years, allowing for larger scales of analyses, but often losing the fine grain of the constituent elements of urban form (like in Berghauer Pont and Haupt 2010, or in Fusco 2016). Our research focuses in particular on the urban street network and the built-up space, which are the aspects of urban form more directly observable by pedestrians moving in urban space. Parcel structure plays a more important role in the historical production of urban form and has been omitted in our research.

In order to identify the elementary spatial units for the analysis of urban form, we consider a new division of urban space resulting from the combination of two elements: the urban street network, a connected set of segments allowing pedestrian movement, and the planar extension of urban space. A generalization of Thysen polygons is thus created around each street segment (topologically delimited by street junctions) to identify the portion of planar space conventionally served by the segment. For several morphological indicators, we only consider a double-sided proximity band within this polygon, in order to approximate visible space (Figure 1).

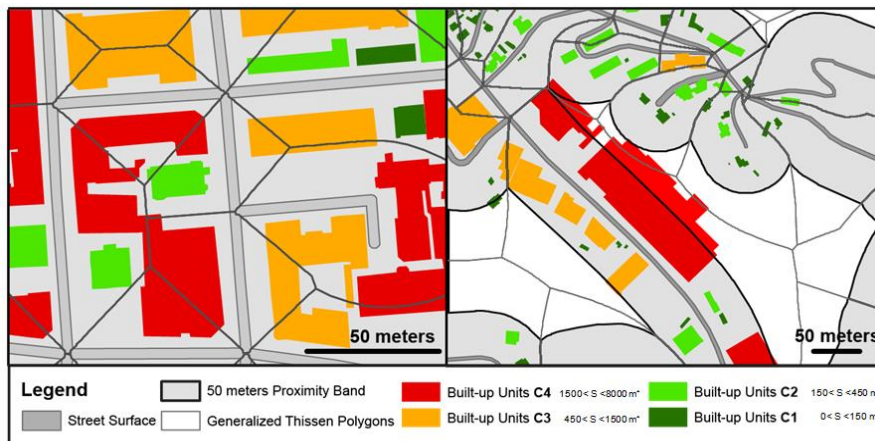


FIGURE 1. Street segment proximity bands and physical footprints of built-up units.

Urban Fabric with Bayesian LINC

The rationale for this spatial unit definition is that a street segment should not be considered the limit, but rather the core of a fragment of urban fabric. This is often the case in European cities where discontinuities in urban fabric normally coincide with double carriageway boulevards (the latter produce two different spatial units in our model) but not with ordinary streets. Moreover, this approach is the most consistent with the pedestrian point of view: when standing in public space, people perceive the urban fabric on both sides of the street and not the elements within the four sides of a city block. In our study area, 99 562 elements were thus identified with 98% of street segment lengths varying between 5 m and 920 m (median value is 77.7 m). When considering the 50 m proximity band area, 98% of values range between 25 m² and 77 550 m² (median value being 3 700 m²).

2.2. A Set of Morphological Indicators

As anticipated, the street network morphology and the built-up forms are the main components of urban fabric considered in this research. Nine indicators, obtained through geoprocessing in GIS, were calculated for each spatial unit and cover five main aspects of urban fabric: network morphology, built-up morphology, network-building relationship, site morphology and site-network relationship. The indicators are presented in more detail in Araldi and Fusco (2016).

TABLE 1. A set of morphological indicators to characterise urban fabric.

Component of Urban Fabric Morphology	Indicator	Definition
Network Morphology	Length	Street segment length between two intersections
	Linearity/Windiness	Ratio between segment length and Euclidean distance
	Local connectivity	Number of degree 1 nodes
		Number of degree 4 nodes
Number of degree 3 or 5 and more nodes		
Built-up Morphology	Built-up Type Coverage Ratio	Ratio between 50 m proximity band surface and 0-150 m ² built-up units surface
		Ratio between 50 m proximity band surface and 150-450 m ² built-up units surface
		Ratio between 50 m proximity band surface and 450-1500 m ² built-up units surface
		Ratio between 50 m proximity band surface and 1500-8000 m ² built-up units surface
		Ratio between 50 m proximity band surface and >8000 m ² built-up units surface
Network-Building Relationship	Street corridor effect	Ratio between length of parallel façades and street length in the 10 m proximity band
	Proximity Band Coverage Ratio	Buildings coverage on the 20 m proximity band
	Proximity Band Building Height	Ratio between building vol. and surf. inside 20 m proximity band
Site Morphology	Surface Slope	Ratio between total and high sloped space-unit ($S > 30^\circ$) in 50m proximity band
Network Site Relationship	Street Acclivity	Computed as segment average of $\arctan(\text{slope})$

SAGEO'2016 – Nice, 6-9 décembre 2016



FIGURE 2. *Built-up types corresponding to physical footprint classes on the Riviera.*

Within this paper we will focus on the indicator describing the contribution to the coverage ratio of the 50 m proximity band around each street segment by the different classes of built-up elements. Built-up elements are either isolated buildings or sets of contiguous buildings. Empirical knowledge lets us establish multiple correspondences between built-up unit physical footprints and possible morphological types within our study area (Figure 2). 0-150 m² built-up footprints are mainly the fact of ordinary single-family independent houses (a) or of their annexes (b: sheds, detached garages). 150-450 m² footprints can correspond to big villas (c), to small series of row-houses (d) or to small block of flats (e). 450-1500 m² correspond mainly to four different typologies: big block of flats (f), small continuously built-up city blocks (i), small retail buildings (h) and long rows of townhouses (g), the latter being particularly rare in our study area. Footprints between 1500 and 8000 m² correspond to big modern housing estates (j), large continuously built-up city blocks (k) and big specialized buildings. Finally, class 5 are exceptional elements in the urban landscape of the study area, i.e. functionally specialized “giant” buildings (l: hospitals, stadiums, convention centres, shopping malls). The indicator is a rate measuring the part of the spatial unit 50 m wide around a street segment covered by each footprint class. The total coverage ratio of the spatial unit is thus the sum of the values calculated for each footprint class.

2.3. *Spatial Statistical Analysis of Morphological Indicators*

Once indicators are calculated for each spatial unit, local spatial clustering techniques are used to identify larger scale urban features (urban fabric). Spatial

clustering indicates where a phenomenon of interest has high/low incidence level, identifying hot/cold spots; many different methods have been developed, in different research fields and perspectives. Local Moran's I indicator of spatial association (LISA, Anselin 1995), based on Moran's I spatial correlation measure (Moran 1948), was identified as a valid geostatistical method. Despite its large application in other research fields, it has so far been used relatively little in the study of urban form (Tsai 2005, Musakwa and Niekerk 2014). In order to test the pedestrian perspective assumption, Araldi and Fusco (2016) analysed and compared the planar application of local Moran's I statistic LISA with the corresponding network-constrained ILINCS (I statistics in Local Indicator of Network-Constrained Clusters, Yamada and Thill 2007, 2010). Their work considered several network and planar neighbourhood depths, always following a topological queen contiguity approach. The case of the street corridor effect indicator was used in order to show that network-constrained LINCS are better suited to capture the feature patterns from a pedestrian point of view. Following this indication, we will carry out spatial analysis of built-up type coverage through a LINCS approach, a network extension of LISA. Adopting a network geometry, the GeoDa software (Anselin 2003) can be used to implement these statistics.

3. The Problem of Population Size in the Spatial Analysis of Rate Values

3.1. Bayesian Correction in Epidemiology

Epidemiology has been one of the first and is still one of the most important fields of application of spatial clustering techniques (Jacquez et al. 1996a, 1996b). Tackling the case of spatial clustering of rate values, Marshall (1991) and Assunção and Reis (1999) highlight that the size of the base population over which the rate is calculated has an impact on the variability of the rate. When dealing with spatial units having strong differences in population size, the variability of the prevalence rate of a given disease in large population units is lower than the variability measured in small population units. This is due to the limits of statistical estimators on small populations. When this dependence of variability on base population size is overlooked, spatial clustering techniques could detect spurious hot/cold spots, linked to the variability of estimations on small populations and not to spatial structure. The same authors thus propose an empirical Bayesian correction taking into account the knowledge of the base population size to correct the *a priori* assumptions on the variability of the rates. This approach, which is now commonly accepted in epidemiology, considers that spatial unit definition (normally administrative units like municipalities in a given region) is independent from the phenomena under investigation (here disease prevalence). Epidemics do not define spatial units and the epidemiological phenomenon (the prevalence rate) can be modelled as a Poisson stochastic process whose outcome (the number of infected people) is directly proportional to the population exposed in the spatial unit. Under this assumption, Empirical Bayesian correction of Moran local I (EBI) is defined as follows (Assunção and Reis 1999):

SAGEO'2016 – Nice, 6-9 décembre 2016

$$z_i = \frac{p_i - b}{\sqrt{a + b/x_i}} \quad (1)$$

$$EBI = \frac{m}{\sum w_{ij}} \frac{\sum w_{ij} z_i z_j}{\sum (z_i - E[z])^2} \quad (2)$$

Where m is the number of spatial units, w_{ij} are the neighbourhood coefficients for units i and j (normally 1 if j is considered neighbour of i , 0 otherwise), z_i are the new deviates for which the Moran local I is calculated and which are derived from the original rates p_i , a and b are Bayesian estimators described in Marshall (1991). The empirical Bayesian correction boils down to reducing the variance of rate values p_i *a posteriori* through standardization coefficients, which have a fixed part and a variable part which is inversely proportional to the population size x_i of each spatial unit.

LISA/LINCS clusters are aimed at verifying whether, on the one hand, the observed rates are the likely result of random outcomes of a spatially homogenous Poisson process with a unique rate or with different rates without spatial structure (null hypothesis) or, on the other hand, if they are the random outcome of a spatial structure of different underlying rates (alternative hypothesis). Bayesian corrections have a direct impact on the calculus of LISA/LINCS clusters: deviations from the study area average values tend to be more significant when they are calculated for units with large populations and less significant when then they are calculated for units with small populations. Accordingly, the usual categories of High-High, Low-Low, High-Low and Low-High spatial patterns tend to be less significant when calculated on units having mainly smaller populations.

3.2. Why could the Analysis of Urban Morphology be Different from the Classical Epidemiological Framework.

At first view, it is easy to conceive a straightforward application of the epidemiological approach to urban morphological rates. The coverage ratio of a given built-up type within a spatial unit would have as base population the surface of the spatial unit. Every square meter of the 50 m proximity band around a street segment is the equivalent of an exposed person in epidemiological terms. In this context, small spatial units should exhibit higher variability than big spatial units, the latter corresponding to long street segments and the former to short street segments. The empirical Bayesian correction of these variabilities based on spatial unit surface amounts to considering as more reliable rates measured around long street segments. But this presupposes that spatial units are independent from the phenomenon under investigation.

Unfortunately, when dealing with urban morphology, spatial unit definition is directly linked to the phenomena under investigation. This is evident in our case as we define spatial units as proximity bands around street segments, which are themselves part of the urban morphology we want to study. Coarser (and often inappropriate) spatial unit definitions like urban neighbourhoods or street blocks (the natural urban grid) also depend on urban morphology. Neighbourhoods are in fact smaller where urbanization is denser and larger where it is less intense. Superposing

Urban Fabric with Bayesian LINC

a conventional grid of decametric/hectometric grain on urban space is very popular in the analysis of land coverage, landscape ecology and, more recently, socioeconomic phenomena. Its use for the analysis of the morphology of urban fabric seems more problematic, as the spatial arrangements of form elements could be partially destroyed by the superimposed observation grid¹. Moreover, we should not see urban space as a coropleth phenomenon. Urban space is the result of the urban colonization of natural and agricultural land. The morphological process is here to be taken into account in order to understand the production of spatial units. As classical morphological analysis shows (Caniggia and Maffei 1979), when a spatial unit approximately defined as we do in our research (land close to a street segment) is relatively saturated (i.e. it has a high land coverage ratio), it becomes economically interesting to intensify the street network in order to develop adjacent natural or agricultural land. Unless strong topographic constraints forbid it, new street segments are added and the average length of the street segments diminishes locally, just as the average surface of the spatial unit (proximity band of 50 m around the street segment). Taking into account that spatial unit definition is linked to morphological phenomena, two alternative hypotheses could be considered against the epidemiological approach.

The first hypothesis considers that new urbanization is historically produced by the repetition of a given building type (with its typical open spaces like interior courts, front and back gardens, parking space, etc.), or of a very limited number of complementary building types. Even if these patterns can later evolve through destruction/reconstruction processes, the contribution of given building types to the land coverage around a street should tend to oscillate within given limits once the unit is totally urbanized. Variability could be much higher when a long street segment proximity band is only partially developed and natural or agricultural land can still be in it (even if variability should fall again for the longest segments, in completely undeveloped land). The understanding of the morphological process could thus suggest an approach that is opposite to the one used in epidemiology. Rates around short and medium-sized segments should be considered as more reliable because the morphological process has produced more stable urban features. Rates around longer segments should be corrected as higher variability characterizes a phase of the morphological process that could be considered as intermediate between natural/agricultural status and full urbanisation. Within this hypothesis of “inverse relation”, we can conceive an empirical Bayesian correction where the base population is not the surface of the area unit, but is inversely proportional to it.

The second alternative hypothesis is a milder revision of the epidemiological approach. The aforementioned link between spatial unit definition and the morphological phenomena would weaken the relationship between the morphological variability and population size (here spatial unit size), without being

¹ Nevertheless, when the focus of the analysis is not the relationship between buildings, street segments and parcels within urban fabric but more abstract aspects of urban form, like fractal dimension of the built-up area, pixel-based analyses are of paramount interest (like in Batty and Longley 1994 or in Frankhauser 1994).

SAGEO'2016 – Nice, 6-9 décembre 2016

able to inverse it. The statistical reduction of morphological rate variability within larger spatial units will be only partially offset by the higher heterogeneity associated with longer street segments in not fully urbanised areas. In this case, a sublinear variation of the total land coverage ratio and of the land coverage ratios by each built-up type could be considered, with an exponent less than one to be applied to the spatial unit surface in order to obtain a convenient EBI.

In the following section we will test these three different approaches to empirical Bayesian correction of ILINCS (EBI-LINCS) using data from the French Riviera.

4. Empirical Evidence from the French Riviera

4.1. The French Riviera: an ideal test area for new morphological analyses

The French Riviera is a conurbation in southern France stretching more than sixty kilometres from the Italian border to the Esterel mountains. Covering around 1500 km², the French Riviera has a population of over one million inhabitants. This space is a unique conjunction of natural and urban landscapes: firstly, the topography, with elevation ranging from the sea level up to 1700 meters of the pre-Alps (passing through hills and valleys differently sloped). Secondly the socio-political and historical influences on the urban planning. Traditional villages are also scattered around three high density urban areas. From east to west, we find: Monaco and its skyscrapers, the most densely populated sovereign nation in the world; the urban agglomeration of Nice with a regular meshed core inspired by the Turin model (Graff 2013), surrounded by hilly and less tightly planned areas. And finally the urban agglomeration of Cannes-Grasse-Antibes characterised by land irregularity together with old city-centres and the car-centred sprawl development of the last 50 years (Fusco 2016). The combination of all these elements produces a sequence of urban centres and peripheral areas of different size and different morphology. This study area gives the opportunity to test the validity of all the morphological indicators of Table 1, of the LISA/LINCS approaches (Araldi and Fusco 2016) and, as far as this paper is concerned, the proposed methodology of empirical Bayesian correction for morphological rates.

4.2. Spatial Patterns of Single Family Houses: Results and Discussion

Within this paper we will focus on the spatial clustering of coverage rates for the smallest footprint class (≤ 150 m²). This is the class which corresponds more strictly to a single building type: ordinary free-standing single-family houses and their annexes. Figure 3 shows the values of land coverage ratios for this building type within the 50 m proximity bands of street segments. The represented area is the central section of the French Riviera, around the city of Nice. The map clearly shows a strong under-representation of this building type in the city-centre of Nice. It is much harder to identify spatial patterns in more peripheral areas based on street connections reflecting the way pedestrians explore urban space. This fully justifies the use of spatial clustering analysis based on the LINCS approach.

Urban Fabric with Bayesian LINCS

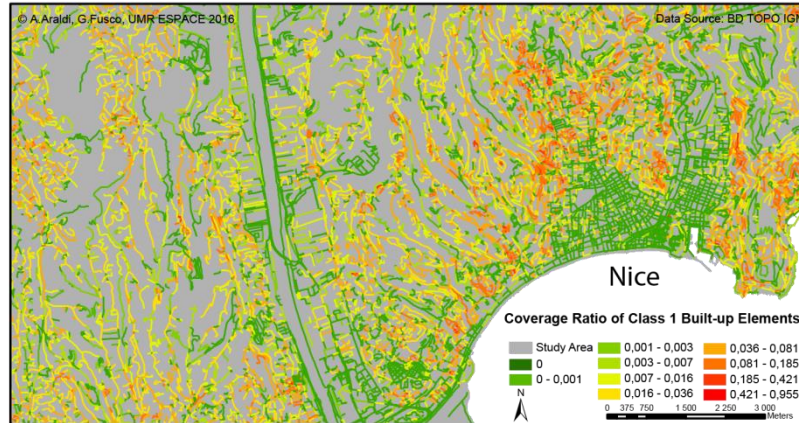


FIGURE 3. Land coverage ratios for single-family houses around the city of Nice.

For this first typo-morphological class, the interpretation of the hot/cold spots of the spatial clustering will be relatively easy. A good spatial clustering should identify as hotspots (High-High) the swaths of residential urban fabric mainly made of individual houses, within planned subdivisions or more spontaneous urbanisation; cold spots (Low-Low) should correspond either to urban fabric made up of other building types or to relatively undeveloped natural and agricultural land. Exceptions within these patterns or transition areas should be identified as Low-High or High-Low features. Finally, more mixed configurations without clear spatial structure should be classified as statistically non-significant in terms of spatial clustering (significance level of 0.05 will be used throughout our analyses). A few preliminary statistical analyses are nevertheless necessary before calculating spatial clustering indicators.

Figure 4.a shows the variation of the variance of single houses land coverage ratio with the increase of the spatial unit surface. The diagram clearly shows heteroscedasticity for the observed morphological rates, pointing to the need of empirical Bayesian correction: the ratio of the variances calculated in the ventiles with maximal and minimal variability is 32.1 and even considering only the 18 ventiles with most homogeneous variance, this ratio is still 7.0. Variance increases over the first three ventiles (with a maximum for the 3rd ventile, corresponding to an average unit surface of 312 m²) and steadily decreases afterwards. This trend invalidates the inverse relation hypothesis: like in epidemiology, and apart for extremely small spatial units (where the possibilities of urbanisation are technically limited to fewer options) rate variability steadily decreases with spatial unit size.

We still have to ascertain whether the classical epidemiological approach or the alternative hypothesis of a sublinear correction is best suited to our empirical data. Figure 4.b and 4.c answer this question. In the former, variance is calculated for empirically Bayesian corrected rates as in the epidemiological approach. Heteroscedasticity remains almost the same. On the contrary, through a Bayesian

SAGEO'2016 – Nice, 6-9 décembre 2016

correction which uses the cubic root of the spatial unit surface (Figure 4.c), we reduce heteroscedasticity considerably: the ratio between the highest and lowest variance is now 13 and even to 3.5 when only the 18 ventiles with most homogeneous variance are considered. This result points to the sublinear correction hypothesis: the morphological process can reduce but not invert the trend of coverage rate variability to decrease with spatial unit surface. Other sublinear corrections were tested within our research (exponents 0.667, 0.5, 0.25 and logarithmic transformation). The cubic root proved the most effective in reducing heteroscedasticity for single-family houses coverage ratio on the French Riviera.

We will however highlight the importance of the residual heteroscedasticity after the empirical Bayesian correction. Once the size of the spatial unit is taken into account (though a sublinear correction), intermediate spatial units still exhibit more rate variability than smaller and bigger ones. The link between the morphological phenomenon and the spatial unit definition is possibly more complex than what a monotonic Bayesian correction with the unit surface can handle. Intermediate spatial units are closely associated with urban fabric. High heterogeneity of urban fabrics in the study area could thus be a possible cause of observed residual heteroscedasticity.

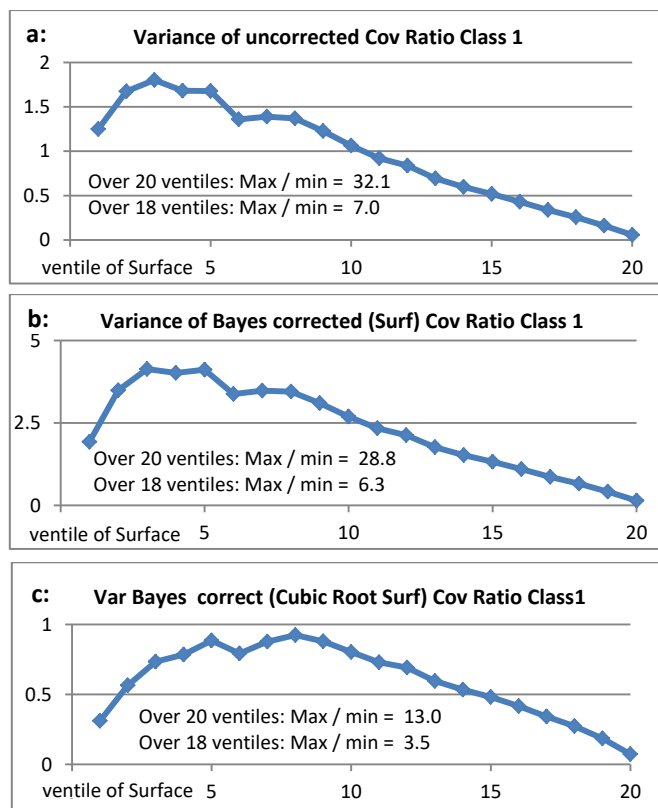


FIGURE 4. Variance of land coverage ratios for single-family houses.

Urban Fabric with Bayesian LINCS

Another important point to be considered before calculating ILINCS statistics is the lack of normality of the empirical distribution of our morphological rates. Griffith (2016) shows how pseudo-significance of Moran I is biased by skewness in data distribution. The same could be said for local Moran I and ILINCS. More particularly, single-family houses coverage ratio is extremely right-skewed on the French Riviera, with 48% of spatial units having the modal value 0, a median value of just 0.0004 and an average value of 0.02, whereas 71% of values are below average (this is not directly visible in Figure 3, where only the central area around the city of Nice is represented). The typical negative deviate from the average value is thus around 0.0004. Positive deviates can be much wider: 0.6% of deviates are two orders of magnitude larger than the typical negative deviate. As the average number of neighbours on each spatial units is around 20 (when a topological radius of 3 is used, see further), a single very large positive deviate can produce a higher than average value in the neighbourhood of a street segment. Now, a random sampling of our empirical distribution will easily obtain at least one very large deviate in 20 independent draws ($p = 20 \times 0.006 = 0.12$, well beyond the significance threshold of 0.05). Conversely, obtaining a lower than average neighbourhood from randomly sampling our empirical distribution is more difficult: almost all draws have to be below the average and no single draw must have a large positive deviate (here p is well below 0.05). As a consequence, when calculating pseudo-significance of ILINCS, HH or LH significance values are underestimated and LL or HL significance values are overestimated.

Given the particular empirical distribution of our data (48% of 0 values), no non-linear transformation of the data could produce a symmetric, almost-normal distribution, not even a rank transformation. Furthermore, EBI correction of rank data would not decrease data heteroscedasticity. As a consequence, ILINCS and EBI-LINCS will be calculated for the original coverage ratio data, and we will be forced to accept spatial clusters with different pseudo-significance levels.

ILINCS were thus calculated following topological contiguity on the street network with a neighbourhood depth of three (which corresponds well to the relatively local features of urban fabric). Values in Figure 5.1 have no Bayesian correction. They identify 32 000 units in Low-Low clusters. In the area represented on the map they correspond to the city-centre of Nice (a), to modern developments in the Var Valley or on the hills, like the IBM centre (b), to natural areas (c) or to old villages. High-High clusters involve more than 10 000 units, both in small subdivisions north of the city-centre (d) and in larger less regular areas on the hills (e). Nevertheless, differences in the variability of rates for different size of spatial units could bias the identification of significant and non-significant spatial clusters.

Figure 5.2 thus proposes a classical empirical Bayesian correction of coverage rates for individual houses, based on the straightforward epidemiological analogy. We observe a big increase of non-significant features. Spatially, significant patterns are severely reduced in urbanised areas (with small or medium-sized spatial units). Low-Low clusters are only marginally reduced: they resist well both in the very connected city-centre and in the undeveloped areas with long street-segments.

SAGEO'2016 – Nice, 6-9 décembre 2016

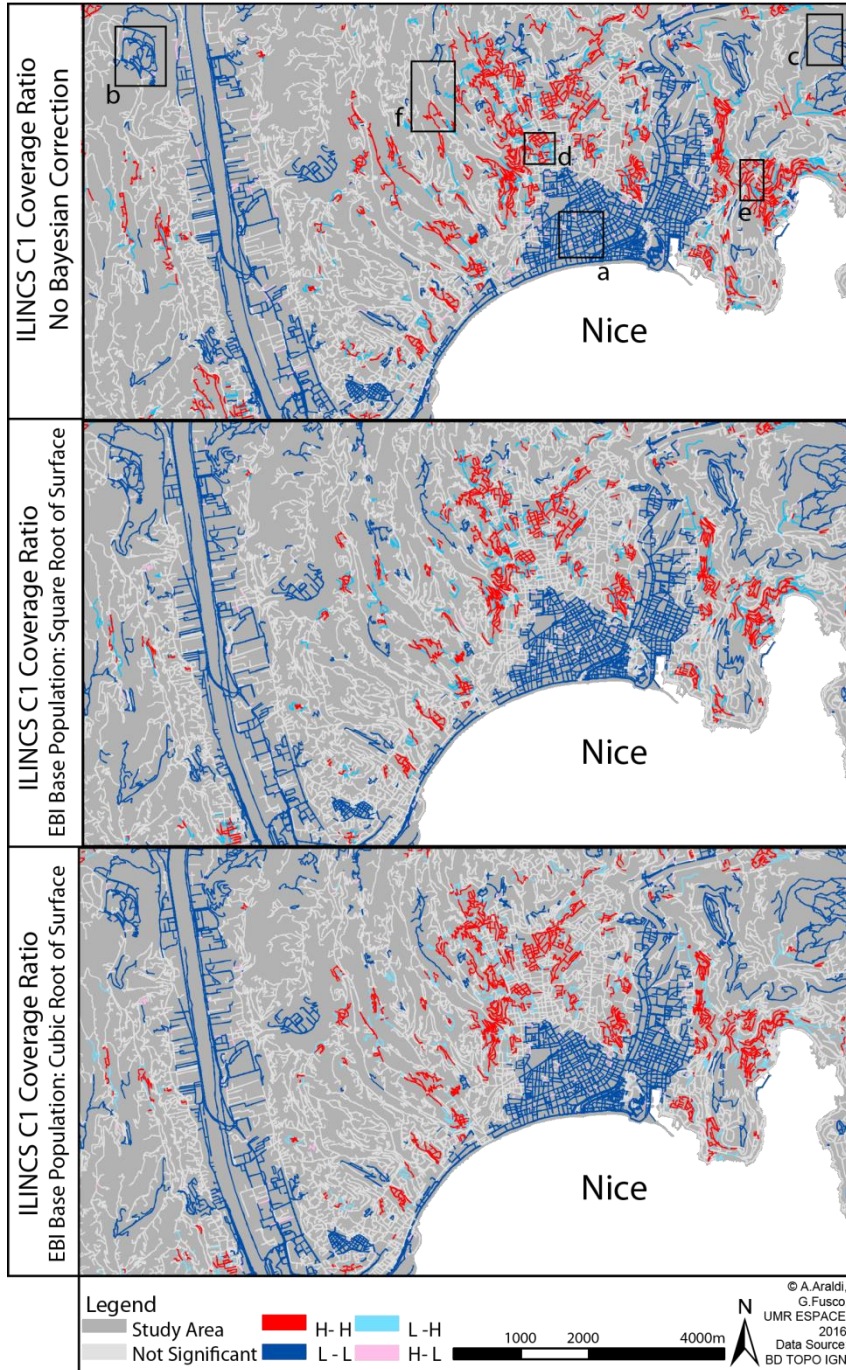


FIGURE 5. *ILINCS* of land coverage ratios for single-family houses.

Urban Fabric with Bayesian LINCS

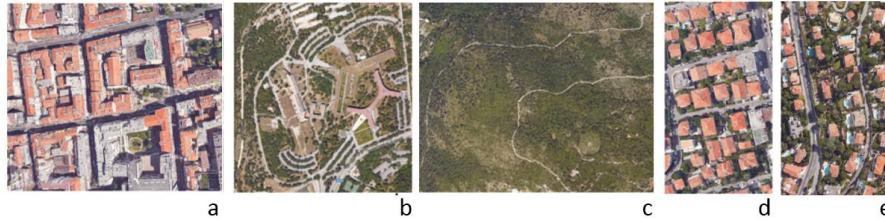


FIGURE 6. Sample of Riviera landscapes within the areas identified in Fig.5.

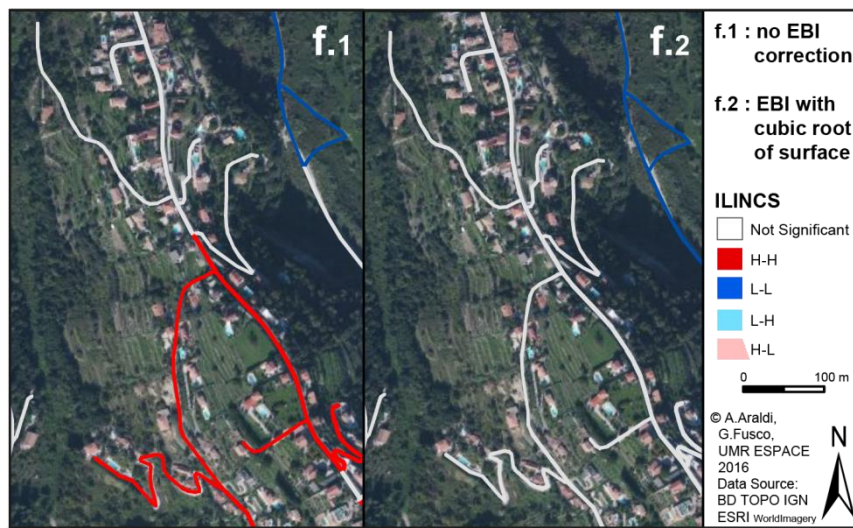


FIGURE 7. The case of a mixed area on the hills around Nice (f in Fig.5).

A few long Low-Low segments even appear and replace shorter ones which are now considered non-significant. The main loss of significant features is observed for the High-High clusters (they diminish of more than 60%): whenever street segments are not particularly well connected, the lower corrected values of short-segments deviates make whole High-High clusters become non-significant. Only the most typical urban subdivisions (d) or suburban concentrations of single-houses remain significant, many tree-like hotspots of individual houses (like e) disappear because of the Bayesian correction to the small branches of the main streets.

Figure 5.3 results from the application of our proposed sublinear empirical Bayesian correction. Overall, compared to Figures 5.1 and 5.2, there is a much slighter reduction of non-significant features. Spatially, most significant High-High and Low-Low clusters are maintained both in urbanised areas and in the less developed hilly peripheries. Sometimes, though, a slight penalization of the deviates of the smaller spatial units, reclassifies as non-significant a few structures that were previously considered as High-High. This is the case, for example, of relatively

SAGEO'2016 – Nice, 6-9 décembre 2016

heterogeneous areas on the hills around the city-centre of Nice (like f). The aerial image in Figure 7 shows no noticeable difference between the southern and the northern section of the area, with class 1 elements (single-family houses) intermingled with class 2 elements (big villas and a few semidetached houses). However, in the south many of the short street segments happen to lead to single-family houses only. In the north, they happen to connect more heterogeneous elements. The uncorrected ILINCS reveal the existence of a significant High-High structure in the south, whereas the corresponding EBI-LINCS, by overlooking the contribution of the smallest street segments, rightly identifies a seamless suburban area with non-significant overrepresentation of class 1 elements. The area in section f is thus clearly distinguished from areas of more pronounced homogeneous overrepresentation of single family houses like d and e.

5. Conclusions and Perspectives.

Specific knowledge of the morphological process led us to critically assess the classical empirical Bayesian correction of rate variability used for LISA/LINCS in epidemiology, because spatial unit definition is linked to the morphological process. We could thus formulate two new proposals: an EBI for morphological rates where the base population is the inverse of the spatial unit surface and an EBI where the base population is a monotonic sublinear function of the spatial unit surface.

The case study of urbanisation in the French Riviera let us verify empirically that the proposed sublinear EBI of individual house coverage rate is better able to reduce heteroscedasticity in the empirical data, to correctly detect typical suburban fabric as hotspots of single-family houses and central-city areas, functionally specialised areas, as well as scarcely developed areas as cold spots. Residual heteroscedasticity is nevertheless still important after the EBI correction. The link between morphological phenomena and the spatial unit definition is possibly more complex than what a monotonic Bayesian correction with the unit surface can handle, especially in the presence of high heterogeneity of urban fabrics, like in our study area.

The spatial analysis of the coverage rate of single houses is only a first step of the research. Other physical footprint classes, having multiple correspondences to morphological types, will have to be cross-analysed with single-family houses. Built-up volumes could then be used to better assess the correspondence between physical footprints and morphological types. The results presented in this paper make us however believe that the best way to take into account size-related variability of the coverage rates will be through an empirical Bayesian correction using as base population a sublinear function of the spatial unit surface.

A complete bottom-up identification of urban fabric types, coherently taking into consideration the pedestrian point of view, will of course be possible only through the combined spatial analysis of all indicators identified in our research design. Street network morphology and network-building relations could be essential to solve the ambiguous correspondences between physical footprints and morphological types. Araldi and Fusco (2016) showed the pertinence of the LINCS approach, already proposed by Yamada and Thill (2007, 2010) in other research

fields, for the spatial analysis of urban morphology. This paper proposes an alternative way of applying empirical Bayesian corrections to morphological rates. These two indications should thus guide the cross-analysis of the whole indicator set.

References

- Anselin L. (1995). Local Indicators of spatial association. *Geog. Analysis*, 27(2): p. 93-115.
- Anselin L. (2003). *GeoDa0.9users guide*. Phoenix, Spatial Analysis Laboratory.
- Anselin L., Kim W.Y., Syabri Y. (2003). Web based Analytical tools for the exploration of spatial data. *Journal of Geographical Systems*. 6(2): p. 197-218.
- Araldi A., Fusco G. (2016). Urban Form from the Pedestrian Point of View: Spatial Patterns on a Street Network. *INPUT 2016*, Turin, Sept. 14th-15th 2016, 6 p. (proceedings in press).
- Assunção R., Reis E. (1999). A New Proposal to Adjust Moran's I for Population Density. *Statistics in Medicine*, vol. 18, p. 2147-2162.
- Batty M., Longley P. (1994) *Fractal Cities: Geometry of Form and Function*, Academic Press
- Berghauser Pont M. Y., Haupt P. (2010). *SPACEMATRIX, Space, Density and Urban Form*. Rotterdam, NAI Publishers.
- Caniggia G., Maffei G.-L. (1979). *Lettura dell'edilizia di base*. Firenze, Alinea.
- Frankhauser P. (1994). *La fractalité des structures urbaines*, Economica, Paris.
- Fusco G. (2016). Beyond the Built-up Form / Mobility Relationship: Spatial Affordance and Lifestyles. *Computer, Environment and Urban Systems*. 60 (2016): p. 50-66
- Graff P. (2013). *Une ville d'exception. Nice dans l'effervescence du 20^e siècle*. Serre, Nice.
- Griffith D.A. (2010) The Moran coefficient for non-normal data, *Journal of Statistical Planning and Inference*. 140 (2010): p. 2980-2990
- Jacquez G M, Grimson R, Waller L, Wartenberg D (1996a). The analysis of disease clusters: Introduction to techniques. *Infection Control and Hospital Epidemiology*, 17, p. 385-397
- Jacquez G M, Waller L, Grimson R, and Wartenberg D (1996b). The analysis of disease clusters: State of the art. *Infection Control and Hospital Epidemiology*, 17, p. 319-327
- Lévy A. (2005). Formes urbaine et significations: revisiter la morphologie urbaine. *Espace et sociétés*, 122(4): p. 25-48.
- Marshall R. J. (1991). Mapping Disease and Mortality Rates using Empirical Bayes Estimators. *Applied Statistics*. 40(2): p. 283-294.
- Moran P. A. P. (1948). Interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*. 10(2): p. 243-251.
- Musakwa, W., Niekerk A. (2014). Monitoring urban sprawl and sustainable urban development using the Moran Index. *Int. Journ. Applied Geospatial Res.* 5(3): p. 1-20.
- Pinon P. (1991). *Lire et composer l'espace public*. Paris, STU.
- Tsai, Y.H. (2004). Quantifying urban form: Compactness versus sprawl. *Urban Studies*, 42(1): p. 141-161.
- Yamada, I. Thill, J.C. (2007). Local indicators of network-constrained clusters in spatial point patterns. *Geographical Analysis*, 39(3), p. 268-292.
- Yamada, I. Thill, J.C. (2010). Local indicators of network-constrained clusters in spatial patterns represented by a link attribute. *Annals Ass'n. Am. Geog.*, 100(2), p. 269-285.



Session 4

Mobilités, flux, trajectoires, modélisation spatio-temporelle

Les réseaux de transport à grande vitesse en France : construction et apports d'un système d'information géo-historique

Françoise Bahoken^{1,2}, Christophe Mimeur³, Martin Koning¹, Thomas Thevenin³.

¹ Institut français des sciences et technologies des transports, des réseaux et de l'aménagement (IFSTTAR-AME / SPLOTT)

² Géographie-Cités / Équipe PARIS (UMR 8504)

³ THEMA-Dijon (UMR 6049 CNRS, Université de Bourgogne)

RÉSUMÉ. *L'originalité de ce Système d'Information Géographique Historique (SIGH) et temporel tient tant dans le type de données mobilisées (il porte sur les réseaux de transport viaires et ferroviaires à grande vitesse) que dans leur profondeur temporelle (de 1975 et 2010). Les possibilités de calcul de matrices de distances kilométriques et/ou temporelles sur réseaux, à l'échelle de la France et à différentes dates, ouvrent d'importantes perspectives d'analyses spatiales. Ils permettent notamment d'apprécier l'effet des investissements réalisés dans l'offre de transport à grande vitesse sur les durées de déplacements interurbains, avec de multiples possibilités d'analyses dans le champ des sciences régionales et des systèmes complexes.*

ABSTRACT. *The originality of this Historical Geographic Information System (SIGH) is that it is focuses on various transport networks (road and railway) at several temporal scale (from 1975 to 2010). The corresponding spatial network offers the possibility to generate spatial and temporal matrices across France, at different dates. These possibilities open important spatial analysis perspectives. They allow for example to assess the effect of investments in high-speed transport supply on the durations of intercity trips, also multiple opportunity in the field of Regional Science and Complex Systems analysis.*

MOTS-CLÉS : *réseau routier, réseau ferroviaire, réseau géo-historique, matrice de distance, matrice de distance-temps.*

KEYWORDS: *road network, railway, geo-historical network, distance matrice, temporal matrice*

Les travaux présentés ont été financés par le projet de recherche BASECOGV (IFSTAR-AME/SPLOTT, Universités Paris-Est Marne la Vallée et de Reims) du PREDIT « Crise, évolution des modes de vie, mobilité et politique de transport » (Convention de recherche n°13-MT-G06-4-CVS-004) ; ils ont été réalisés en collaboration avec l'UMR n°6049 Théma-Dijon.

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse

1. Introduction

Les évolutions des modes de transports utilisés par les Hommes, de leurs choix d'itinéraire et des pratiques aux lieux de destination constituent des sujets d'étude d'autant plus importants que les possibilités d'analyses spatiales ont été récemment décuplées, grâce à la généralisation des Systèmes d'information géographique (SIG) notamment. Si la portée opérationnelle de ces outils n'est plus à démontrer, leurs potentialités spécifiques pour analyser des données de mobilités et leurs interactions avec les dynamiques socio-économiques locales demeurent encore largement sous-exploitées. Qu'ils portent sur la demande ou sur l'offre de transport (les infrastructures et les services), peu de travaux semblent ainsi mobiliser des SIG de réseaux, qui plus est dans une perspective historique, *a contrario* des travaux anglo-saxons (Gregory et Hell, 2007). Pourtant, l'utilisation des SIG Historiques (SIGH) n'est pas dénuée d'intérêts, que ce soit pour les historiens, les géographes ou les économistes.

Cette communication présente l'élaboration d'un SIGH décrivant les réseaux de transport routiers et ferroviaires interurbains en France entre 1975 et 2010. Ce corpus fut élaboré dans le cadre du contrat de recherche BASECOGV financé par le PREDIT (Koning et al., 2015) visant à appréhender certains effets économiques territoriaux liés aux investissements successifs dans les infrastructures ferroviaires à grande vitesse. Si divers auteurs (Dancoisne, P., 1984 ; Dupuy, G et Stransky, V., 1996) ont comparé les performances des réseaux routiers et ferroviaires à une date donnée et/ou suivi l'évolution temporelle d'un réseau particulier, rares sont les études (à notre connaissance) qui contrastent les deux dimensions : bimodale et historique.

Nous décrivons tout d'abord la constitution de l'état du réseau, en 2010, qui servira de base à la reconstitution d'états antérieurs (Section 2). La nécessité de disposer de matrices de distances kilométriques et/ou temporelles sur réseaux nous a conduit à transformer les réseaux en graphes spatiaux, puis à les enrichir (Section 3). Une première analyse portant sur l'évolution des performances comparées de déplacements (route *versus* fer) entre une sélection de 231 unités urbaines (UU) est finalement présentée (Section 4).

2. La reconstruction de réseaux routiers et ferroviaires à grande vitesse

Les distances parcourues et les temps de parcours entre des couples de lieux d'origine et de destination (OD) occupent un rôle central dans les analyses des mobilités spatialisées, notamment au travers de la fonction d'impédance spatiale. En effet, alors que pendant des siècles les déplacements étaient caractérisés par une « *lenteur homogène* » (Ollivro, 2000), les réseaux de transport nés des innovations techniques ont instauré l'essor de la « *rapidité différenciée* » (*ibid.*), avec un bouleversement des rapports de l'Homme avec son espace. Les espaces de proximité se sont élargis et l'horizon des déplacements s'est étendu. C'est pourquoi il est

SAGEO'2016 – Nice, 6-9 décembre 2016

important que la mesure soit effectuée sur les différents réseaux physiques de transport reliant des OD et non seulement à vol d'oiseau. Si la mesure de distance euclidienne est souvent privilégiée dans l'analyse des mobilités, elle n'en demeure pas moins théorique et insuffisamment précise. Disposer de mesures de distances-temps sur réseau suppose de connaître, outre la position (x,y) des OD, celle de leurs liaisons orientées (Lij), autrement dit de disposer d'un graphe spatial (G) symbolisant ces réseaux.

2.1. La constitution de l'état du réseau en 2010

Les données élémentaires mobilisées dans le SIGH portent à la fois sur des nœuds, qui correspondent aux UU, que sur les liens qui les unissent, c'est-à-dire sur infrastructures de transport à grande vitesse. La première étape de l'analyse a donné lieu à la construction de l'état du réseau pour l'année 2010 : le fichier de nœuds correspond aux centres géométriques des UU analysées dans le projet BASECOGV; le fichier de liens étant formé des tronçons viaires et ferroviaires à grande vitesse. La seconde étape fut la reconstitution des réseaux pour 1975, 1982, 1990 et 1999.

2.1.1. Les nœuds du réseau : les unités urbaines

Le choix des UU comme unité spatiale élémentaire se justifie par le fait que nous souhaitons disposer d'une définition plus large et plus dense de l'agglomération urbaine, notamment au regard de celle d'aires urbaines (AU) dont les densités et les superficies sont moindres. En 1999, 18% du territoire est intégré dans une UU contre 41% pour une AU. D'autre part, les UU sont une maille territoriale pertinente pour analyser les interdépendances socio-économiques liées à l'offre de transports interurbains. L'échantillon retenu comporte 231 UU comptant au moins 20000 habitants en 1999. Pour des raisons évidentes de continuité spatiale des réseaux, les UU insulaires (en Corse ou dans les DOM-TOM) ont été écartées.

Le zonage territorial en UU de l'INSEE fut reconstitué à partir d'une agrégation attributaire du fond communal de GEOFLA© de l'IGN, grâce à la table de passage fournie par l'INSEE. Malgré les redécoupages des périmètres des communes et des UU ayant pu intervenir au cours du temps, seule la géographie de 1999 de l'INSEE a été considérée. La distribution spatiale des 231 UU considérées, dans l'ensemble des UU françaises, est présentée sur la Figure 1.

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse

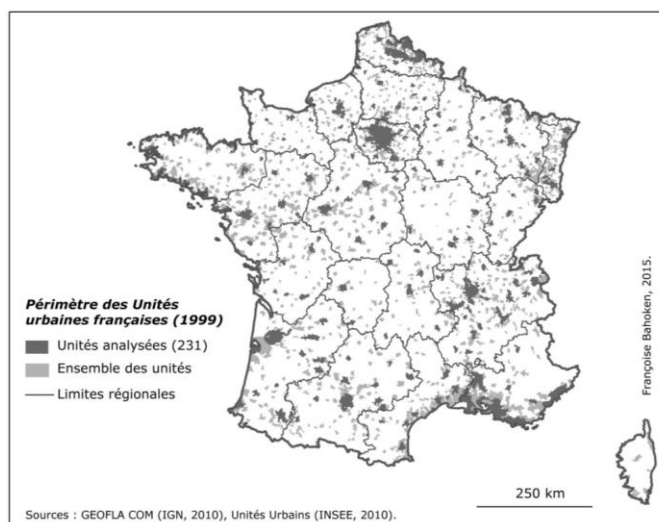


Figure 1. Localisation des unités urbaines analysées

2.1.1. Les liens ou tronçons du réseau en 2010

La construction des linéaires des réseaux routiers et ferroviaires a été réalisée à partir de sources IGN et SNCF Réseau (ex RFF).

Pour les tronçons routiers, nous avons opté pour une définition extensive de la grande vitesse à partir d'une requête par sélection de tronçons de la base GEOROUTE500 de l'IGN. Furent considérées les voies principales à caractère autoroutier ; les bretelles d'accès ; les routes nationales et les liaisons principales, incluant des routes départementales. Une partie du réseau routier principal intra-urbain a en outre été intégrée afin que les centres géométriques des UU puissent être connectés au réseau (phase d'approche). En dehors des zones urbaines, le réseau local d'appoint (initialement constitué pour compléter les linéaires discontinus en raison de la construction des autoroutes) n'a pas été retenu.

Pour le réseau ferroviaire, les lignes à grande vitesse (LGV) correspondent aux tronçons spécialement conçus pour accueillir des trains à grande vitesse (TGV). Il s'agit de lignes électrifiées où la vitesse de circulation commerciale est supérieure ou égale à 250km/h, ainsi que de certaines autres voies classiques ferroviaires électrifiées pouvant accueillir un TGV.

2.2. La reconstitution des réseaux historiques

C'est à partir des fichiers vectoriels de voies routières et ferroviaires de 2010 que les réseaux historiques (1975, 1982, 1990, 1999) ont été reconstitués. Pour le réseau

SAGEO'2016 – Nice, 6-9 décembre 2016

ferroviaire, la base de données FRANcE, développée par Théma, décrivant l'ensemble des tronçons du XIX^{ième} jusqu'en 2000 a été mobilisée. Ce n'est que dans un troisième temps que la connexion des UU aux différents tronçons des réseaux fut réalisée.

2.2.1. La reconstitution des réseaux routiers historiques

La reconstitution des réseaux routiers a été précédée d'une importante opération de géo-référencement portant sur un corpus de cartes anciennes scannées, décrivant le réseau routier entre 1975 et 2010. Quatre séries de cartes de l'IGN ont ainsi été mobilisées, à savoir :

- la série 901, pour la France entière, au 1/10 00 000 ;
- la série Rouge des cartes régionales au 1/250 000, pour les années 1975, 1982 et 1990 ;
- la série Top250 des cartes régionales au 1/250 000, pour l'année 1990 ;
- la série R** des cartes régionales au 1/250 000, pour les années 2006 et 2010.
- La couverture régionale étant incomplète pour les années antérieures à 2006, des fonds complémentaires ont été ponctuellement mobilisés.

Après leur intégration sous QGis, les scans de ces cartes ont fait l'objet d'un géo-référencement en Lambert93, la couche de référence étant le réseau vectoriel de 2010. Pour chacune des années, une procédure d'actualisation de la géométrie du réseau routier de 2010 a ainsi été mise en œuvre, de manière systématique, pour garantir la comparabilité complète des différentes couches, ainsi que la saisie par comparaison visuelle, à partir d'un carroyage régulier, avec des erreurs très ponctuellement supérieures à 2 pixels.

C'est ainsi que la couche de voies routières de 1999 correspond à celles de 2010 diminuées des tronçons qui n'étaient pas visibles sur les cartes régionales de 1990 et ainsi de suite pour l'ensemble des années. Le Tableau 1 présente l'évolution des linéaires de voies routières répertoriées dans le SIGH.

Classe administrative	1975	1982	1990	1999	2010
Autoroutes	4 039	6 153	7 771	9 720	11 377
Route nationales	9 159	9 287	9 403	9 593	10 771
Autres routes	47 568	47 704	48 521	49 297	51 604
Ensemble	60 766	63 145	65 695	68 610	73 752

Source : BASECOGV, 2015.

Tableau 1. Evolution des linéaires routiers (en km) par classe administrative

2.2.2. La reconstitution des réseaux ferroviaires à différentes dates

La reconstitution des réseaux ferroviaires s'appuie sur la base de données FRANcE de Théma. Ce corpus répertorie à l'échelle nationale l'ensemble des tronçons du réseau ferroviaire pour la période 1830-2000. Challenges traditionnels

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse

des SIGH, la diversité et l'hétérogénéité des sources historiques (Gregory et Hell, 2007) ont impacté les choix d'implémentation à l'intérieur de la base de données. Pour autant, la richesse des informations collectées nous donne accès à un large éventail d'attributs.

FRANcE a été mobilisée conjointement à la carte de RFF (SNCF Réseaux), aux cartes régionales de l'IGN et au fichier BD ROUTE500(c) de l'IGN pour l'année 2010. L'état du linéaire a d'abord été identifié visuellement, à partir de la carte de RFF. La création des réseaux antérieurs a ensuite été obtenue – comme pour le réseau routier – par une comparaison visuelle. Les tronçons à usage du fret n'ont pas été retenus. Nous avons aussi supposé que les tronçons de 2000 demeuraient ouverts en 2010, date pour laquelle nous avons intégré, à partir du fichier GEOROUTES500©, la LGV Est inaugurée en 2007. Le tableau 2 présente l'évolution des linéaires de voies ferrées.

Classe administrative	1975	1982	1990	1999	2010
LGV	-	300	705	1274	1826
TGV	-	1601	4739	5044	7906
Corail	5782	5782	5782	5782	4819
TER	17726	15971	12575	12106	10200
RER	10	24	47	62	62
Local	225	182	223	216	216
Autres	2994	1895	2113	2027	1805
Ensemble (km)	26737	25755	26183	26511	26834

Source : BASECOGV (2015).

Tableau 2. Evolution des linéaires ferroviaires (en km) par classe administrative

2.1.3. La connexion des unités urbaines aux réseaux

Les UU concernées ont été connectées à un tronçon de voie routière et/ou ferroviaire, aux différentes dates, à partir de leur barycentre.

Cette connexion a été considérée comme effective – y compris sur les lignes classiques – dès lors qu'elle dispose d'un accès à une autoroute ou à une voie routière express ; à une gare de TGV ou de voie ferroviaire. En raison de « l'effet tunnel » caractéristique des LGV, toutes les gares se trouvant sur le tracé d'une infrastructure n'y sont pas nécessairement connectées, certaines d'entre elles n'étant d'ailleurs que des gares de transit ou de fret. La connexion de l'UU au réseau à grande vitesse a, en outre, été difficile à estimer, dans la mesure où certaines autoroutes sont éloignées des agglomérations. Dans ce cas, la connexion au réseau autoroutier a été considérée comme effective dès que l'UU est située à moins de 10 km à vol d'oiseau d'un échangeur permettant l'accès à une autoroute. Les accès autoroutiers situés à plus de 20 km du réseau principal d'une UU ont donc arbitrairement entraîné l'exclusion de l'UU car considérées comme trop éloignées du réseau principal.

SAGEO'2016 – Nice, 6-9 décembre 2016

Pour déterminer les UU connectées au réseau de TGV, nous nous sommes appuyés sur les gares répertoriées dans GEOROUTE500©, situées dans le périmètre des UU, avec la liste des gares desservies par un TGV en 2015 fournie par la SNCF. Pour chacune des années, ont été supprimées les gares hors réseau ou qui ne sont pas encore mises en service public. La connexion des UU au réseau de LGV fut plus délicate, dans la mesure où certaines gares sont construites à une distance importante de l'agglomération à laquelle elles sont théoriquement rattachées (par exemple, celles de Haute-Picardie et de Lorraine TGV)¹. Pour tenir compte de ces particularités, une UU a été considérée comme desservie par une LGV si 1) une gare nouvelle (dédiée uniquement aux dessertes LGV) se trouve dans l'UU et/ou 2) il existe des liaisons régulières et dédiées, *via* d'autres modes de transport, entre l'UU et une gare nouvelle (comme le cas de la gare de Lorraine TGV desservie par un bus vers/ depuis Nancy et Metz).

	1975	1982	1990	1999	2010
Linéaire de réseau routier (km)	60766	63 145	65 695	68 610	73 752
dont autoroutes	4039	6 153	7 771	9 720	11 377
UU desservies par autoroutes (%)	28.6	51.5	59.7	82.3	92.6
Linéaire de réseau ferroviaire (km)	26737	25 755	26 183	26 511	26 834
dont LGV	0	300	705	1 274	1 826
dont TGV, sur lignes classiques	0	1 601	4 739	5 044	7 906
UU desservies par LGV (%)	0.0	3.3	3.5	5.6	8.7
UU desservies par TGV (%)	0.0	10.4	27.3	31.6	45.9

Source : BASECOGV (2015).

Tableau 3. Evolution des kilométrages des réseaux viaire et ferroviaire et de la connectivité des UU à ces réseaux (1975-2010)

Une vérification générale de la connectivité de l'ensemble des UU à un point du réseau viaire ou ferroviaire, au moins, fut ensuite réalisée. Des corrections manuelles ont été effectuées sur le réseau routier, afin d'assurer la connexité du réseau sans laquelle le calcul des itinéraires complets eut été impossible. Le Tableau 3 compare les linéaires kilométriques disponibles pour chaque réseau et chaque date ainsi que la connectivité des UU à ces réseaux.

Au final, l'architecture de la base de données est présentée sur la Figure 2.

¹ Les choix de localisation des gares de TGV tiennent, en effet, pour beaucoup d'impératifs économiques (Facchinetti-Mannone et al., 2010). Ils sont liés au coût et à la disponibilité du foncier généralement plus accessible dans les espaces ruraux qu'en périphérie urbaine ; aussi environnementales, liées à un impératif de diminution de la pollution sonore des riverains.

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse 6

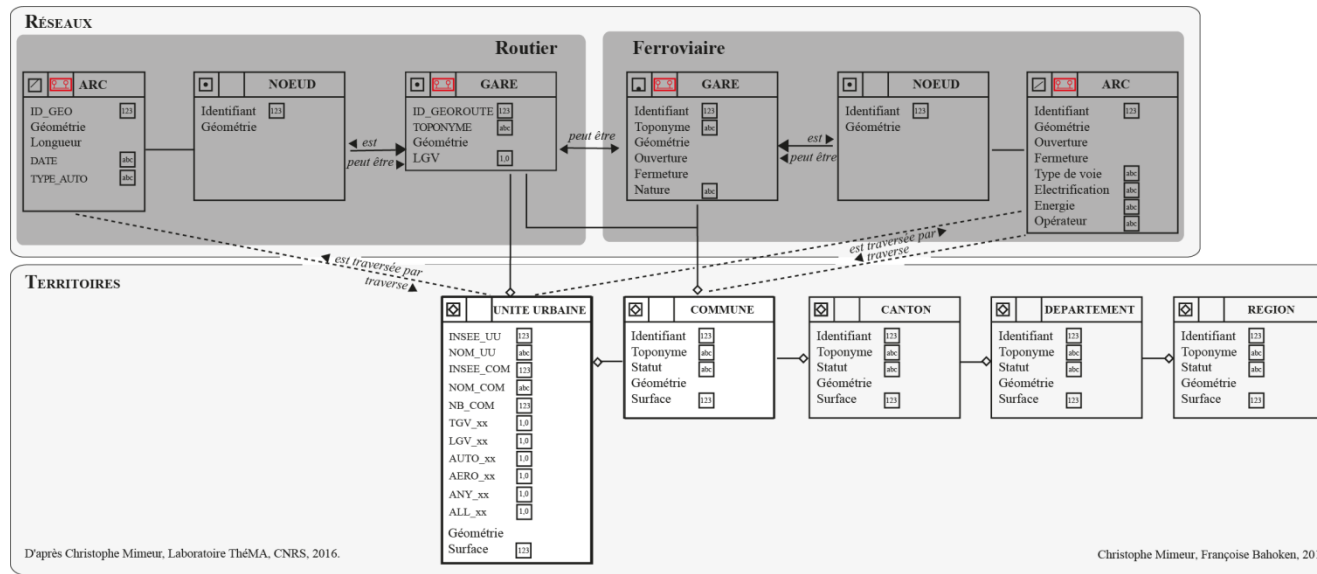


Figure 2. Architecture générale du SIGH

SAGEO'2016 – Nice, 6-9 décembre 2016

Le SIGH recense 27 000 km de voies ferroviaires en 2010, dont 1800 km de LGV et 7 900 km de voies classiques sur lesquelles peuvent circuler des TGV ; et 74 000 km de voies routières, dont 11400 km d'autoroutes. Si le réseau viarie a augmenté de 17% entre 1982 et 2010, celui de voies ferrées apparaît relativement stable, à +4%, en raison notamment de la fermeture des lignes trop vétustes et/ou peu utilisées. La Figure 3 décrit l'évolution de ces réseaux, entre 1975 et 2010.

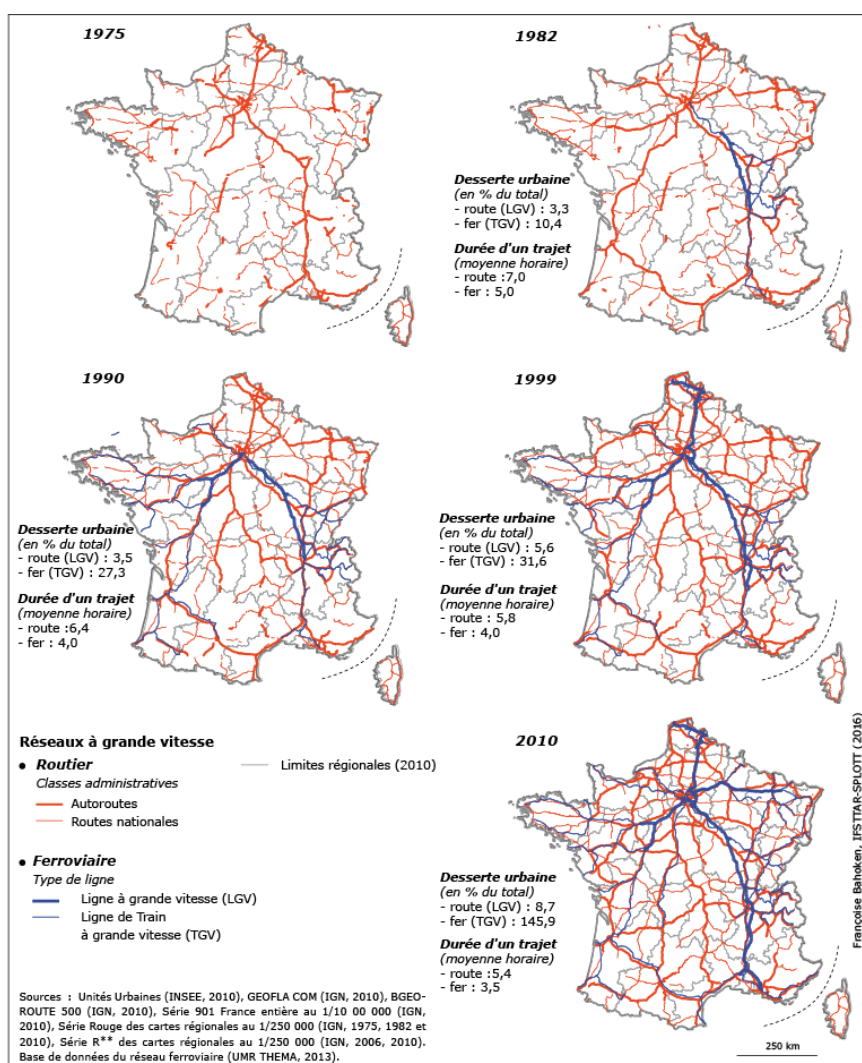


Figure 3. Evolution des réseaux routiers et ferroviaires de 1975 et 2010

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse

On observe une césure initiale de la France qui s'estompe avec le temps. Les UU de Paris de Lyon et de Marseille sont en effet connectées au réseau autoroutier dès 1975 et au réseau TGV à partir de 1981. Ces réseaux sont ensuite étendus vers le nord (avec l'autoroute A1 vers Lille et la LGV Nord) et encore plus à l'est (avec l'autoroute A4 Metz-Nancy et LGV Est-européenne) tandis que, dans le même temps, les lignes rapides de l'ouest du pays demeurent à l'état de projet (les autoroutes A10-A81 vers Rennes, A36 de Beaune à Mulhouse ainsi que la LGV Atlantique et le TGV Bordeaux à Perpignan).

3. La construction de matrices de distances-temps sur réseaux

Dans le cadre de BASECOGV, l'analyse économétrique visait à mettre en relation les niveaux d'emplois au sein des UU avec des indicateurs d'accessibilité aux emplois et aux aménités. C'est pourquoi il a été nécessaire de calculer des distances-temps de parcours sur les réseaux, aux différentes dates. Pour ce faire, il a été nécessaire de transformer les réseaux en graphes ; la deuxième étape étant celle du calcul d'un plus court chemin sur ces graphes.

3.1. Des réseaux de transport aux graphes spatiaux

Le calcul de distances parcourues sur un réseau est une opération souvent réalisée pour déterminer un itinéraire optimal, généralement le plus court, en empruntant un système de transport donné. L'approche s'inscrivant dans le cadre d'analyse de la théorie des graphes, le réseau doit être transformé en un graphe spatialisé.

3.1.1. La construction de graphes spatiaux

Définition du graphe spatialisé

Le graphe spatialisé simple $(G)=(N, L)$, est un objet mathématique formé d'un ensemble N de couples de nœuds (n) , d'origine (i) et de destination (j) . Ces N nœuds sont mis en relation par des liens (L) , au sein d'un graphe qui n'est pas nécessairement connexe.

Pour tout $(i) \neq (j)$

$$L \subset \{(i, j) \mid ni \in N\}$$

Le graphe issu d'un réseau de transport est souvent orienté (i, j) et valué, ce qui signifie que des attributs supplémentaires (w) sont associés à la description des arcs.

Ainsi, $(G)=(N, L, w)$ ou (w) est un attribut du couple de lieux (i, j) , une mesure de longueur du lien, de vitesse ou de coût de transport, par exemple. Cette mesure caractérise une fonction de poids associée au lien (L) qui va pondérer la mesure totale de distance parcourue.

$$w : L \rightarrow R$$

Le passage du réseau au graphe est une opération réalisée automatiquement, fondée sur la topologie et la sémantique des réseaux. L'approche majoritairement

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse

distance) et la valeur d'espacement, la première nécessitant que quatre conditions³ soient remplies.

Parmi les différentes méthodes de calcul des itinéraires, notre choix s'est porté sur l'algorithme de Dijkstra-Moore (1959), généralement utilisé dans le calcul des itinéraires routiers, pour un graphe connexe valué ou non. Ce choix se justifie par sa robustesse et son efficacité pour calculer des itinéraires sur des réseaux spatiaux, quelle que soit la mesure de distance considérée.

3.1.2. Les outils disponibles, le choix de H2Network pour OrbiGis

Divers outils permettent le calcul de parcours dans un graphe et la génération des matrices OD, mais tous ne sont pas adaptés à nos besoins : le calcul de distances sur des données massives, non sélectionnées *a priori*, la restitution de fichiers cartographiques et des matrices OD correspondantes.

Nombre de SIG intègrent aujourd'hui des modules de cheminements dans des graphes et le calcul plus ou moins aisé de distances sur réseaux⁴. Ces extensions permettent la construction (et/ou la vérification pour un fichier réseau importé) d'une structure topologique de graphe à partir d'une couche réseau (un *shapefile*), la vérification de sa géométrie, la suppression des lacunes, la vérification de la connexité, de manière à ce que le cheminement puisse être calculé sur tout ou partie du graphe. D'autre part, elles permettent le calcul de plus courts chemins kilométriques ou temporels, en fonction de critères complémentaires de la table des attributs (la longueur des tronçons par exemple).

Notre choix s'est porté sur *H2Network X* (Bocher, E, Petit, G., 2014), solution comparable sur le plan technique aux possibilités de QGIS. L'outil procure un environnement de développement riche sous OrbisGIS, grâce à ses bibliothèques de fonctionnalités d'analyse spatiale et de géovisualisation. Les patrons d'applications (*templates* de scripts SQL) réalisés à partir de H2Gis (Gouge et al., 2014) autorisent aussi l'analyse et la gestion de graphes spatiaux, et leur combinaison avec les potentialités classiques d'un SIG. Sur le plan pratique, l'intérêt réside dans l'accès direct aux scripts formant le *distancier* IFSTTAR – IRSTV–Ecole Centrale de Nantes (Bocher et al., 2015), développé dans le cadre d'un groupe de travail du projet Belgrand – GEBD, ainsi qu'au support des développeurs pour affiner le calcul des matrices. Sur le plan méthodologique, *H2Network X* autorise l'articulation d'informations portant à la fois sur le réseau (des caractéristiques des liens et des nœuds) que sur leur contexte, c'est-à-dire l'environnement du nœud ou du tronçon.

³ Pour un ensemble N de lieux, D² est une fonction réelle de distance (ou métrique) si et seulement si elle satisfait aux quatre conditions suivantes : 1) non négativité (D_{ij} ≥ 0) ; 2) identité (si D_{ij} = 0, alors i = j) ; 3) symétrie (D_{ij} = D_{ji}) ; 4) inégalité triangulaire : D_{ik} ≤ (D_{ij}) + (D_{jk}).

⁴ Les extensions les plus connues sont Chronomap© pour Mapinfo©, Network Analysis© pour Arcgis©, Odomatrix© (Hilal, 2010) ou encore TransCad©. Toutes étant payantes nous avons fait le choix de mobiliser des ressources libres et ouvertes. Trois extensions pour SIG libres sont généralement utilisées : *Pgrouting* (PostGreSQL / PostGis), *Graphe routier* (Qgis) et *H2Network* (OrbiGis).

SAGEO'2016 – Nice, 6-9 décembre 2016

Il est donc possible de prendre en compte des informations qui portent à la fois sur l'environnement traversé par le réseau (le type de territoire concerné, selon qu'il est situé en zone urbaine, péri urbaine, ...) que sur les attributs des tronçons (des variables de pondération des vitesses liées au type de territoires) ou des nœuds (prise en compte de l'inter modalité).

3.1.3. *La construction des matrices*

Etant donné l'existence du *distancier*, la construction des matrices consiste à faire tourner l'algorithme correspondant, sur un réseau routier ou ferroviaire donné, à partir d'une couche de linéaire et d'une couche surfacique (de laquelle vont être extraits les points (x,y) d'origine ou de destination). La première étape consiste à connecter des points d'OD, typiquement les barycentres des mailles pour générer le graphe à partir duquel le plus court chemin va être calculé. La procédure générale est décrite dans E. Bocher et al. (2015). A chaque itinéraire est *in fine* associé un nombre total de kilomètres parcourus sur le réseau ainsi que la durée du voyage. Les matrices correspondantes résultent de l'agrégation, par une fonction somme, des valeurs de distances (ou de durée) des différents tronçons formant l'OD. Le critère du plus court chemin entre OD portant soit sur les distances kilométriques, soit sur des temps de transport, les graphes ont fait l'objet d'un enrichissement sémantique.

3.2. *L'enrichissement sémantique des graphes*

L'enrichissement sémantique concerne l'ajout de caractéristiques supplémentaire à notre réseau, des variables qualitatives et quantitatives qui complètent la nomenclature des fichiers IGN.

3.2.1. *L'intégration de vitesses moyennes de circulation sur les tronçons*

Les tronçons routiers ont été enrichis par le calcul d'une mesure de vitesse de déplacement. La vitesse variant nécessairement selon la classe de la route (autoroute, route nationale, ...) et le moment de la journée (heures de pointe, heures creuses) nous avons utilisé deux familles de critères : les paramètres définis par M. Hillal (2010)⁵ dans Odomatrix© et les attributs GEOROUTE500©.

Nous avons donc estimé la vitesse en termes d'heures creuses ou pleines, à partir des attributs 'Vocation' et 'type de voie' des fichiers IGN. Les vitesses ainsi calculées ont ensuite été pondérées en considérant 8 heures de pointe (6-10h et 16-20h) et 16 heures creuses. Si la plage des heures pleines peut apparaître large, nous

⁵ Les vitesses de circulation estimées par M. Hilal utilisent des paramètres liés au type de voie et à l'environnement de l'infrastructure routière. Cette pondération se justifie par le fait que, pour un tronçon donné, la vitesse moyenne de circulation est en pratique plus faible que la vitesse légale autorisée dans l'environnement concerné. L'auteur mobilise les trois critères suivants : l'environnement, interprété comme la position de la commune dans une AU (au sens de l'INSEE), par exemple, dans la commune centre ; la tranche de population du pôle urbain et le type de voie routière.

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse

avons préféré sous-estimer les vitesses routières plutôt que de les surestimer, la vitesse moyenne sur un axe routier était rarement égale au maximum légal autorisé. Le tableau 4 présente l'évolution des vitesses moyennes sur les deux types de réseaux, en fonction des classes administratives.

Ces vitesses moyennes sont pondérées par la longueur des tronçons, car les bretelles d'accès aux autoroutes, souvent situées en milieu urbain nécessitent d'être distinguées des tronçons autoroutiers de rase campagne.

Classe administrative	1975	1982	1990	1999	2010
Réseau routier					
Autoroutes	103	111	114	116	117
Routes nationales	83	73	83	82	83
Autres routes	64	56	64	64	65
Ensemble	70	64	73	74	75
Réseau ferroviaire					
LGV	-	178	221	207	202
TGV	-	113	127	134	174
Autres voies ferrées	88	96	103	111	106
Ensemble	88	98	111	121	132

Source : BASECOGV (2015).

Tableau 4. *Vitesses (en km/h) pondérées par la longueur des tronçons*

3.2.2. La pénalité temporelle des « 200 kilomètres/30 minutes » pour le fer

Les premiers calculs des temps de parcours réalisés par voie ferroviaire se sont avérés très éloignés de la réalité, du fait des modalités spécifiques des voyages réalisés en train. L'application directe du principe de mesure du « plus court chemin » sur le graphe ne s'est pas révélée pertinente, les voyages en trains étant fréquemment soumis à des ruptures de charges, à l'inverse des déplacements par la route.

A défaut d'avoir su modéliser finement, et autrement, les temps imputables aux ruptures de charge ou d'attente en gare, nous avons défini une pénalité de retard aux longs trajets (donc hors TGV), selon une règle des « 200 kilomètres/30 minutes » (Koning et al., 2015). Ainsi, dès qu'un déplacement en train présente une portée supérieure à 200 kilomètres, nous lui avons ajouté 30 minutes de pénalité par tranche de 200 kilomètres. L'exception des TGV permet de limiter la pénalisation des dessertes à très longue distance, tout en garantissant une bonne approximation des durées globales des voyages interurbains.

SAGEO'2016 – Nice, 6-9 décembre 2016

3.2.3. *Les mesures de distances temporelles disponibles*

L'analyse des déplacements interurbains entre les 231 paires d'UU est obtenue à partir de matrices OD carrées et fermées (231*231), dont la construction s'inscrit dans deux grandes familles de distances : kilométriques ou temporelles. Trois types de mesures temporelles sont ici proposées : la durée du parcours 1) sur le réseau routier, 2) la durée de parcours sur le réseau ferroviaire, avec les pénalités liées aux éventuelles interconnexions et 3) la durée minimale de parcours pour une OD donnée entre les réseaux routiers et ferroviaires. Les durées globales de déplacements sont présentées dans le Tableau 5.

Temps de parcours	1975	1982	1990	1999	2010
Par la route (en heures)	7.9 (1.5)	7,0 (1,2)	6,4 (1,1)	5,8 (0,9)	5,4 (0,8)
Temps train (en heures)	8.3 (1.5)	6,3 (1,2)	5,7 (1,1)	5,3 (1,0)	4,8 (0,9)
Temps minimum (en heures)	7.5 (1.4)	6,2 (1,1)	5,6 (1,1)	5,1 (1,0)	4,7 (0,9)
Avantage route (en %)	28.2 (44.9)	27,5 (19,9)	23,4 (17,4)	26,4 (17,5)	21,6 (16,1)

Les écarts-types sont entre parenthèses.

Source : BASECOGV (2015).

Tableau 5. *Evolution des durées de parcours interurbains (en heures) entre 1975 et 2010*

Quel que soit l'indicateur considéré, on observe les temps de parcours interurbains ont considérablement baissé entre 1975 et 2010 (de -42% à -31%). Ce résultat s'explique par les investissements publics réalisés ces dernières décennies pour le développement des infrastructures de transport à grande vitesse, aussi par la manifestation d'externalités de réseaux, autrement dit par l'augmentation mécanique des bénéfices imputables à la présence d'un réseau à grande vitesse, en lien justement avec l'augmentation de sa densité.

Pour l'année 2010, on observe un avantage systématique des parcours réalisés par voie mode ferroviaire (voir Tableau 5) : la durée moyenne d'un déplacement en train est de 4,8 heures, tandis que le temps de déplacement en voiture est de 5,4 heures. Malgré la pénalité des « 200 kilomètres/30 minutes », qui a entraîné une augmentation de 37% en moyenne de la durée de voyages par le train, on n'observe donc pas de préférence massive pour les trajets en automobile, la route n'étant privilégiée que pour 22 à 28% des liaisons. La comparaison de ce résultat, avec les pratiques réelles des voyageurs réalisées à longues distances sur le territoire français valide alors l'hypothèse générale selon laquelle la composante temporelle (la durée du voyage) du coût généralisé des déplacements (i.e. du coût réel de déplacement sur lequel les individus sont supposés baser leurs choix modaux), n'explique qu'une partie des choix de mobilité individuels. Autrement dit, notre SIGH devrait dans le

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse

futur inclure des éléments sur les dépenses monétaires liées aux mobilités individuelles.

4. Premières analyses des performances comparées (route/fer) des parcours interurbains

Pour présenter l'intérêt de ce SIGH, nous proposons une courte analyse descriptive des performances des différentes possibilités de déplacement interurbaines. Des analyses complémentaires sont disponibles dans Notot (2015), et dans Bahoken et al. (2016). Il est tout aussi intéressant d'analyser l'évolution de la connectivité des 231 UU analysées aux différents réseaux de transport, en lien avec l'évolution des temps de déplacements qui en découle (voir Tableau 6).

Mode de desserte des UU Temps de déplacement	Autoroutes		LGV		TGV	
	Oui	Non	Oui	Non	Oui	Non
Par la route (heures)	5,8 (1,0)	6,9 (1,2)	5,2 (0,6)	6,2 (1,2)	6,0 (1,0)	6,2 (1,2)
Temps train (heures)	5,3 (1,1)	6,1 (1,3)	4,3 (0,5)	5,5 (1,2)	5,3 (1,0)	5,6 (1,3)
Temps minimum (heures)	5,1 (1,0)	6,0 (1,2)	4,2 (0,5)	5,4 (1,1)	5,2 (1,0)	5,5 (1,2)

Les écarts-types sont entre parenthèses. Les trajets considèrent les pénalités temporelles pour les interconnexions en gare, pour le mode ferroviaire.

Source : BASECOGV, 2015.

Tableau 6. Différences de temps de parcours selon la connectivité aux réseaux

Si 52% des territoires urbains étaient reliés au réseau autoroutier en 1982, on observe que plus de 90% le sont désormais en 2010. Cela souligne, une fois de plus, l'incroyable (densification du) maillage du réseau routier français. Pour ce qui est des LGV ferroviaires, elles sont finalement peu nombreuses puisqu'elles ne concernent que 9% de l'échantillon en 2010. En dépit des nombreux investissements dont elles ont pu faire l'objet, peu d'espaces géographiques apparaissent en être dotés et, de fait, bénéficieraient de leurs supposés bienfaits sur le développement socio-économique local. Ce constat mérite toutefois d'être nuancé, par la prise en compte de l'ensemble des UU accueillant des TGV, que ces derniers circulent sur des lignes dédiées ou sur des voies ferrées classiques (ce qui est le cas pour une partie non négligeable de leurs déplacements) : on observe d'ailleurs que la part des UU desservies par un TGV est passée de 10% en 1982 à 46% en 2010.

Les temps de parcours moyens apparaissent assez différenciés selon que les UU ont accès à un réseau à grande vitesse ou non. Les UU desservies par des autoroutes présentent des durées moyennes de voyages de 5,8 heures, contre 6,9 heures pour les autres. Pour les UU connectées au réseau de LGV, les temps de parcours sont à peine inférieurs : de 5,2 heures (et 6,2 heures pour les non connectées). En fait, seule

SAGEO'2016 – Nice, 6-9 décembre 2016

la connexion (ou non) au réseau TGV semble jouer un vrai rôle dans la réduction des temps de déplacements.

Cependant, en dépit d'une connectivité aux réseaux de transports bien plus forte pour les très grandes UU (60% d'entre elles sont desservies par TGV sur la période 1982-2010 : 93% par autoroutes et 11% par LGV), les différences des temps de parcours moyens ne sont pas très importantes par rapport aux autres UU. Certes, les UU de 200 000 habitants et plus en 2010 mettent systématiquement moins de temps pour relier les autres territoires, mais les écarts ne sont pas si prononcés que cela. En effet, les temps de parcours estimés sur ces réseaux contemporains sont corrélés à 75% avec les estimations réalisés sur des réseaux historiques (XIX^{ème} siècle).

5. Conclusion

L'innovation apportée par ce SIGH porte sur l'extension spatiale et temporelle de ces données et sur la bi-modalité : la base de données peut en effet faire l'objet d'analyses réalisées sur le temps long, en combinant transport ferroviaire et routier pour cette partie contemporaine. L'analyse menée sur les paires d'UU n'est pas limitative, il est tout à fait possible de refaire les calculs similaires sur un autre jeu de données, à la seule condition que les nouvelles localités considérées soient physiquement connectées au réseau analysé. Des travaux non publiés ont d'ailleurs été réalisés en ce sens sur des réseaux de poste et des réseaux de voies romaines. BASECOGV décrit non pas la demande de transports interurbains, mais uniquement l'offre potentielle de transport inter urbains à grande vitesse. Il pourrait être intéressant de complexifier l'offre par la fréquence de desserte et d'envisager la demande à partir de nouvelles sources, historiques et contemporaines. De même que la sémantique de ces réseaux pourrait être enrichie par des données concernant la demande de transport, idéalement, les effectifs de voyageurs (ou de fret) qui pourraient être reconstitués pour chacun des itinéraires.

Cela étant, des analyses spatio-temporelles sont actuellement envisagées sur les graphes spatiaux. Deux directions sont actuellement privilégiées : d'une part, la caractérisation de la structure de ces réseaux à l'aide d'indicateurs dédiés issus de la théorie des graphes ; d'autre part, une analyse spatio-temporelle visant à mettre en perspective l'évolution des changements structurels du réseau à grande vitesse, à partir de l'appréhension du réseau ferroviaire historique constitué au XIX^{ème} siècle. Cela permettrait en outre de tester certaines variables instrumentales sur un temps long.

Remerciements

Les auteurs souhaitent remercier Vivien Notot, stagiaire en L3Pro Carthagéo, ENSG-Université Paris 1 (2013-2014) ; Carlos Olarte-Bacarès, Chargé de Recherche IFSTTAR sous contrat (2014-2015) et Olivier Bonin (Chargé de recherche IFSTTAR), pour l'aide à la construction de la base de données.

BASECOGV, une base de données géo historique sur les réseaux à grande vitesse

6. Bibliographie

Bahoken, F., Koning, M., Mimeur, C., Olarte-Bacares, C., Thevenin, T. (2016), Les temps de parcours interurbains en France : une analyse géo-historique, *Transports*, n°495, pp. 17 -25.

Bocher E., Petit, G. (2014), *H2Network : un outil pour la modélisation et l'analyse de graphes dans le système d'Information Géographique OrbisGIS*, rapport du Lot 92 – GEBD Grand Equipement Bases de données J11-06, 92 p.

Dancoisne, P., 1984. *Théorie des graphes et constitution du réseau ferré français*. Paris I - Sorbonne, Paris.

Dupuy, G., Stransky, V., 1996. Cities and highway networks in Europe. *Journal of Transport Geography*, 4, pp. 107–121.

Dijkstra, E. W. (1959), A note on two problems in connexion with graphs, in *Numerische Mathematik*, n°1, pp. 269-271.

Facchinetti-Mannone, V., Bavoux, J.-J. (2010), L'implantation des gares TGV en France : tensions interscalaires, jeux d'acteurs et recompositions spatiales, *Belgeo. Revue belge de géographie* 9–22. doi:10.4000/belgeo.6638

Gregory, I.N., Ell, P.S. (2007) Historical GIS: technologies, methodologies, and scholarship. *Cambridge University Press*.

Kasraian, D., Maat, K., Stead, D., van Wee, B. (2016) Long-term impacts of transport infrastructure networks on land-use change: an international review of empirical studies. *Transport Reviews*, pp.1-21.

Ollivro, J. (2000) L'homme à toutes vitesses: de la lenteur homogène à la rapidité différenciée. *Presses universitaires de Rennes*, Rennes.

Gouge, A., Bocher, E., Fortin, N., Petit, G. (2014), H2Network: A tool for understanding the influence of urban mobility plans (UMP) on spatial accessibility, in: *Proceedings of the 3rd Open Geospatial Research & Education symposium*.

Hilal, M. (2010), ODOMATRIX. Calcul de distances routières intercommunales, in: *Cahier des Techniques de l'INRA*, Numéro spécial « Méthodes et outils de traitement des données en sciences sociales. Retours d'expériences », pp. 41-63.

Huriot, J.-M., Perreur, J., *Distances, espaces et représentations : une revue*, Document de travail, Institut de Mathématiques Economiques, 35 p.

SAGEO'2016 – Nice, 6-9 décembre 2016

Thevenin, T., Schwartz, R., Sapat, L (2013), Mapping the Distortions in Time and Space: The French Railway Network 1830 – 1930, Historical Methods, *in: Journal of Quantitative and Interdisciplinary History*, n°46, pp. 134-143.

Koning M., Bahoken F., Bazin Benoit S., Beckerich C., Blanquart C., Delaplace M., Joignaux G., Savy M. (2015) *BASECOGV – Dessertes TGV et dynamiques économiques locales : un éclairage à partir de la distinction entre territoires productifs, résidentiels ou intermédiaires*, Rapport de recherche pour le PREDIT, 197 p.

Notot, V. (2014), *Cartographie des retombées économiques indirectes de l'implantation des lignes à grande vitesse, intégration et analyse de données d'études dans un SIG*, Rapport de stage de fin d'études, Licence 3 Professionnelle Géomatique et Environnement, ENSG, 48 p.

Formalisation spatio-temporelle de perturbations maritimes pour la simulation du réseau maritime

Martin Tanguy, Aldo Napoli

CRC- Centre de recherche sur les Risques et les Crises, MINES ParisTech, PSL University.

CS 10207 rue Claude Daunesse 06904 Sophia Antipolis Cedex, France
martin.tanguy@mines-paristech.fr, aldo.napoli@mines-paris.fr

RESUME. L'objectif de cet article est de proposer une méthodologie pour formaliser les risques maritimes via leurs caractéristiques spatio-temporelles. En effet, l'espace maritime connaît des risques, qu'ils soient de nature anthropique (marrée noires, collisions entre navires, piraterie maritime...) ou naturels (tempêtes, icebergs, récifs...). Ces risques ont un influence sur le réseau de transport maritime : ils peuvent générer des accidents mais également influencer sur l'usage de l'espace maritime pour le transport.

ABSTRACT. The goal of this paper is to propose a methodology to formalize maritime risks by the use of their spatial and temporal characteristics. Indeed, the maritime space known risky events, like anthropogenic (black tide, collisions with ships, maritime piracy ...) or natural risks (storms, icebergs, reefs ...). These risks have an impact on the maritime transportation: they can generate accidents but also influence the use of maritime space for transportation.

MOTS-CLES : Risques maritimes, modélisation spatio-temporelle, simulation multi-agent.

KEYWORDS: Maritime risks, spatiotemporal modelling, agent-based simulation.

SAGEO'2016 – Nice, 6-9 décembre 2016

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

L'essor des transports à une échelle globale s'inscrit dans un phénomène de mondialisation. 90% des biens échangés à travers le monde sont effectués par voie maritime. La maritimisation se définit comme un processus de dépendance économique des états au trafic maritime due à ce phénomène de mondialisation (Vigarié, 1983). Ce phénomène transforme les territoires, et principalement les espaces côtiers mais a également comme conséquence une territorialisation des espaces maritimes, à savoir une utilisation et gestion des espaces pour les activités humaines. L'utilisation de cet espace a augmenté depuis 1970 et principalement le transport, que ce soit via porte-conteneurs ou tanker. Cela forme ainsi un réseau de transport maritime lié à l'emplacement des ports dans le monde. Cependant ce réseau s'étend sur un espace vaste et dynamique, l'espace maritime, qui peut générer des risques pour l'usage de l'espace à des fins de transports. L'objectif de cet recherche est de proposer une démarche méthodologique pour formaliser un environnement de simulation et évaluer de la vulnérabilité du transport maritime par la simulation des risques comme objet spatio-temporel. En effet, les perturbations ayant des natures hétérogènes, cet article propose une formalisation des risques maritimes comme objet spatio-temporel afin de les simuler, dans un second temps, dans un environnement de simulation du réseau maritime. La première partie présente les différents risques existant en mer et la notion de perturbation pour le transport maritime, la seconde partie présente la notion d'objet spatio-temporel et sa formalisation et enfin la troisième partie s'intéresse à la formalisation des perturbations comme objet spatio-temporel et leurs usages dans le cadre d'un système de simulation du réseau maritime.

1.1. *Le réseau maritime face au risque*

L'espace maritime est un espace en constamment en mouvement. Cette spécificité est décrite par la notion d'hypermobilité (Parrain, 2012). Cela signifie que les composants du territoire sont dynamiques, rendant ces espaces particulièrement incertains pour le transport maritime, ce qui induit un risque que ce soit en pleine mer ou dans les espaces littoraux. Elliott et al. (2014) proposent une typologie des risques en mer qu'ils soient de nature anthropologique ou naturels. Ces risques sont répartis en 14 catégories en fonction des causes des aléas : sont distingués les risques naturels (facteurs climatiques, hydrologiques, tectoniques...) et anthropiques (biologiques, technologiques...). Ces risques ont des caractéristiques spatio-temporelles propres. Certains risques sont statiques, comme les récifs, mais certains sont plus complexes et dynamiques. Cela regroupe des risques comme les tempêtes dont la dynamique est liée aux conditions atmosphériques ou la piraterie maritime liée à des problèmes de nature géopolitique. L'IMO (International Maritime Organisation) classifie en deux catégories les risques en mer du point de vue de l'activité humaine en mer : il s'agit de la sécurité (les risques liés à l'activité même, cela concerne les risques technologiques, les collisions entre navire...) et la sûreté liée à des actions volontaires et délibérées de nuire (piraterie, terrorisme en mer).

Formalisation spatio-temporelle de perturbations maritimes

Ces risques sont d'autant plus difficiles à prendre en compte du fait que la dynamique évolue en fonction du trafic maritime sur le réseau. Ces risques ont un impact en provoquant des perturbations sur le réseau maritime et ont pour conséquence un coût supplémentaire pour le transport, en raison du temps supplémentaire et de la distance nécessaire pour atteindre le port de destination. Les perturbations sont définies par une perte de performance du médiateur spatial pour relier les différentes parties de la structure spatiale (Zavitsas, 2012). Cela signifie que ces perturbations sont des événements contraignant l'usage de l'espace pour le transport. La vulnérabilité peut être définie comme un impact, comme le degré de dégradation de la perturbation, ou comme une capacité à réagir et à réorganiser la structure spatiale après la perturbation (Veyret et Reghezza, 2005). Plusieurs approches existent pour étudier la vulnérabilité, toutes liées à la notion d'accessibilité, à savoir la capacité d'un lieu à pouvoir être rejoint (Rodrigue, 2013). En effet, l'espace maritime est un médiateur spatial et son efficacité est liée aux opportunités de déplacement offertes (Gleyze, 2005).

1.2. Les aléas comme perturbation du réseau

Le rôle d'un réseau de transport est de relier efficacement des lieux entre eux afin d'optimiser son utilisation. Dans le cas du transport maritime, Berle et al. (2011) définissent pour principal enjeu du transport maritime, la recherche de maximisation du coût de transport face aux différentes trajectoires possibles dans l'espace maritime. Ainsi la notion de vulnérabilité d'un réseau de transport se focalise sur les impacts possibles liés à des pertes d'efficacité des réseaux, construits par les possibilités offertes par l'espace. Berdica et Mattsson (2007) s'intéressent à la vulnérabilité en considérant, pour un réseau de transport donné, la susceptibilité que des incidents entraînent une réduction considérable du niveau de service du réseau. Ce niveau de service correspond à l'efficacité du réseau à remplir son rôle de médiateur spatial (Gleyze, 2005). Pour mesurer ces pertes d'efficacité, Taylor et Susilawati(2012) montrent que l'ensemble des mesures existantes repose sur le concept d'accessibilité à l'espace(mesures locales ou globales) comme la facilité d'atteindre dans un réseau la destination désirée (Niemeier, 1997). Ainsi les perturbations peuvent dans un réseau donné, soit rendre inaccessible la destination désirée soit rendre plus difficile (plus long) le chemin entre deux localisations dans le réseau. La question principale est ainsi de savoir comment adapter des mesures propres aux graphes à la spécificité spatiale du réseau maritime : en effet, si les nœuds représentent les ports et plateformes du réseau, les liens ne peuvent être pris en considération, à savoir que deux nœuds peuvent être liés par une infinité de liens à travers la surface que représente l'espace maritime. En effet, dans le cadre classique d'une analyse de réseau par théorie des graphes, une perturbation est modélisée par ablation d'un composant voire d'une partie du réseau (Murray et Grubestic, 2007). Cependant du fait des possibilités offertes par l'espace maritime pour les navires, la prise en compte des perturbations demande une modélisation spécifique de celle-ci. Ainsi, les perturbations doivent également se voir attribuer des propriétés spatiales adaptées : une géométrie, permettant de définir l'extension de la perturbation dans l'espace maritime, mais également une temporalité,

SAGEO'2016 – Nice, 6-9 décembre 2016

indiquant l'évolution de cette spatialité dans le temps. Ainsi, les perturbations peuvent se définir comme des objets spatio-temporels.

1.3. Objectif de recherche

Cette recherche ayant pour objet d'évaluer la vulnérabilité du réseau maritime face à un risque, l'objectif est de simuler des perturbations spatio-temporelles afin de tester l'influence de celles-ci sur le réseau maritime. Ces perturbations, modélisées sous forme d'objets spatio-temporels ont une forme, une extension spatiale, une temporalité ainsi qu'une influence sur l'accessibilité à l'espace pour un navire donné. Pour cela, le réseau étant ici considéré comme l'ensemble des déplacements effectués dans l'espace maritime, il est ainsi nécessaire de décrire des comportements individuels pour les navires liés aux niveaux d'accessibilité offert par l'espace maritime. Deux étapes sont nécessaires pour la construction des comportements : premièrement, le choix d'une destination du voyage à partir d'un point de départ dans le réseau. Deuxièmement, le comportement spatial du navire au travers du médiateur spatial. Ces comportements sont liés aux niveaux d'accessibilité de l'espace maritime et permettent, à posteriori d'évaluer la vulnérabilité du réseau maritime au travers des caractéristiques spatio-temporelles des différents comportements spatiaux des navires, évoluant en fonction des scénarios de perturbations dans l'espace maritime. Cela peut entraîner des configurations inattendues du réseau maritime en fonction de la perturbation.

2. Objet et simulation spatio-temporelle

Pour comprendre l'influence de ces perturbations, il convient de se questionner sur l'aspect spatio-temporel propre aux caractéristiques de l'espace maritime, en effet, celles-ci ont des natures diverses : elles peuvent se déplacer (icebergs) se diffuser (marées noires) et possèdent dans chaque cas de figure des caractéristiques propres. Ainsi afin de formaliser les perturbations, il est nécessaire de définir les objets spatio-temporels, les dynamique associées et la prise en compte des objets dans une simulation spatialement explicite.

2.1. Objets spatio-temporels

Pour simuler les effets des perturbations sur le réseau maritime, il est nécessaire d'avoir recours à des principes de modélisation spatio-temporelle. Deux approches en modélisation spatio-temporelles existent : la conceptualisation de l'espace en tant qu'objet (perceptive discrète) et en tant que champs (perceptive continue). Le point de vue dit discret, ou centré-objet, est que chaque objet dans l'espace possède ses propres délimitations, ses frontières, dans l'espace et le temps. Ils sont distincts entre eux et de l'ensemble de l'espace, qui est alors un support d'objets spatio-temporels. (Mathian and Sanders, 2014). C'est dans cette perspective que cette recherche s'inscrit. Peuquet (1994) définit un objet géographique par 3 axes interdépendants : le Quoi (Nature de l'objet représenté), le Où (Situation et dimension de l'objet dans l'espace) et le Quand (Processus et événements de l'objet dans le temps). Le premier

Formalisation spatio-temporelle de perturbations maritimes

axe, le Quoi, cherche à déterminer la nature de ce qui est modélisé. Deux types d'objets pouvant être distingués : les objets « *bona fide* » correspondant à des objets dits naturels ayant une existence physique et matérielle et les objets « *fiat* » construit pour la modélisation, correspondant à des limites arbitraires, socialement construits par et pour l'homme, ils correspondent aux limites politiques, administratives, juridiques ou sociales. Ces objets possèdent des données attributaires associées à l'objet spatio-temporel. Le second axe, le Où, cherche à définir l'emplacement et la forme des objets spatiaux. En soi, cet axe n'est pas une spécificité de la modélisation spatio-temporelle mais peut-être vu comme une base : c'est la modification des géométries simples spatiales des objets (points, lignes, réseaux, polygones ou volumes) qui va permettre de modéliser le changement. En effet, la localisation des objets peut se modifier dans le temps. Il faut ainsi distinguer les objets fixes, qui n'admettent pas de modification spatiale au cours du temps et les objets mobiles, qui peuvent se déplacer dans l'espace. Le dernier axe, Quand, est certainement le plus important dans le cadre de la modélisation spatio-temporelle. En effet, c'est par le croisement entre les questions Où et Quand que les notions d'évènements (une singularité dans l'espace et le temps) et de processus, à savoir ce qui entraîne la modification des objets géographiques (Livet et al., 2010). Ces modifications sont à l'origine de la dynamique des objets.

2.2. Dynamique des objets spatio-temporels

Les trois axes (Quoi, Où, Quand) sont repris par Claramunt et al., (1999) en y ajoutant les axes Comment et Pourquoi. Sans ces axes, le modèle reste description, à savoir qu'il observe et décrit des changements aux travers les objets. L'axe Comment permet de dépasser ce stable pour expérimenter des processus en testant des hypothèses pour comprendre les processus du changement à partir de simulations. Enfin, l'axe Pourquoi s'intéresse aux causes et permet d'émettre des théories et tester des lois générales. Ce dernier type est principalement utilisé pour les modèles en géographie physique. Ainsi c'est principalement l'axe Comment qui se révèle être le plus pertinent dans le cadre de cette étude, à savoir faire apparaître des processus et les tester via la dynamique des objets spatio-temporels. Cette dynamique des objets spatio-temporels est liée à la prise en compte du changement par et pour ces objets dans l'espace. En effet, ces objets apparaissent, disparaissent et se modifient au cours du temps. Les objets peuvent être caractérisés en fonction de leurs rapports au temps. Grenon et Smith (2004) distinguent deux types d'entités. Les entités SNAP : celle-ci sont durables, persistantes dans le temps et ne sont pas modifiées en tant qu'entité. Ce sont les propriétés de celle-ci qui peuvent évoluer, mais celle-ci garde leur cohérence (ex : Une ville, un individu). Les secondes, les entités SPAN (perdurantes) prennent en considération le temps par des événements et des processus, à savoir que l'entité admet un début, une fin et une suite d'évènements. Ici, l'intérêt est placé d'avantage sur la trajectoire de l'objet dans l'espace. Pour Livet(2010) la notion d'évènement se focalise sur le changement d'état de l'entité spatiale et la notion de processus aux causes entraînant les modifications de l'entité, que ce soit les caractéristiques spatiale, attributaires ou relationnelle de l'entité géographique. Ces deux types d'entités admettent cependant

SAGEO'2016 – Nice, 6-9 décembre 2016

une évolutivité. Ces évolutions dans le temps décrivent des processus spatio-temporels (Claramunt and Thériault, 1995). Ceux-ci décrivent premièrement l'évolution des entités seules : les changements d'état (apparition ou disparition de l'entité) ou leur maintien dans un temps donné (stabilité). Ces processus amènent également des transformations de ces entités : par expansion (augmentation de la surface de l'objet), contraction (diminution de la surface de l'objet), déformation (changement de forme de l'objet mais conservation de l'aire de la surface), déplacement (modification de la place de l'objet dans l'espace) et rotation par changement d'angle de l'objet dans l'espace. Ces objets peuvent être considérés et modélisés différemment au sein d'un modèle spatial. Ruas et Sanders (2015) distinguent quatre types d'espace supports pour les objets géographiques en fonction des objectifs même de la modélisation : l'espace peut être un support à une configuration d'objets et se modifier en fonction de l'apparition, modification et disparition d'objets, il peut également être un support fixe dans lequel s'inscrit des trajectoires d'objets mobiles, un support dans lequel se sont les interactions d'objets dans le temps qui sont décrits (flux entre les parties de l'espace) ou être un espace maillés où les subdivisions de l'espace évoluent dans le temps. Dans une approche orienté-objet, c'est cependant la seconde, qui est le plus souvent utilisée.

2.3. Simulation d'objets spatio-temporels

Deux types d'approches sont possibles afin d'explorer les dynamiques des objets dans l'espace : une approche statistique, basée sur la mise en évidence de similarité entre les différents objets spatio-temporels et l'approche complexe, focalisé sur les interactions existantes entre les agents entre eux et l'environnement partagé. Ces deux approches ont des objectifs différents : dans le premier, l'approche statistique, l'objectif du modèle est de mettre au point un modèle reproduisant les faits observés dans la nature, cela demande alors de comparer empiriquement le modèle à des données existantes. Dans le second, l'intérêt est principalement porté sur la mise en place de règles qui vont permettre au modèle de reproduire au mieux des mécanismes existants pour arriver à une présentation simplifiée du réel. Il est possible de distinguer deux grands types de système de simulation spatialement explicites Marceau and Benenson (2012) : les automates cellulaires et les systèmes multi-agents. Les automates cellulaires sont des maillages de cellules représentant l'espace dans lequel chaque cellule possède un état. En fonction de règles de transition, basée sur le voisinage topologique des cellules, celle-ci voient leurs états se modifier. Ces systèmes sont principalement utilisés pour étudier les phénomènes de diffusion mais également les flux de véhicules ou de piétons. La seconde approche est l'utilisation de systèmes multi-agents. Cette approche agent a pour objectif non pas l'identification des variables déterminantes d'un problème mais la compréhension des mécanismes à l'origine de l'émergence du phénomène observé. Le principe de la modélisation à base d'agents est de modéliser des systèmes complexes en adoptant une approche bottom-up par une modélisation centrée sur les individus. Un agent est une entité qui a ses propres objectifs et des capacités dans un environnement partagé (Ferber, 1997). Les systèmes multi-agents sont spatialement explicites, à savoir que les agents évoluent dans un environnement spatialisé. Les

Formalisation spatio-temporelle de perturbations maritimes

agents ont un emplacement dans l'environnement, et les systèmes à base d'agents sont adaptés à la modélisation spatiale et pour la simulation. Selon Pumain (2003), les systèmes multi-agents sont 3 atouts majeurs de la géographie: démontrer le rôle des acteurs individuels dans les changements globaux et observer les transformations des territoires et l'adaptation des territoires du fait des interactions socio-économiques existantes. Les systèmes multi-agents sont utilisés dans le cadre des transports : Davidsson et al. (2005) dresse un état de l'art de plus de 56 approches multi-agents pour simuler les réseaux de transports. Cette étude met en évidence le manque de système multi-agent appliqué au transport maritime, en dehors de systèmes s'intéressant aux échanges entre terminaux (ports), ne prenant pas en compte l'espace entre les points de départ et d'arrivée. Cependant plusieurs études existent depuis pour formaliser la navigation et l'espace maritime au travers de la simulation. Fournier (2005) formalise des comportements agents en fonction de l'environnement maritime pour les navires. Vaněk et al. (2013) formalise un système multi-agents pour étudier la piraterie au travers trois groupes d'agents (tanker, pirates, navires de sécurité). D'autres études s'intéressent au transport de conteneurs (Sinha-Ray et al, 2011) ou à l'énergie (Zavitsas, 2011) mais ne prennent pas en considération l'accessibilité à l'espace maritime nécessaire à l'évaluation de la vulnérabilité.

3. Proposition d'une démarche de modélisation pour l'évaluation des risques maritimes

L'objectif de cet article est d'élaborer une démarche de modélisation des perturbations en mer pour l'évaluation des risques par simulation. Ce système est basé sur les interactions existantes entre d'une part les perturbations, modélisés en tant qu'objets spatio-temporels s'inscrivant dans une structure spatiale, et d'autre part les navires, modélisés sous forme d'agent, spatialement explicite, se déplaçant vers leur but en fonction des possibilités offertes par l'environnement maritime. La démarche de modélisation se base sur des scénarios représentant l'évolutivité d'une perturbation spatio-temporelle au sein de l'espace maritime. Ces scénarios peuvent être modélisés sous forme d'objets spatio-temporels ayant une spatialité, une temporalité ainsi qu'une influence sur l'accessibilité à l'espace pour les navires. Pour définir ces scénarios il convient en premier lieu d'explicitier et formaliser les perturbations comme un objet spatio-temporel. L'objectif d'un tel système est de permettre de simuler les perturbations ainsi que les navires et analyser leurs interactions au travers d'un environnement de visualisation permettant l'évaluation de la vulnérabilité du réseau maritime.

3.1. Définition des perturbations comme trajectoire spatio-temporelle

Une perturbation peut être formalisée comme objet spatio-temporel à partir de deux caractéristiques spatiales : un centroïde et une extension spatiale, ainsi qu'une caractéristique temporelle, sa durée et enfin une caractéristique d'intensité via son

SAGEO'2016 – Nice, 6-9 décembre 2016

impact sur l'accessibilité à l'espace pour un navire donné. Ces 4 axes sont représentés sur la figure suivante :

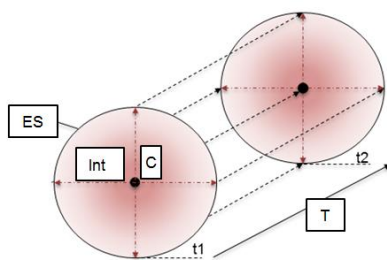


FIGURE 1. Caractéristiques d'une perturbation comme objet spatio-temporel

Centroïde(C) : le centroïde est le centre de la perturbation, ayant un niveau d'impact sur l'accessibilité. Il peut rendre compte d'un centroïde unique (point – ex : un iceberg), mais peut également être de forme linéaire voire polygonale dans la mesure où la forme de la perturbation n'admet pas de centroïde unique (ex : un champ de glace en Arctique). **Temporalité(T)** : la temporalité se définit comme la durée de la perturbation. Dans le cadre d'un pas de temps non contenu, c'est l'ensemble des pas de temps nécessaires à la manifestation de la perturbation. La temporalité peut être persistante, périodique ou unique. Si elle est persistante cela signifie que la perturbation n'a pas de fin à partir de son commencement. Elle est périodique si elle admet un début, une fin mais qu'elle peut malgré tout survenir à nouveau. Enfin, elle peut être unique, commencer et finir sans que son existence implique qu'elle se déclenche de nouveau. **Extension spatiale(ES)** : l'extension spatiale de l'objet correspond à la zone totale d'influence à partir de la localisation du centroïde de la perturbation. L'extension spatiale peut prendre plusieurs formes en fonction de la dynamique spatiale de la perturbation. Certaines perturbations n'ont cependant pas d'extension et peuvent être uniquement formalisées spatialement à partir du centroïde. **Intensité(Int)** : L'intensité définit la répartition spatiale de l'influence de la perturbation sur l'accessibilité à l'espace des navires. Celui-ci peut n'être qu'une singularité, à savoir, être uniquement compris dans le centroïde. L'impact peut également être homogène sur l'ensemble de l'extension spatiale. Enfin, il peut également être inégalement réparti : soit être décroissant ou décroissant à partir du centroïde, ou n'avoir pas de logique linéaire.

3.2. Représentation des perturbations dans une structure spatiale

Les perturbations peuvent ensuite avoir des évolutions distinctes en fonction de ces caractéristiques spatio-temporelles. A partir de la typologie de Claramunt et al (1995) quatre catégories de dynamiques spatiales peuvent être distinguées : premièrement, les perturbations statiques qui ne connaissent pas de dynamique et n'évoluent pas dans le temps. Elles peuvent toutefois être persistantes, ou bien admettre un début ou une fin. Deuxièmement, les perturbations en déplacement : elles admettent un déplacement du centroïde pendant la durée de vie de la perturbation. Troisièmement, les perturbations en expansion/rétraction correspondant aux perturbations qui ont un centroïde qui n'admet pas de translation dans l'espace mais leur extension spatiale évolue au cours du temps. Et enfin, les perturbations en déplacement et diffusion : ces perturbations se diffusent et se déplacent également dans l'espace. A partir de ces critères une perturbation peut être définie par une fonction générique permettant de définir une perturbation à partir d'une trajectoire spatio-temporelle, composé de plusieurs étapes à temps discret (Pos). Ces trajectoires peuvent être formalisées sous la forme : $\mathbf{P}(\mathbf{Pos}, \mathbf{T}, \mathbf{C}, \mathbf{ES}, \mathbf{Int})$, P étant la perturbation, Pos le numéro de l'étape,

TABLE 1. Exemple d'une trajectoire de perturbation spatio-temporelle

Pos	T	C	ES	Int
1	10	A11	4	-8
2	10	E9	3	-6
3	10	G4	2	-4
4	20	K2	1	-2

Celle-ci décrit alors un déplacement, ainsi qu'une rétraction de la perturbation. Elle peut être visualisée au travers d'une structure spatiale, dans le cas présent une grille, utilisée comme espace support à la perturbation (Figure 2) :

SAGEO'2016 – Nice, 6-9 décembre 2016

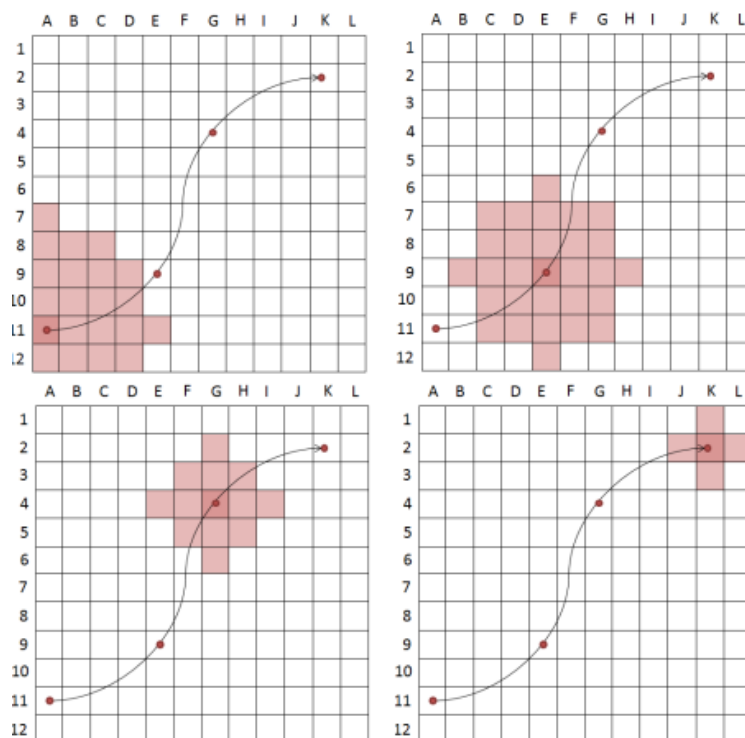


FIGURE 2. Représentation spatiale de la perturbation pour chaque pas de temps

3.3 Influence des perturbations sur les navires

Ces perturbations réduisant l'accessibilité à l'espace, il est nécessaire de formaliser les comportements spatiaux des navires en fonction de l'espace et du niveau d'accessibilité afin d'évaluer l'influence sur la forme spatio-temporel du réseau maritime. Pour cela il est nécessaire de déterminer des logiques de déplacement pour les navires dans l'espace afin de construire les trajectoires, et donc la structure géométrique des liens du réseau. Un comportement spatial peut être défini par un objectif (un lieu à atteindre) et un chemin pour rejoindre ce lieu, en fonction du niveau d'accessibilité de l'espace. Plusieurs mesures de distance peuvent être utilisées pour définir le comportement spatial des navires. Premièrement, la distance topologique. Celle-ci correspond au nombre de cellules qu'un navire doit traverser pour atteindre l'objectif de destination. Cependant, cette distance ne prend pas en compte le niveau d'accessibilité de l'espace. Pour cela d'autres mesures de distances permettent de pondérer le comportement spatial des navires en fonction de l'accessibilité : la distance absolue pondérée et la distance absolue relative. La première correspond au chemin ayant la meilleure accessibilité globale, en tenant compte de la somme de tous les endroits traversés dans la structure spatiale. Pour un navire, cela signifie que le choix du chemin le plus accessible est entre sa position

Formalisation spatio-temporelle de perturbations maritimes

réelle dans la matrice et la destination à atteindre. La seconde construit la trajectoire en fonction de l'accessibilité moyenne du, ce type de chemin peut rallonger fortement la distance topologique. L'ensemble des distances pouvant être utilisés par les navires est représentés dans l'exemple suivant (Figure. 3) :

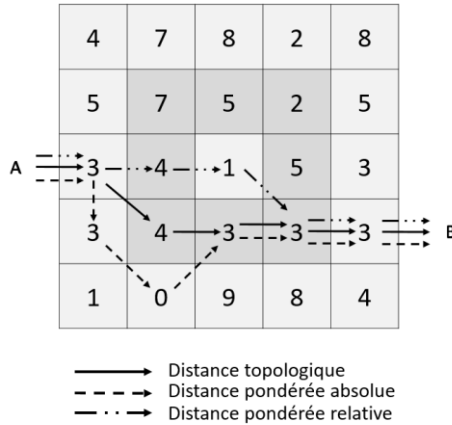


FIGURE 3. Différents comportements spatiaux en fonction du niveau d'accessibilité au travers d'un espace maillé

La sélection de la distance pour le comportement spatial dépend de la capacité des navires et leur connaissance de l'environnement maritime (la connaissance des valeurs d'accessibilité). La première distance, topologique, mesure le chemin le plus court en fonction de la distance euclidienne dans la matrice (ce qui ne fonctionne pas en cas de distance de Von Numann ou de Moore). La seconde, la distance pondérée absolue, est utilisé pour la prise en compte de l'effet de accessibilité dans l'espace sur le comportement spatial et donc de mesurer le plus court chemin temporel. La dernière, la distance relative pondérée peut être utilisée dans des cas imprévisibles comme les perturbations.

3.3 Formalisation du système de simulation des risques en mer

Ces différents composants forment un modèle pour la simulation du réseau maritime en fonction des perturbations spatio-temporelles. Les perturbations peuvent être déclenchées par l'utilisateur afin de tester des scénarios et évaluer leur impact sur le réseau maritime. Pour formaliser le modèle, un diagramme des différentes classes objets du modèle peut être fait. Pour cela nous avons utilisé un diagramme de classe UML afin de préciser la nature spatio-temporel des éléments du modèle :

SAGEO'2016 – Nice, 6-9 décembre 2016

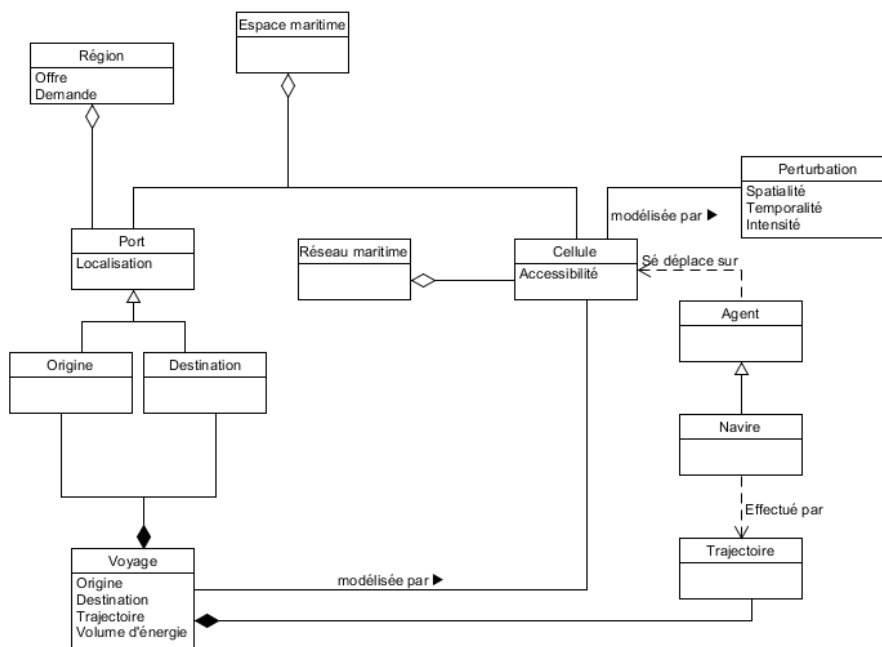


FIGURE 4. Diagramme de classe de l'ensemble du modèle

L'ensemble de ces éléments forment ainsi un modèle pour la simulation de la vulnérabilité du réseau maritime face à un risque de perturbation. En fonction des scénarios de perturbations, il est possible de mesurer l'impact de celle-ci à partir des modifications que celle-ci entraînent sur la formation du réseau maritime et sa réorganisation en fonction des perturbations. Les différents composants du modèle sont définis de la manière suivante

TABLE 2. Composants du modèle de simulation

Composant du modèle	Description
Espace maritime	Il correspond au médiateur spatial, à savoir l'ensemble de l'espace pouvant être utilisé à des fins de transport.
Région	Les régions correspondent à l'ensemble des Etats du monde ayant un niveau d'offre et de demande en biens et service nécessitant un échange par voie maritime.
Port	Localisés à l'interface entre l'espace maritime et l'espace terrestre, les ports sont les portes d'entrée et de sortie du médiateur spatial. Ils délimitent ainsi le comportement spatial des navires dans l'espace maritime.

Formalisation spatio-temporelle de perturbations maritimes

Voyage	Un voyage est la réalisation d'un échange à l'intérieur du médiateur spatial. C'est un objet spatio-temporel borné par une origine, une trajectoire déterminé par l'accessibilité et une destination.
Réseau	Le réseau maritime représente l'ensemble des voyages effectués dans l'espace maritime. Il est défini typologiquement par les origines et destinations de ces voyages et géométriquement par les trajectoires spatio-temporelles des voyages.
Cellule	Subdivision de l'espace correspondant à une zone d'égale accessibilité dans l'espace.
Perturbation	Scénarios des aléas testés par l'utilisateur afin de mesurer l'impact d'une perturbation sur le réseau maritime.

4. Conclusion

Cet article décrit l'élaboration d'une méthodologie pour l'évaluation de la vulnérabilité du réseau de transport maritime. Celle-ci repose sur la notion d'accessibilité, centrale dans la relation existante entre l'espace et la vulnérabilité. Cette vulnérabilité est mesurée via la formalisation de scénarios de perturbations spatio-temporelles. Ces scénarios permettent de simuler ainsi le comportement des navires et donc l'émergence de structures du réseau liées aux risques maritimes. Pour formaliser ce réseau, un modèle conceptuel de données spatiales est utilisé pour décrire les composantes et le fonctionnement du système. Ce modèle permet de simuler l'effet des pertes d'accessibilité dues aux risques sur les comportements spatiaux des navires. Cette modélisation est construite en plusieurs étapes : la modélisation de l'espace comme support aux interactions des navires et perturbation, la modélisation du réseau à partir de l'emplacement des ports et des trajectoires des navires et la modélisation des comportements des navires et la modélisation spatio-temporelle des perturbations. Ce modèle sera ensuite simulé sur un outil de simulation multi-agent tel que NetLogo. A partir des simulations réalisées, des indices et des indicateurs pourront être mis en place pour évaluer la vulnérabilité du réseau maritime et être testé sur des espaces réels pour analyser les comportements des navires dans les zones maritimes les plus sensibles.

Bibliographie

- Berdica, K., Mattsson, L.-G., 2007. Vulnerability: A Model-Based Case Study of the Road Network in Stockholm, in: Murray, A.T., Grubestic, T.H. (Eds.), *Critical Infrastructure*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 81–106.
- Berle, Ø., Asbjørnslett, B.E., Rice, J.B., 2011. Formal Vulnerability Assessment of a maritime transportation system. *Reliab. Eng. Syst. Saf.* 96, 696–705.
doi:10.1016/j.ress.2010.12.011
- Claramunt, C., Parent, C., Spaccapietra, S., Thériault, M., 1999. Database Modelling for Environmental and Land Use Changes, in: Stillwell, J., Geertman, S., Openshaw, S.

SAGEO'2016 – Nice, 6-9 décembre 2016

- (Eds.), *Geographical Information and Planning*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 181–202.
- Claramunt, C., Thériault, M., 1995. Managing Time in GIS An Event-Oriented Approach, in: Clifford, J., Tuzhilin, A. (Eds.), *Recent Advances in Temporal Databases*. Springer London, London, pp. 23–42.
- Davidsson, P., Henesey, L., Ramstedt, L., Törnquist, J., Wernstedt, F., 2005. An analysis of agent-based approaches to transport logistics. *Transp. Res. Part C Emerg. Technol.* 13, 255–271. doi:10.1016/j.trc.2005.07.002
- Elliott, M., Cutts, N.D., Trono, A., 2014. A typology of marine and estuarine hazards and risks as vectors of change: A review for vulnerable coasts and their management. *Ocean Coast. Manag.* 93, 88–99. doi:10.1016/j.ocecoaman.2014.03.014
- Fournier, F., 2005. *Intégration de la dimension spatiale au sein d'un modèle multi-agents à base de rôles pour la simulation ; Application à la navigation maritime*. PhD thesis, IRENav.
- Gleyze, J.-F., 2005. *La vulnérabilité structurelle des réseaux de transport dans un contexte de risques*.
- Grenon, P., Smith, B., 2004. SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spat. Cogn. Comput.* 4, 69–104. doi:10.1207/s15427633scc0401_5
- Livet, P., Muller, J.-P., Phan, D., Sanders, L., 2010. Ontology, a Mediator for Agent-Based Modeling in Social Science. *J. Artif. Soc. Soc. Simul.* 13. doi:10.18564/jasss.1538
- Mathian, H., Sanders, L., 2014. *Objets géographiques et processus de changement: approches spatio-temporelles*.
- Marceau, D., Benenson, I. (Eds.), 2012. *Advanced Geo-simulation Models*. BENTHAM SCIENCE PUBLISHERS.
- Murray, A.T., Grubestic, T.H., 2007. Overview of reliability and vulnerability in critical infrastructure, in: *Critical Infrastructure*. Springer, pp. 1–8.
- Niemeier, D.A., 1997. Accessibility: an evaluation using consumer welfare. *Transportation* 24, 377–396.
- Parrain, C., 2012. *La haute mer : un espace aux frontières de la recherche géographique*. *EchoGéo*. doi:10.4000/echogeo.12929
- Peuquet, D.J., 1994. It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Ann. Assoc. Am. Geogr.* 84, 441–461. doi:10.1111/j.1467-8306.1994.tb01869.x
- Ruas, A., Sanders, L., 2015. Regards croisés sur la modélisation des dynamiques spatiales. *Rev. Int. Géomat.* 25, 275–300. doi:10.3166/riig.25.275-300
- Taylor, M.A.P., Susilawati, 2012. Remoteness and accessibility in the vulnerability analysis of regional road networks. *Transp. Res. Part Policy Pract.* 46, 761–771. doi:10.1016/j.tra.2012.02.008
- Vaněk, O., Jakob, M., Hrstka, O., Pěchouček, M., 2013. Agent-based model of maritime traffic in piracy-affected waters. *Transp. Res. Part C Emerg. Technol.* 36, 157–176. doi:10.1016/j.trc.2013.08.009
- Veyret, Y., Reghezza, M., 2005. Aléas et risques dans l'analyse géographique. *Ann. Mines* 61–89.
- Zavitsas, K., 2012. *The vulnerability of the petroleum supply chain*. Imperial College London.

Interaction Homme Machine pour l'orientation spatiale des malvoyants

Said Boularouk^{1,2}, Didier Josselin^{2,1}, Eitan Altman^{3,1}

1. Laboratoire d'Informatique d'Avignon, UAPV

said.boularouk@univ-avignon.fr

2. UMR ESPACE 7300, UAPV, Avignon

didier.josselin@univ-avignon.fr

3. INRIA, Sophia Antipolis

eitan.altman@sophia.inria.fr

RÉSUMÉ. Pour aider les malvoyants à compenser la perte visuelle et à appréhender l'environnement urbain extérieur, nous proposons un outil interactif basé sur l'automatisation de la langue via une interaction homme-machine. Nous présentons un simulateur, développé en CSharp, pour aider les personnes malvoyantes à connaître les lieux urbains extraits de la base de données géographiques d'OpenStreetMap. Le principe est de transmettre les informations spatiales par un module de synthèse vocale et recevoir les requêtes de le malvoyant par un module de reconnaissance vocale.

ABSTRACT. To help blind people to compensate visual perception and to better understand their outdoor environment, we developed a method using human-computer interaction via Text-to-Speech. It helps visually impaired people to know surrounding places from OpenStreetMap data by hearing. The principle is to convey spatial information by voice synthesis and receive requests from blinded people by voice recognition.

MOTS-CLÉS : Synthèse vocale, reconnaissance vocale, aveugle, malvoyant, cartographie, sig, sémiologie sonore, bases de données, appréhension de l'espace, interaction homme machine, géolocalisation, GPS, Haversine

KEYWORDS: speech synthesis, speech recognition, blind, visually impaired, mapping, gis, semiotics sound, databases, apprehension of space, human computer interaction, geolocation, GPS, Haversine

SAGEO'2016 - Nice, 6-9 décembre 2016

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

Selon l'Organisation Mondiale de la Santé, il existe près de 285 millions de personnes qui présentent une déficience visuelle, dont 45 millions sont malvoyants et 180 millions souffrent de handicap visuel. 90 % des personnes qui présentent une déficience visuelle vivent dans un pays à faible revenu, 82 % des malvoyants sont âgés de 50 ans et plus¹. Les études démontrent que l'immobilité des malvoyants est une conséquence directe de leur handicap visuel. Pour remédier à ce problème, l'appréhension de l'espace par un champ perceptif autre que la vision s'avère nécessaire (Josselin *et al.*, 2016).

Les malvoyants ont besoin d'aide humaine ou matériel pour se déplacer. D'un côté, pour éviter les obstacles, de l'autre pour s'orienter dans l'espace. L'identification des lieux avoisinants est primordiale pour tout déplacement sécurisé et confortable. Cette identification est difficile, voire impossible pour le malvoyant seul dans un environnement méconnu. L'appréhension de l'espace passe par la connaissance des composantes de l'environnement. Nous ne considérons dans cet article que les déplacements extérieurs et excluons du champ les déplacements *indoor*.

En milieu urbain extérieur, les éléments de l'environnement qui peuvent intéresser les malvoyants (lieux à atteindre) sont représentés par des points (magasins, lieux de culture et de loisir, centres de santé, services de sécurité, administrations, etc.). Chaque composante est identifiée par son nom, sa localisation GPS (et son adresse). Sur certaines cartes en ligne, les composantes contiennent les détails du service qu'elle offre et d'autres caractéristiques techniques et urbaines. Au vu des progrès de la cartographie et de l'apparition du GPS, la collecte des positions et les détails des composantes géographiques est une chose acquise, à une précision variable selon les conditions de collectes de l'information (nombre de satellites, zones enclavées, qualité du matériel, mouvement). Dans cette article, nous proposons une solution adaptée aux malvoyants, qui illustre comment l'automatisation de la parole peut aider à construire une représentation de l'espace relativement comparable à la représentation «cartographique».

Au Japon, l'entreprise Human Tech of Japan a développé un système de navigation pour les personnes malvoyants, en utilisant le GPS et le smartphone sous *android* (Kushwaha, Kushwaha, 2011). Le système utilise la synthèse vocale du téléphone (TTS ou Text To Speech) pour aider le malvoyant dans ses déplacements (Tenbrink *et al.*, 2013). En Corée du Sud, une équipe de Hanyang Cyber University, a développé un autre système simple pour smartphone sous *android* permettant la navigation vocale (Cha *et al.*, 2013). D'autres systèmes de locomotion pour malvoyants existent déjà, comme la canne électronique d'aide à la détection d'obstacles (Cbel, Visioprtonic, etc.). Citons également,

1. <http://www.who.int/fr/>

Interaction Homme Machine pour l'orientation spatiale des malvoyants

par exemple, le système d'aide à la navigation Okeneea² dédié aux malvoyants, développé en France.

L'objectif de cette recherche est de proposer une application dédiée aux malvoyants pour appréhender l'espace, basée sur la géolocalisation, la recherche de lieux dans un environnement proche et le dialogue homme machine. Nous présentons un simulateur géographique permettant à un utilisateur malvoyant de se positionner dans un milieu urbain et chercher les lieux qui l'entourent dans un rayon choisi, modifiable par simple commande vocale. L'utilisateur peut également sauvegarder ses lieux préférés dans une liste de favoris, dans la perspective de tracer un itinéraire dont les repères sont importants et connus. Le simulateur peut transmettre la distance, l'adresse et les détails de chaque lieu. Cet outil, à l'état de prototype, peut potentiellement être utilisé par le malvoyant pour préparer ses déplacements et, à terme, pendant son déplacement lui-même pour la recherche de trajets adaptés. Dans son état actuel, le prototype ne permet pas encore de calculer des itinéraires optimaux adaptés aux malvoyants.

2. Le malvoyant dans l'espace

A cause du handicap visuel et des difficultés de déplacement dus à la cécité, les connaissances spatiales d'une personne malvoyante restent limitées, mais elles sont de même nature que celles des voyants (Hatwell, 2006). L'approche écologique de Gibson (Baltenneck, 2010; Virole, 2009; Hatwell, 2006) propose de situer la cognition au croisement de l'humain et de l'environnement dans lequel il évolue, en fonction de ce qu'il lui permet (ou non) de faire. Cette notion est très pertinente dans le cadre du déplacement d'une personne malvoyant, car elle nous apprend que la lecture de l'environnement est en partie individuelle. La théorie de Gibson a abouti à deux concepts principaux : le champ optique ambiant et les *affordances*. Le principe d'affordance désigne à la fois toutes les possibilités d'action sur un objet et la capacité d'un objet à suggérer sa propre utilisation. William Gaver considère l'affordance comme l'ensemble des propriétés de l'environnement offertes à un individu pour accomplir certaines actions (Gaver, 1992). Quand une personne malvoyante ne trouve pas dans le stimulus visuel l'information nécessaire pour se déplacer, parce que son champ de vision est déficient ou quasiment inexistant, elle se concentre sur les affordances reçues à travers les autres champs de perception.

Le concept d'affordance montre en effet «*la façon dont les organismes se couplent avec leurs environnements, (...)* et décrit une relation structurelle entre la perception et l'offre de sens de l'objet» (Virole, 2009). L'extension du concept des affordances au langage nous permet de comprendre que la per-

2. <http://www.okeneea.com/actualites/decouvrez-kapteo-une-revolution-dans-le-quotidien-des-personnes-deficientes-visuelles/>

SAGEO'2016 – Nice, 6-9 décembre 2016

ception est fondamentalement orientée vers la recherche de signification. Comprendre un mot est avant tout comprendre le sens qu'il renvoie. De même, un bruit entendu est signifiant dès lors qu'il correspond à une adaptation de l'organisme à son environnement, "*langage d'orientation naturaliste*" (Virole, 2009). D'ailleurs, il n'est pas nécessaire de faire appel à des analyses approfondies du fait linguistique pour observer que chez l'enfant, les mots sont appris parce qu'ils sont d'abord associés à des situations concrètes. Le support de la perception n'a pas vraiment d'importance si le «sens» de l'objet ciblé ou de l'action voulue est compris par le récepteur.

Certaines aides techniques (GPS, balises sonores...) peuvent compenser la déficience visuelle pour l'interprétation de l'espace. Elles sont utilisées différemment selon le type du handicap visuel de chaque individu (CFPSAA, 2010). L'automatisation de la parole peut jouer le rôle de transmetteur de l'information spatiale extraite d'une carte géographique. Grâce à la perception auditive, le malvoyant peut connaître l'emplacement et la distance des lieux et leur orientation relative.

Une enquête réalisée auprès des malvoyants par l'Institute of Gerontology et AIST au Japon (Miura *et al.*, 2011) a montré que les meilleurs moyens de réception des informations spatiales et de l'environnement durant leur mobilité sont les suivants : 48,8 % des participants malvoyants préfèrent la parole, 34,1 % choisissent le *pitch*³ et l'*harmonie sonore*⁴ et 31,7 % préfèrent les vibrations. L'étude ne montre pas de différence particulière entre les choix des participants malvoyants et les participants complètement aveugles. Cela signifie, très logiquement, que la perception auditive est le mode privilégié et commun chez les malvoyants pour compenser la perte visuelle pour l'orientation et l'appréhension de l'espace.

De Saussure (De Saussure, 1997) définit le signe comme signifié (sens) et signifiant (aspect physique du sens). La réalité d'une chose et sa signification (signe) est le résultat d'une convention entre les individus au sein d'une communauté. Cette convention est utile dans le domaine de la linguistique informatique et dans toute représentation graphique (exemple : légende d'une carte). La *représentation graphique* est une forme de codification d'information en signes graphiques, le *langage verbal* est aussi une forme de codification en signes sonores. La codification est tout simplement une transcription d'une information connue par l'intermédiaire d'un système de signes quelconques. Alors, pourquoi ne pas codifier l'information spatiale en signes sonores, plus spécifiquement dans notre cas, en texte parlé de synthèse ?

3. Ton ou hauteur d'une note, en lien avec sa fréquence ; plus le pitch augmente plus la note est aiguë.

4. En musique, l'harmonie sonore peut être considérée à la fois d'un point de vue général et théorique, mais également via l'écoute : un accord harmonieux est agréable à entendre.

3. Système d'Information Géographique vocal

Plusieurs arguments militent en faveur de l'interaction entre le sonore et le vocal. Le développement des technologies de la numérisation sonore arrive aujourd'hui à un niveau de développement très important. Le langage verbal, forme de codification en signes sonores, est le code de communication privilégié chez l'humain. Le développement de la reconnaissance et de la synthèse vocale par ordinateur accélère la recherche autour de l'interaction homme-machine. Enfin, les données vectorielles sont stockées dans des bases que l'on peut interroger et fouiller programmatiquement avec un temps de réponse très court, ce qui peut compenser la dimension spatiale de la vue.

On mesure l'efficacité d'une représentation par la facilité que rencontre le lecteur à la lecture d'un dessin. Cette facilité peut se mesurer par la différence entre le temps de perception d'une construction efficace et la perception d'une construction inefficace. Le sens auditif, mode de perception préféré des malvoyants, peut être une alternative pour recevoir l'information. En fait, la différence majeure entre la perception visuelle et la perception auditive est que la première est *spatiale*, tandis que la seconde est *fréquentielle* –la fréquence étant directement liée à sa période ($f = 1/T$)– et donc *temporelle*. Les deux sens participent de manière complémentaire à une perception de courte période, voire synoptique, et une perception plus chronologique donc séquentielle. Par ailleurs, le système nerveux central a la possibilité d'extraire du signal transmis par le nerf optique, des informations non spatiales comme la couleur, la texture ou la fréquence. De même, l'oreille a la possibilité d'extraire des signaux temporels et fréquentiels transmis par le nerf auditif et des informations relatives à l'espace (Jouhaneau, 2013). Il est possible de passer d'une représentation graphique à une représentation sonore, dont l'objectif reste de véhiculer l'information géographique aux malvoyants. Certes, les règles de construction d'une représentation sonore sont différentes de celles d'une représentation graphique. Dans notre étude, nous séparons d'abord le contenu (l'information) du contenant (le moyen de transmission de l'information). Ce qui nous importe, c'est la validité et l'efficacité de la transcription du contenu, c'est à dire la rapidité et l'intelligibilité de sa perception.

4. Un simulateur cartographique vocale

Sur une carte, on trouve des primitives géométriques (points, lignes et zones) qui sont indépendantes de la dimension et de la constitution de la tâche qui les rend visible (Bertin, 1973). Par exemple, un nœud est l'un des éléments de certains modèles de données (e.g. OpenStreetMap). Mais sa représentation, en vue de faire passer un message ou de le présenter comme un objet particulier (point localisé, intersection de routes d'un réseau, hub diffusant les flux...), peut varier sémiologiquement. Dans notre système, à part l'environnement de géométrie multiple dans lequel se trouvent les lieux, nous gérons

SAGEO'2016 – Nice, 6-9 décembre 2016

essentiellement un ensemble de points de l'espace, définis par leur latitude, leur longitude, leur identifiant et un vecteur contenant les détails de ce point (nom, type de services, adresses, propriétaire..). Ces données nous intéressent le plus, car nous souhaitons les transmettre aux malvoyants afin qu'ils appréhendent leur environnement et qu'ils puissent planifier leurs déplacements de façon sécurisée. Pour être compréhensible, ces informations doivent être transformées d'un langage formel en un langage humain, via un traitement de *parsing*, puis transmises au malvoyant dans sa propre langue, et ce de la façon la plus efficace. En complément, GoogleMap offre une API de géolocalisation inverse (Reverse geocoding) pour obtenir, à partir des coordonnées GPS, l'adresse du lieu. Notre système combine donc des données géographiques libres (OSM) et des APIs de différents domaines en ligne (OSM, GoogleMap) et fait interagir, dans un même environnement, un SIG en ligne et un système de Text-To-Speech (cf. Figure 1).

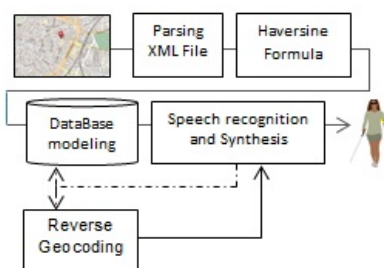


FIGURE 1. Schéma de principe et workflow

4.1. Extraction des données

Nous prenons la carte en ligne d'OSM comme source de données géographiques (Codescu *et al.*, 2011). Nous utilisons le module de reconnaissance et de synthèse vocale de Microsoft sous .NET pour l'interrogation et la communication des données pour le malvoyant.

Dans notre exemple, l'utilisateur est positionné à l'Université d'Avignon (marque rouge sur la Figure 2). À partir d'un fichier XML d'OSM (cf. Figure 3) et des attributs des objets localisés, nous pouvons extraire les positions de l'individu et des lieux environnants (boutiques, musée, soins de santé) qui sont représentés sur la carte, sous forme de *node* (nœud). Cela nous permet de calculer la distance entre chaque nœud et l'utilisateur malvoyant. En outre, nous pouvons obtenir toutes les informations importantes sur les endroits représentés par les *tags* (balises).

Le transfert des métadonnées du fichier OSM vers la base de données *Postgres* peut se faire via l'API *OSM2pgsql*. Il est important que le temps de réponse de

Interaction Homme Machine pour l'orientation spatiale des malvoyants

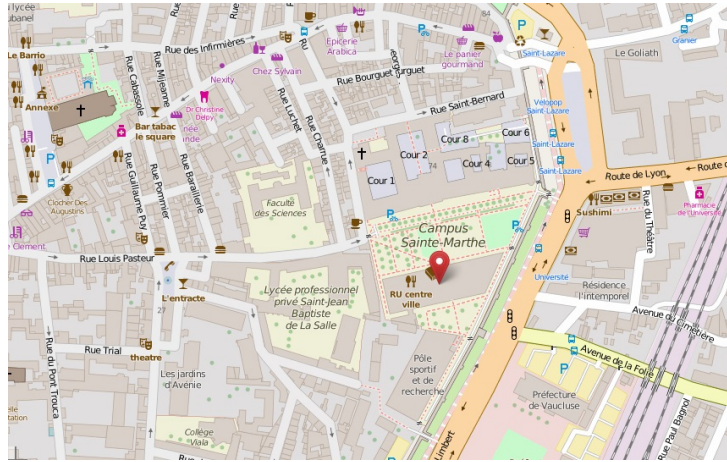


FIGURE 2. Localisation d'un point d'observation

```
<?xml version="1.0" encoding="UTF-8"?>
- <osm license="http://opendatacommons.org/licenses/odbl/1-0/"
  attribution="http://www.openstreetmap.org/copyright"
  copyright="OpenStreetMap and contributors" generator="CGImap 0.4.0 (8701
  thorn-02.openstreetmap.org)" version="0.6">
  <bounds maxlon="4.8056900" maxlat="43.9456000" minlon="4.8055300"
    minlat="43.9454300"/>
  - <node version="3" lon="4.8055510" lat="43.9454611" uid="538005"
    user="EddieJ" timestamp="2013-08-17T06:40:57Z" changeset="17379879"
    visible="true" id="2133489487">
    <tag v="parking_space" k="amenity"/>
    <tag v="1" k="capacity:disabled"/>
    <tag v="surface" k="parking"/>
    <tag v="disabled" k="parking_space"/>
    <tag v="survey 2013" k="source"/>
    <tag v="yes" k="wheelchair"/>
  </node>
  - <node version="1" lon="4.8056773" lat="43.9454860" uid="188930"
    user="J-Louis ZIMMERMANN" timestamp="2013-07-12T17:03:56Z"
    changeset="16928235" visible="true" id="2382554717">
    <tag v="waste_basket" k="amenity"/>
    <tag v="Mairie d'Avignon" k="operator"/>
    <tag v="survey 2013" k="source"/>
  </node>
  - <node version="2" lon="4.8121984" lat="43.9495394" uid="2117145"
    user="abdeldjalil" timestamp="2015-09-18T17:16:29Z"
    changeset="34107480" visible="true" id="2499310786">
    <tag v="Traiteur Garip" k="name"/>
    <tag v="deli" k="shop"/>
    <tag v="survey 2013" k="source"/>
    <tag v="no" k="wheelchair"/>
  </node>
  - <node version="4" lon="4.8119055" lat="43.9495545" uid="1068293"
    user="x0r0" timestamp="2014-05-04T15:01:37Z" changeset="22127250"
```

FIGURE 3. Exemple de fichier XML d'OSM sur Avignon

chaque requête soit court, si l'on veut une représentation efficace, d'où l'utilisation des bases de données pour stocker l'information environnementale.

SAGEO'2016 – Nice, 6-9 décembre 2016

	id	lat	lon	id_node	tags
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	1	545669099	53579766	152546	{highway,crossing}
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	2	545829480	53527143	152590	{crossing,traffic_signals,highway,traffic_signals}

FIGURE 4. API OSM vers postgres et extrait de la table Nodes avec les Tags

La table des nœuds contient la *latitude* et la *longitude*. On peut donc appliquer les lois trigonométriques pour calculer la distance et l'orientation (Zdeb, 2010). On fait appel à l'API Google de géolocalisation inverse pour trouver les adresses des lieux. La table *Nodes* contient aussi le vecteur *Tags*. Ce dernier contient les détails introduits par les contributeurs d'OSM concernant chaque lieu (cf. Figure 4).

4.2. Calcul des distances entre le malvoyant et les lieux

Nous appliquons les lois de la trigonométrie dans l'espace. La formule *Haversine* nous permet de calculer la distance entre deux endroits en utilisant leur latitude et de leur longitude (Zdeb, 2010). Elle donne une distance à vol d'oiseau approximative, mais acceptable pour l'appréhension de l'espace géographique. Selon la formule *Haversine*, on calcule plusieurs paramètres a , puis c et la *distance*:

$$a = \sin^2(\delta\phi/2) + \cos\phi_1 \cdot \cos\phi_2 \cdot \sin^2(\delta\lambda/2) \quad (1)$$

$$c = 2 \cdot \arctan 2(\sqrt{a}, \sqrt{1-a}) \quad (2)$$

$$distance = R \cdot c \quad (3)$$

Avec

- $R = 6371000$ est le rayon de la terre en mètre ;
- ϕ_1 est latitude du malvoyant, (lat_1) en radian ;
- ϕ_2 est latitude du lieu (nœud) représenté sur OSM ; (lat_2) en Radian ;
- $\delta\phi = lat_2 - lat_1$ la différence des latitudes en Radian ;
- $\delta\lambda = lon_2 - lon_1$ la différence des longitudes en Radian ou (lon_2) est la longitude du lieu et (lon_1) est la longitude de le malvoyant.

On peut déduire la distance⁵ :

$$distance = \arccos(\sin(\phi_2) \cdot \sin(\phi_1) + \cos(\phi_2) \cdot \cos(\phi_1) \cdot \cos(\delta\lambda)) \cdot R \quad (4)$$

5. <http://www.movable-type.co.uk/scripts/latlong.html>

Interaction Homme Machine pour l'orientation spatiale des malvoyants

L'orientation du lieu (*bearing*) par rapport au malvoyant est calculée comme suit :

$$\theta = \arctan 2(\sin \delta\lambda \cdot \cos \phi_2, \cos \phi_1 \cdot \sin \phi_2 - \sin \phi_1 \cdot \cos \phi_2 \cdot \cos \delta\lambda) \quad (5)$$

Le résultat est un angle en radian entre $-\pi$ à π .

4.3. Recherche de tous les lieux dans un rayon donnée par le malvoyant

Supposant que le malvoyant cherche tous les lieux situés dans une distance d , les formules précédentes peuvent nous aider à calculer les distances en formulant les requêtes *sql*. En traduisant les latitudes et longitudes en distance relative d , nous pouvons ainsi trouver tous les lieux à l'intérieur du cercle dont le rayon est d .

$$r = d/R \quad (6)$$

où R est le rayon de la terre.

Les latitudes minimale et maximale à la distance d sont (Matuschek, 2010) :

$$lat_{min} = lat - r \quad (7)$$

$$lat_{max} = lat + r \quad (8)$$

Les longitudes minimale et maximale à la distance d sont:

$$lon_{min} = lon - \delta lon \quad (9)$$

$$lon_{max} = lon + \delta lon \quad (10)$$

où :

$$\delta lon = \arcsin(\sin(r)/\cos(lat)) \quad (11)$$

Le calcul des coordonnées GPS minimales et maximales pour d défini d'avance nous permet de formuler des requêtes de sélection vers la table des noeuds :

```
SELECT * FROM nodes WHERE (Lat >= lat_min AND Lat <= lat_max)
AND (Lon >= lon_min AND Lon <= lon_max)
```

4.4. Synthèse vocale et fonctionnalités du simulateur

Les caractéristiques recherchées du système sont les suivantes:

- Le moyen de communication principale est la voix ;
- Le système est *prudent* ; il arrive parfois que le système confonde les mots proche en consonance (ex: bus - base) ou interprète un bruit en un ordre ; le système demande la confirmation du choix du lieu, la sauvegarde des positions et le rayon de recherche ;

SAGEO'2016 – Nice, 6-9 décembre 2016

- Le système cherche l'*efficacité* ; pour que le processus de clarification sémantique (*grounding*) n'affecte l'efficacité du système en allongeant le temps d'interaction, on demande la confirmation de la tâche ;
- Le système est *flexible* ; le malvoyant peut changer sa position par ordre vocal, avant de chercher les lieux les plus proches de sa nouvelle position ;
- Le système est *multi-processus* ; pour rendre le système plus efficace, on crée deux processus de reconnaissance vocale, un pour les options du système, l'autre pour rechercher les lieux, le schéma de la figure 5 montre les attributs linguistique de chaque processus.

```
// To define list of options and list of search.independent each from other.
private Choices optionChoices = new Choices();
private Choices searchChoices = new Choices();

// Two speech recognition Engine, one for options the other to select and manipulate data from DB
private SpeechRecognitionEngine search_option_recEngine = new SpeechRecognitionEngine(CultureInfo.CurrentCulture);
private SpeechRecognitionEngine option_recEngine = new SpeechRecognitionEngine(CultureInfo.CurrentCulture);

// Speech synthesis, translate Text to Speech
SpeechSynthesizer speechSynth = new SpeechSynthesizer();
PromptBuilder promptBuild = new PromptBuilder();
```

FIGURE 5. création des commandes vocales sous C#

La synthèse vocale est la technique informatique de synthèse sonore qui permet de transformer un texte en parole artificielle. Le physicien théoricien britannique Stephen Hawking est l'une des personnes les plus connues utilisant ce genre de technique pour communiquer. Il faut noter que «*les documents numériques accessibles aux malvoyants par lecteur d'écran nécessitent que les images soient décrites*» (Allaire, 2012). Parmi les lecteurs de contenu d'écran sur le web, citons *TalkBack* de Firefox, *ChromeVox* de Chrome ou *VoiceOver* d'Apple.

La Figure 6 montre les deux processus STT (Speech To Text) que nous avons implémentés : à gauche, il concerne les options du système ; à droite, il concerne la recherche dans la base des données. Pour que le malvoyant passe d'une commande de catégorie (A) à une commande de catégorie (B), il doit activer le processus de reconnaissance correspondant. Voici un exemple de dialogue.

- Le malvoyant dit *Option Système* ; le simulateur répond *Options Système activées* ;
- Le malvoyant dit *Plus grand* ; le simulateur répond *Le rayon de recherche est de 600 mètres* ;
- Le malvoyant dit *Chercher dans la base* ; le simulateur répond *Recherche dans la base activée* ;
- Le malvoyant dit *Restaurant* ; le simulateur répond *Nous avons trouvé N restaurants* ;
- En revanche, si le malvoyant dit *Plus grand*, le simulateur ne répond pas car cet attribut n'appartient pas à *Chercher dans la base*.

Interaction Homme Machine pour l'orientation spatiale des malvoyants

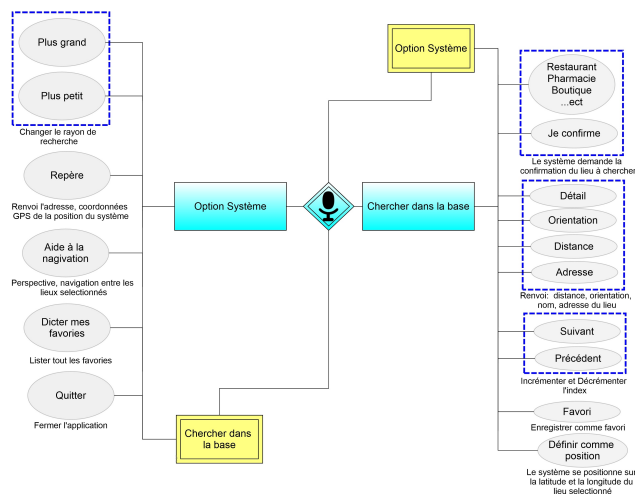


FIGURE 6. *Modèle de dialogue Homme-Machine dédié aux malvoyants*

Pour résumer, le modèle de dialogue est simple. Le malvoyant peut choisir entre deux grandes fonctions :

Option système A ; les options sont : *augmenter ou diminuer le rayon de recherche, connaître son propre emplacement et lire la liste de ses lieux favoris.*

Chercher dans la base B ; la recherche dans la base exige une ontologie de recherche pour réduire les mots clés et réduire le temps de réponse. Il faut noter que les attributs d'OpenStreetMap sont en anglais et dans un langage formel. De plus, OSM permet aux contributeurs d'ajouter des attributs personnalisés. Donc une méthode sémantique de fouille dans la table des données est nécessaire pour extraire le maximum d'information et réduire suffisamment la liste des attributs vocaux du simulateur. Le processus de reconnaissance asynchrone du système (choix personnel) implique de définir la liste des mots utilisables.

4.5. *Algorithme de transcription vocale*

La figure 7 montre les cinq principales phases de construction de notre SIG vocal : *le choix de la zone géographique, l'extraction des données vectorielles, l'insertion des données dans une base Postgres, la conversion vers une base de données Mysql, l'exploitation des données par le simulateur vocal.* Le schéma de la figure 8 présente l'architecture et le fonctionnement du simulateur. Il montre la tâche de chaque commande, le défilement d'informations depuis la table des données *Mysql* jusqu'au malvoyant et la tâche de chaque commande vocale. Nous ne le détaillerons pas dans cet article.

SAGEO'2016 – Nice, 6-9 décembre 2016

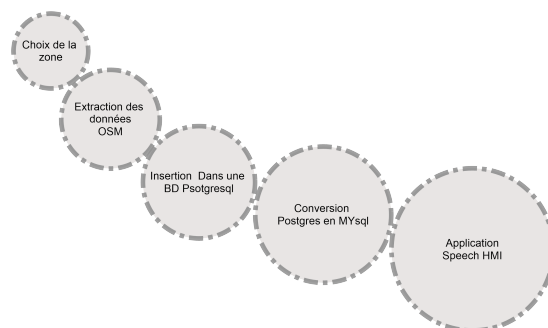


FIGURE 7. *Cycles de développement de l'information*

4.6. Exemple de requêtes

Positionné à l'université d'Avignon, l'utilisateur envoie une requête vocale pour trouver des restaurants. La figure 9 montre les restaurants situés dans un rayon de 500 m. L'interface est organisée autour d'une boussole qui donne l'orientation. Il est également possible d'intégrer dans la requête une condition d'angle de recherche. À terme, celle-ci sera reliée à un capteur embarqué par le malvoyant pour qu'il puisse disposer des informations vocales lors de son déplacement réel.

5. Conclusion

Notre proposition permet potentiellement, en utilisant des données de type OSM au sein d'un SIG vocal, d'aider les malvoyants à se mouvoir dans leur environnement. Ce prototype sera testé prochainement avec des personnes d'associations de malvoyants. Les différences de perception et d'usage entre malvoyants et aveugles seront évaluées. En effet, l'intelligence artificielle a fait progresser l'interaction homme-machine, ce qui peut être bénéfique à la compensation du handicap chez certaines personnes à mobilité réduite. Le temps de réponse du simulateur, présenté dans cet article, est très court (moins de 2 secondes), mais on ne peut pas encore conclure avec certitude que la représentation sonore, ou plus précisément vocale, est de même efficacité qu'une représentation graphique. En effet, le nombre des attributs d'OpenStreetMap risque d'allonger le temps de réponse et aussi de rendre le système plus confus. La construction d'une ontologie pour la recherche d'informations à partir d'une base de données est ainsi probablement nécessaire pour remédier à ces problèmes. L'application de la synthèse et la reconnaissance vocales montre tout de même qu'il est possible de représenter les données vectorielles par la parole sans perdre le sens de l'information qu'elles portent, avec un système souple et adaptatif. En perspective, nous pensons placer le simulateur vocal au cœur

SAGEO'2016 – Nice, 6-9 décembre 2016

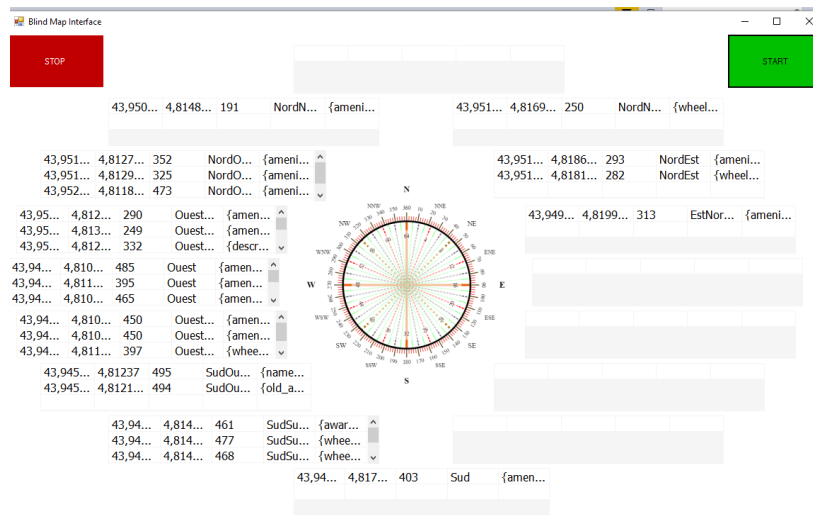


FIGURE 9. Recherche de restaurants dans un rayon de 500 m (localisation dans l'enceinte de l'Université d'Avignon)

- CFPSAA. (2010). *Les besoins des personnes déficientes visuelles*. Rapport technique. Confédération Française pour la Promotion Sociale des Aveugles et Amblyopes.
- Cha J. S., Lim D. K., Shin Y.-N. (2013). *Design and implementation of a voice based navigation for visually impaired persons*.
- Codescu M., Horsinka G., Kutz O., Mossakowski T., Rau R. (2011). DO-ROOM: activity-oriented search and navigation with openstreetmap. *GeoSpatial Semantics: 4th International Conference, GeoS, Brest, France, may 2011*.
- De Saussure F. (1997). *Cours de linguistique générale* (C. Bally, A. Sechehaye, Eds.).
- Gaver W. W. (1992). The affordances of media spaces for collaboration. *Proceeding CSCW '92 Proceedings of the 1992 ACM conference on Computer-supported cooperative work* Pages 17-24.
- Hatwell Y. (2006). Appréhender l'espace pour un enfant aveugle. *CAIRN.INFO*.
- Josselin D., Saidi A., Roussel D., Boularouk S., Bonin O., Altman E. *et al.* (2016). Sonorous cartography for sighted and blind people. *Proceeding of AGILE'2016, Helsinki, 14-17 June, 2016*. (6 pages, awarded best paper)
- Jouhaneau J. (2013). Perception de l'espace et immersion - perception auditive. *Techniques de l'ingénieur*.
- Kushwaha A., Kushwaha V. (2011). Location based services using android mobile operating system. *International Journal of Advances in Engineering & Technology*.
- Matuschek J. P. (2010). *Finding points within a distance of a latitude longitude using bounding coordinates*. Consulté sur <http://janmatuschek.de/LatitudeLongitudeBoundingCoordinates>

Interaction Homme Machine pour l'orientation spatiale des malvoyants

- Miura T., Ebihara Y., Sakajiri M., Ifukube T. (2011). Utilization of auditory perceptions of sounds and silent objects for orientation and mobility by visually-impaired people. IEEE International. (Institute of Gerontology, the University of Tokyo, Bunkyo-ku, Japan)
- Tenbrink T., Wiener J. M., Claramunt C. (Eds.). (2013). *Representing space in cognition: Interrelations of behaviour, language, and formal models*. Oxford University Press, Linguistics, 08 Exploration in Language and Space.
- Virole B. (2009). *Surdit  et sciences humaines* (Chapitre 8, L'Harmattan  d.).
- Zdeb M. (2010). Driving distances and times using sas and google maps. SAS Global Forum 2010.

Méthodologie de localisation de systèmes fixes d'identification de plaques minéralogiques sur le réseau routier

Marie Trotta¹, Jean-Paul Donnay²

1. Institut Belge pour la Sécurité Routière, Bruxelles
marie.trotta@ibsr.be
 2. Unité de Géomatique, Université de Liège
jp.donnay@ulg.ac.be
-

RÉSUMÉ. Malgré un recours croissant à la technologie des caméras ANPR, en particulier dans les investigations criminelles, la littérature et l'expérience des forces de police avec lesquelles nous avons travaillé, semblent montrer qu'il n'existe pas de méthode systématique afin de localiser le plus adéquatement possible ces installations coûteuses. Partant d'un ensemble de délits localisés, relevant de quatre catégories de vols perpétrés sur une année au sein de la province de Liège (Belgique), nous proposons une méthodologie exploitant, dans une analyse multicritère, la densité de délits, la densité de voirie et la distance entre les sites des caméras. L'approche multicritère permet d'ajuster la pondération des facteurs selon les types de délits. En outre, la méthodologie est itérative, dans la mesure où les sites des caméras sont localisés par phases successives. Un critère de validation des résultats est fourni sous la forme d'un taux de couverture du territoire considéré, qui est fonction du nombre de sites de caméras retenus et du temps de parcours depuis chacun de ces sites.

ABSTRACT. Despite increasing use of ANPR cameratechnology, especially in criminal investigations, literature and the experience of the police with whom we worked, seem to show that there is no systematic method to locate as adequately as possible these expensive facilities. Given a set of four categories of robberies over one year localized within the province of Liège (Belgium), we propose a methodology based on a multi-criteria analysis, making use of density offenses, road density and the distance between the sites of the cameras. Multi-criteria approach can adjust the factorweightings for different types of offenses. In addition, the methodology is iterative, insofar cameras sites are located in successive phases. One result validation criteria is provided in the form of a territorial coverage rate considering the number of selected camera sites and the travel time from each of these sites.

MOTS-CLÉS : caméras ANPR, profilage géographique, ingénierie du trafic, estimation à noyaux de densité, analyse multicritère

KEYWORDS: ANPR cameras, geographic profiling, traffic engineering, Kernel Density Estimation (KDE), multi-criteria analysis

Localisation des caméras ANPR

1. Principes et utilisation des caméras ANPR

Les caméras dites « ANPR », pour *Automatic Number Plate Recognition*, constituent un moyen d'identifier les plaques minéralogiques des véhicules en temps réel et, moyennant une connexion à une base de données adéquate, de repérer et éventuellement de reconstituer des portions de trajets de véhicules sur le réseau routier. Leur utilisation a été imaginée dans les années 1990, mais leur banalisation n'est apparue qu'à partir des années 2000 tant pour l'analyse criminelle (PA Consulting, 2004) que pour la gestion du trafic (Friedrich *et al.*, 2008). Il existe des solutions mobiles, les caméras étant installées sur des véhicules, ou des installations fixes sur des poteaux ou des portiques. Une caméra couvre en pratique une voie dans un ou dans deux sens de circulation. Les données qu'elle fournit, géolocalisées et minutées, peuvent être enregistrées à des fins de traitements ultérieurs. Les services de police peuvent ainsi utiliser ces caméras de plusieurs manières, tant en temps réel, qu'en temps différé. Dans ce dernier cas, les données, issues de caméras fixes, sont stockées dans un entrepôt où des applications, de type OLAP et fouille de données, permettent de répondre à des préoccupations décisionnelles, notamment en matière de criminalité (College of Policing, 2013).

C'est dans ce contexte des installations fixes et du traitement de données a posteriori que la problématique de la localisation optimale des caméras ANPR se pose avec le plus d'acuité car il faut, avec un nombre limité d'installations, couvrir le territoire ciblé et innervé par le réseau routier. De plus, pour des questions évidentes de coûts, il est tentant de vouloir exploiter l'infrastructure installée, à la fois pour les besoins de la gestion du trafic et ceux de la recherche criminelle. Or les deux applications ne réclament pas nécessairement les mêmes critères de localisation des caméras.

2. Un état de l'art lacunaire...

Si les caméras ANPR font l'objet de publications relativement nombreuses, tant du point de vue technique que dans les médias, la littérature scientifique en la matière est par contre assez rare et très déséquilibrée. Outre les articles strictement techniques, décrivant le matériel et ses caractéristiques de fonctionnement, la plupart des contributions scientifiques portent sur les algorithmes de reconnaissance des caractères alphanumériques des plaques d'immatriculation (Rhead *et al.*, 2012 ; Gurney *et al.*, 2013). La problématique de la localisation des caméras n'est pas abordée explicitement, d'une part parce que bon nombre d'études utilisent des caméras mobiles, et d'autre part parce que dans le contexte de la gestion de trafic, les sites dits « stratégiques » sont identifiés soit de manière empirique (essentiellement selon les comptages routiers et les statistiques d'accidents de la route – PA Consulting, 2004 ; Highways England, 2016), soit sur base des théories de l'ingénierie du trafic (Homburger *et al.*, 2007), éventuellement réduite aux transports publics (Matysiak *et al.*, 2015).

Les instances publiques responsables des infrastructures ne sont pas avares de publications à destination des services techniques. Ainsi en Grande-Bretagne, le

SAGEO'2016 – Nice, 6-9 décembre 2016

Home Office a publié des standards nationaux pour les données, l'infrastructure et l'accès aux données des caméras ANPR (Home Office, 2014). En outre, le pays dispose d'un centre national dédié à la gestion et au traitement des données (National ANPR Data Centre) récoltées par plusieurs milliers de caméras établies sur le territoire national. Vis-à-vis du grand public, il n'est pas rare de voir publier sur Internet des listes de caméras localisées (ANPR et autres sans distinction, par exemple Hounslow, 2016), d'autant que leur installation a posé et pose toujours débat quant à la protection de la vie privée (Home Office, 2011). Néanmoins, le rôle croissant des caméras ANPR dans la lutte contre la criminalité – et aujourd'hui le terrorisme – fait que leur nombre et leur localisation précise, et a fortiori les critères prévalant à leur installation, ne sont pas divulgués. Le message est toujours le suivant : « *In line with national policy, we do not disclose details of our fixed locations as this information is likely to be of benefit to offenders and if known could reduce the value of ANPR to policing* » (Police.UK, 2016). Pour l'anecdote, le même message est repris *in extenso* sur les sites d'autres pays utilisant la même technologie (par exemple CCTV Dubai, 2015).

La dimension géographique du problème n'est pourtant pas ignorée. La manière dont les caméras ANPR peuvent aider la recherche criminelle est exposée succinctement pour justifier auprès du grand public leur installation et leur densité : « *Geographical searches produce a list of VRMs [Vehicle Registration Marks] that have appeared in a particular area, or areas, within specified timeframes. Each geographical area is defined by a centre point and a circle drawn around it using a defined radius. All VRMs that appear at a location, or locations, within the specified timeframe are included in the result set. This search might help to identify associates of an offender, or the vehicles they are known to use. Where more than one vehicle may have been used to leave the scene of an incident, this may help to link a VRM witnessed at the scene to that of an identified associate.* » (College of Policing, 2013). Mais une fois encore, les méthodes utilisées pour sélectionner les localisations adéquates ne font l'objet d'aucun commentaire.

3. Objectif et plan de l'étude

3.1. Objectif

L'objectif poursuivi par cette étude est le positionnement des sites de caméras fixes de surveillance ANPR sur le réseau routier d'un territoire administratif donné. L'opportunité nous en a été donnée par les services du gouverneur de la Province de Liège (Belgique) qui, pour répondre à la demande des polices locales, ont souhaité disposer d'une méthodologie de localisation des caméras ANPR assurant une couverture de l'ensemble du territoire de la province de Liège (3.862 km²). L'installation des caméras ANPR doit permettre de suivre des trajets empruntés par

Localisation des caméras ANPR

les véhicules suspects, d'alimenter les procédures de profilage géographique¹ exploitées dans la résolution des délits graves et, accessoirement, de participer à la surveillance et la gestion de la circulation routière en général.

Étant donné le coût et le temps d'installation de telles caméras, la sélection des sites favorables doit en outre être proposée par phases successives, tout en garantissant un taux de couverture croissant (Trotta et Donnay, 2015).

3.2. Données de l'application

Au moment de l'étude², les différentes catégories de vols constituaient la préoccupation majeure des services de police locaux. Bien qu'il ne s'agisse pas du type de délit idéal pour la problématique de la localisation des caméras ANPR (cf. section 4.1), les données étaient disponibles et en nombre suffisant pour tenter l'établissement d'une méthodologie. Les services de police nous ont ainsi fourni les données relatives à quatre types de vols : (1) cambriolages qualifiés dans les habitations, (2) vols à main armée dans les commerces, (3) vols avec violence dans les habitations et (4) vols de véhicules, soit au total 12.353 faits recensés et localisés entre octobre 2013 et 2014 sur le territoire de la province de Liège (table 1).

Type de délits	Nombre de faits localisés	Nombre de sites localisés	% de sites
Cambriolages qualifiés dans les habitations	10.387	4.844	71,3
Vols à main armée dans les commerces	150	126	1,9
Vols avec violence dans les habitations	121	121	1,8
Vols de véhicules	1.695	1.694	25,0
Total	12.353	6.785	100,0

TABLE 1. Nombre de délits par catégories, entre octobre 2103 et octobre 2014

D'autre part, nous avons accès à la base de données du référentiel PICC de la Région wallonne (SPW, 2016), reprenant toutes les informations requises sur le réseau de voirie (à l'exception du trafic) et les adresses postales couvrant le territoire provincial. Cependant, pour une question de confidentialité, la localisation des délits ne nous a été transmise qu'au niveau du nom de rue, et comme plusieurs délits peuvent apparaître par rue, ce sont au total 6.785 « sites » de délits qui ont été géocodés en coordonnées Lambert belge 72 sur les axes de voirie (table 1).

¹«Geographic profiling [...] was originally defined as a criminal investigative methodology for analysing the locations of a connected series of crimes in order to determine the most probable area of offender residence» (Rossmo, 2000).

²Avant les épisodes dramatiques liés aux activités terroristes sur le territoire belge.

SAGEO'2016 – Nice, 6-9 décembre 2016

Il ne nous a pas été possible de disposer des adresses postales des auteurs des délits solutionnés, ni des données sur les comptages routiers sur la période considérée, toutes données qui auraient permis d'affiner les valeurs de certains paramètres utilisés dans la méthode. De même, la localisation des portiques déjà installés et équipés de caméras ANPR sur le territoire de la province de Liège et celui des provinces et pays limitrophes, n'a pu être renseignée que de façon très partielle. Ces divers manquements dans les données auraient un impact certain sur les résultats si ceux-ci devaient être considérés du point de vue opérationnel. Cependant, l'étude avait tout une portée méthodologique et les données ne sont considérées que comme un cas d'école.

3.3. Plan

L'étude est réalisée en trois étapes et incorpore, dans la dernière étape, un processus itératif de sélection des sites où peuvent être installées les caméras :

- L'analyse des zones concentrant un nombre important de délits (*hotspots*), selon les catégories de délits sélectionnées par les forces de police.
- L'identification des itinéraires susceptibles d'être fréquemment utilisés pour se rendre à / sortir de ces *hotspots* depuis tout point de la province – sur le réseau routier – et depuis les frontières de la province.
- Le positionnement des sites de caméras sur les tronçons routiers les plus utilisés dans les itinéraires d'accès/sortie des *hotspots*, en tenant compte de la densité de délits et de la distance entre les sites de caméras. Comme énoncé ci-dessus, ce positionnement doit se faire par phases (10, 20, 30, 40 et 50 sites).

4. Analyse des données criminelles

4.1. Types de délits concernés

Tous les types de délits ne sont évidemment pas susceptibles d'être enregistrés par des caméras ANPR. L'analyse criminelle présuppose, de la part du criminel, un déplacement suffisamment long (en kilomètres et/ou en temps) au moyen d'un véhicule immatriculé de façon reconnaissable. De plus, l'approche repose sur l'une des deux hypothèses suivantes :

- Soit l'identification d'un véhicule à des endroits et des temps déterminés permet de faciliter le traçage d'un véhicule suspect et, finalement, de délimiter une zone de repli vraisemblable du criminel, aussi limitée que possible. En ce sens, l'installation d'un réseau de caméras ANPR est susceptible d'alimenter les procédures de profilage géographique de l'analyse criminelle.
- Soit l'identification récurrente d'un même véhicule dans les environs d'un site criminel et à des moments jugés propices permet de suspecter le véhicule d'avoir participé à la préparation et/ou la réalisation du délit.

Il résulte de ces contraintes et objectifs, que l'efficacité des caméras ANPR dans les investigations criminelles dépend, bien sûr, du nombre de caméras installées, mais

Localisation des caméras ANPR

aussi, d'une part, de la « densité » du réseau routier (connectivité et longueur cumulée sur un territoire de superficie donnée) et, d'autre part, de la propension des criminels à effectuer de longs trajets. Les criminels agissant à proximité de leur lieu de résidence, au comportement dit « maraudeur » (Meaney, 2004 ; Paulsen, 2007), ne sont donc pas la cible idéale des investigations utilisant les caméras ANPR. Or, les cambriolages dans les habitations relèvent surtout de cette catégorie (Snook, 2004), ce qui explique pourquoi le jeu de données sélectionné pour nous par la police ne constitue pas le meilleur échantillon qui soit.

4.2. Identification des « hotspots » de criminalité

En première analyse, on pourrait considérer que les endroits où apparaissent de nombreux délits constituent de bons candidats à la localisation des caméras ANPR. En réalité, il ne faut pas se focaliser sur ces endroits eux-mêmes, mais plutôt sur les chemins qui y mènent ou qui permettent de s'en échapper. Il reste que ces « nœuds de concentration des délits » ou *hotspots* doivent d'abord être identifiés. La technique utilisée en analyse criminelle pour les mettre en évidence est bien maîtrisée, et consiste à transformer le nuage d'événements ponctuels localisés en une surface continue par un estimateur à noyaux de densité (*KDE* pour *KernelDensity Estimation*), puis à faire correspondre les sommets de cette surface aux *hotspots* recherchés (Chainey *et al.*, 2008, Johansson *et al.*, 2015). Le *KDE* permet d'établir la densité de faits dans un voisinage, lui-même défini par une distance de recherche qualifiée de largeur de la bande passante (figure 1). L'estimateur calcule la densité comme une somme pondérée de faits où chaque poids lié à un fait est inversement proportionnel à la distance séparant la localisation du fait à celle du centre du noyau. La valeur est calculée de proche en proche, par convolution du voisinage dans une grille couvrant l'espace d'étude. La distribution des poids dépend de la fonction de décroissance choisie, tandis que la valeur de densité est également influencée par la taille de la grille spatiale et la largeur de la bande passante.

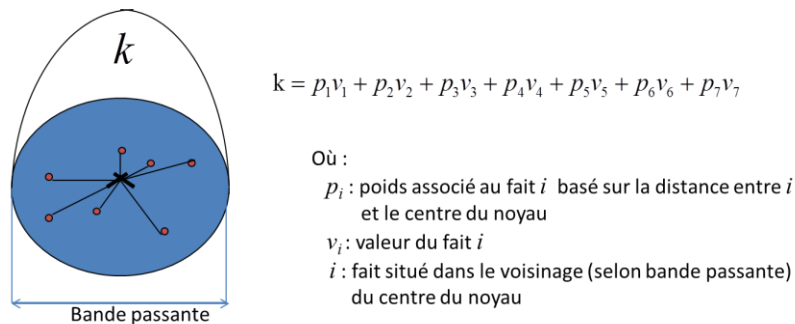


FIGURE 1. Principe du calcul d'une densité de noyau (fonction gaussienne)

SAGEO'2016 – Nice, 6-9 décembre 2016

Ce dernier paramètre est jugé important par la littérature (Chainey, 2013) et requiert dès lors un calibrage approprié selon la thématique analysée. Dans cette étude, nous postulons que la bande passante ne peut être supérieure à la distance moyenne parcourue sur le réseau routier par les criminels, afin d'être en mesure de discerner des *hotspots* associés à différentes origines.

Ces valeurs varient fortement d'un territoire à l'autre mais sont généralement de l'ordre de 6 km pour les délits considérés (Vandeviveret *al.*, 2015). Nous avons procédé à une estimation des distances parcourues par les auteurs des faits sur base des statistiques des faits résolus fournies par la police (nous n'avons pas eu accès au détail des données, mais seulement à des valeurs statistiques calculées à notre demande). On notera tout d'abord que le pourcentage de faits résolus (8,2%) est faible et pourrait ne pas être représentatif de l'ensemble des comportements. En effet, une origine plus lointaine de l'auteur pourrait être associée à une difficulté additionnelle de résolution de l'enquête.

Pour l'échantillon de faits retenu par la police, la distance moyenne parcourue par les auteurs est de 6,7 km et la valeur médiane se situe dans la classe 4 à 6 km (figure 2). Conformément à ce que propose la littérature, la bande passante choisie de 2,5 km est inférieure à ces deux valeurs. Les distances moyennes pour chaque type de faits sont supérieures à 5 km et les distances médianes supérieures à 2,5 km. Cependant, une incertitude persiste à propos des cambriolages dans les habitations, qui représentent la plus grande proportion de faits (> 70 %). Leur valeur médiane est comprise dans la catégorie de distances de 2 à 4 km, c'est-à-dire du même ordre de grandeur que la bande passante choisie. Cette situation limite n'est pas pénalisante pour la suite de la procédure, mais illustre bien le caractère circonscrit de ce type de délits et, par voie de conséquences, sa faible adéquation avec les techniques d'investigation utilisant des caméras ANPR. Le KDE utilisant une fonction gaussienne est calculé avec cette bande passante de 2,5 km sur les éléments linéaires du réseau de voirie (KDE linéaire utilisant les distances le long du réseau) préalablement rastérisés à une résolution de 30 m.

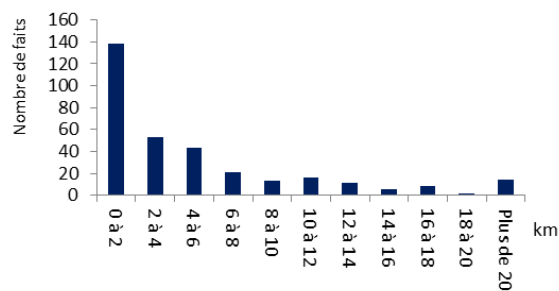


FIGURE 2. Répartition des distances parcourues par les auteurs de cambriolages

Localisation des caméras ANPR

Les quatre surfaces de densité ainsi engendrées présentent une grande stabilité quant à la position de leur maxima. Cela s'explique aisément par la nature des délits considérés qui, en dernière analyse, sont largement corrélés à la densité de population. Ceci nous a incités à construire une surface de densité globale, prenant en compte tous les faits localisés avec les mêmes paramètres, mais non limitée cette fois aux seuls éléments linéaires. C'est au départ de cette surface globale que sont identifiés et localisés les *hotspots*. La procédure utilisée pour ce faire consiste tout d'abord à inverser la surface (les maxima locaux deviennent les minima locaux) puis à extraire les « bassins versants » autour de chaque minimum local en limitant l'extension des bassins à 8 km pour assurer la couverture complète du territoire. 37 *hotspots* sont ainsi identifiés, concentrés dans les parties les plus densément peuplées de la province et situés à proximité immédiate des voies rapides (figure 3).

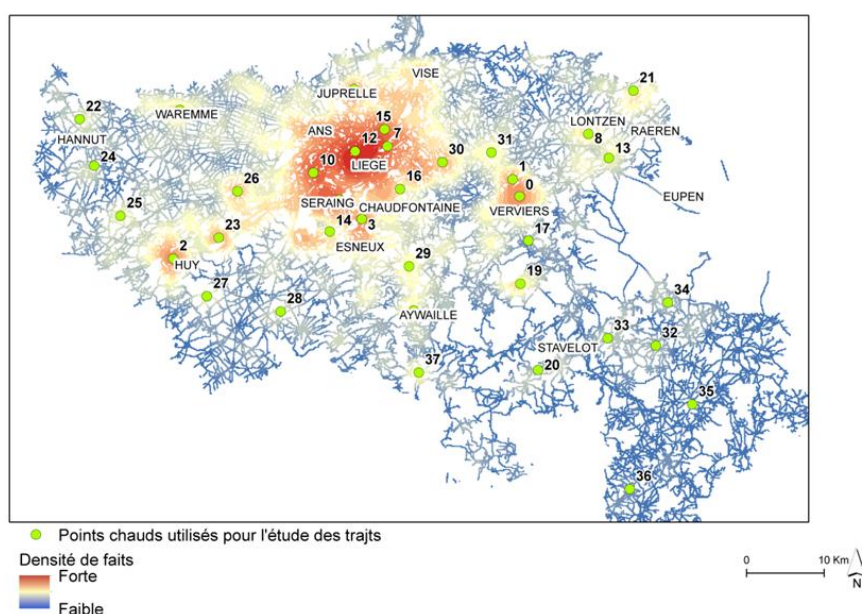


FIGURE 3. Zones de concentration et identification des points chauds (*hotspots*) pour tous les types de délits sur la province de Liège

5. Identification des plus courts chemins vers les *hotspots*

L'étape suivante consiste à repérer les tronçons et nœuds du réseau fortement empruntés pour se rendre vers les *hotspots*. Il faudrait pour cela envisager tous les chemins possibles depuis les origines, en nombre limité (les *hotspots*), vers toutes les destinations possibles au sein de la province (tous les croisements du réseau routier). Pour limiter la complexité du problème, seuls les plus courts chemins entre

SAGEO'2016 – Nice, 6-9 décembre 2016

les origines et destinations ont été considérés³. De plus, en l'absence de données précises qui auraient permis l'application des méthodes de l'ingénierie du trafic⁴, seuls sont pris en compte pour la construction des plus courts chemins, la topologie du réseau, le sens de circulation permis sur l'axe de voirie et la vitesse maximale autorisée par tronçon.

Afin de générer les plus courts chemins depuis chaque nœud (croisement) de la voirie, l'algorithme de « recherche les ressources les plus proches » est utilisé. « *Cet algorithme recherche une ou plusieurs ressources les plus proches d'un incident d'après le temps de trajet, la distance et autres coûts, et génère en sortie le meilleur itinéraire, la ressource choisie et les directions entre l'incident et la ressource.* » (ESRI, 2015).

Le hotspot est ici considéré comme l'unique ressource (origine), et chaque nœud comme un incident (destination). Le temps de trajet sur chaque tronçon (arc entre 2 nœuds) est calculé d'après la vitesse maximale autorisée sur celui-ci. Le temps de traitement étant dépendant du nombre de destinations, celles-ci ont été réduites de manière à ce que les nœuds soient au moins distants entre eux de 100 m, ce qui laisse quand même 50.590 destinations à prendre en considération. Les plus courts chemins vers toutes les destinations sont ainsi construits au départ des 37 hotspots identifiés à l'étape précédente.

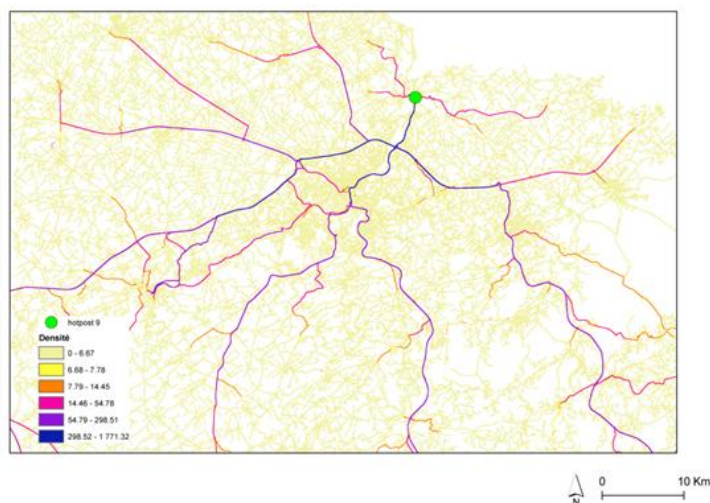


FIGURE4. *Densité de plus courts chemins vers un hotspot. Les autoroutes sont logiquement mises en évidence mais aussi des tronçons à vitesse plus faible ne pouvant être identifiés a priori*

³ Sur un territoire plus limité et strictement urbain, la prise en compte d'un indice de redondance sur les chemins les plus courts devrait être envisagé (Sevtsuk&Mekonnen, 2012).

⁴ Pour rappel, aucune donnée de trafic n'a pu être obtenue pour réaliser cette étude.

Localisation des caméras ANPR

La fréquence d'utilisation de chaque tronçon, soit le nombre de fois qu'un tronçon participe à un plus court chemin, est ensuite calculée vis-à-vis de chaque *hotspot* (figure 4). Le processus est réalisé en mode maillé à une résolution de 30 m, et il met en évidence les goulots d'étranglements (passages obligés) du réseau routier. Les inévitables effets de frontière d'un territoire limité induisent également une plus forte fréquence d'utilisation des nœuds en situation centrale, mais d'autant moins perturbante que le territoire est étendu et la densité de voirie élevée.

La densité globale des plus courts chemins vers tous les *hotspots* est obtenue par sommation entre pixels homologues des 37 images de densités calculées à l'étape précédente. Comme seules les routes avec un trafic important sont en principe candidates à un site ANPR, seuls les tronçons dont la valeur de densité globale est supérieure à un seuil sont sélectionnés. Le choix de la valeur seuil est arbitraire mais doit assurer la mise en évidence des axes innervant toute la province, en dehors des autoroutes (aux valeurs de densité systématiquement très élevées) et en particulier dans les parties périphériques de la province. La valeur seuil de 90 a été choisie en tenant compte de ces considérations.

6. Positionnement des sites ANPR

La sélection des sites de caméras doit être opérée de façon itérative, par groupes de 10 sites, jusqu'à un total de 50 sites, tout en conservant à chaque étape les sites préalablement sélectionnés. Ceci doit permettre d'échelonner dans le temps l'installation de l'infrastructure, tout en améliorant au fur et à mesure son efficacité. Rappelons qu'un « site ANPR » peut accueillir une ou plusieurs caméras ANPR selon la configuration de la voirie et les capacités techniques des caméras utilisées.

Les dix premiers sites ont été positionnés sur les autoroutes à l'entrée de la province (8), ainsi que sur les deux localisations présentant les plus fortes densités de plus courts chemins situées respectivement aux abords des deux agglomérations principales de la province (Liège et Verviers).

La sélection des sites suivants, par groupe de 10, repose sur la valeur d'un indice multicritère calculé en chaque pixel de la voirie. L'analyse multicritère reprend deux facteurs préalablement calculés : la valeur de la densité globale de délits obtenue par KDE, et la densité globale des plus courts chemins vers tous les *hotspots*. Ce dernier facteur reste contraint par la valeur seuil de 90 discutée ci-dessus. Un troisième facteur est introduit dans l'analyse pour tenir compte de la distance aux sites déjà sélectionnés aux itérations précédentes. La méthode d'agrégation multicritère est la plus simple (équation 1), dans la mesure où elle facilite l'interprétation des scénarios utilisant des facteurs et des poids distincts. Les facteurs sont préalablement ramenés sur un même intervalle et leurs poids sont normalisés (somme des poids unitaire).

$$IC = \left(\sum_{i=1}^n p_i F_i \right) \prod_{j=1}^m C_j \quad (1)$$

où :

SAGEO'2016 – Nice, 6-9 décembre 2016

IC : indice multicritère*p* : poids des facteurs*F* : contributions des facteurs*C* : valeurs (binaires) des contraintes

Les valeurs des poids présentées au tableau 2 sont déterminées selon une procédure analytique hiérarchique (AHP ; Saaty, 1980), en fonction de la structure de préférence des acteurs travaillant sur l'étude (membres de la police et chercheurs). Pour rappel, le vecteur des poids correspond au premier vecteur propre de la matrice d'inertie formée par les comparaisons de l'importance relative des facteurs deux à deux (importance du facteur en ligne vs. du facteur en colonne). Les importances relatives sont établies sur une échelle de 9 paliers (4 de part et d'autre de l'équivalence) et fixées par les acteurs de l'étude (« structure de préférence des acteurs »). Le calcul des poids peut être implémenté dans n'importe quel tableur ou logiciel statistique, tandis que le calcul des valeurs de l'indice est réalisé au moyen des fonctions d'algèbre de cartes d'un quelconque logiciel SIG capable de traiter des données maillées. La totalité de la procédure est disponible parmi les fonctionnalités d'aide à la décision spatiale de plusieurs logiciels (ex. ArcGIS, Idrisi, QGIS, etc.).

Critères	Poids des facteurs
F1 : distance aux caméras positionnées	P1 : 0,692
F2 : densité des plus courts chemins	P2 : 0,231
F3 : valeur de la densité globale du KDE	P3 : 0,077
C ₁ : densité des plus courts chemins supérieure à 90	

TABLE 2. *Facteurs et contrainte de l'analyse multicritère*

Un poids prépondérant est ici attribué au facteur de distance (F1) afin d'assurer au mieux une couverture de l'ensemble de la province. Il implique que le positionnement des sites n'est pas majoritairement dicté par la densité de délits – ce qui est conforme à la remarque faite au début de la section 4.2 – l'objectif étant d'assurer le suivi du véhicule sur un trajet au sein de la province. Dans un contexte différent (nature différente des délits, densification locale des sites ANPR, etc.), la simplicité et la robustesse du modèle multicritère utilisé permettent bien entendu de modifier cette structure de préférence, tant sur le nombre de facteurs que sur leurs poids respectifs.

Les sites sont donc positionnés d'après les résultats de l'analyse multicritères, en sélectionnant les 10 sites présentant les meilleures valeurs d'indice, et en régénérant le facteur de distance à chacune des itérations. Sont ainsi déterminés les 20, 30, 40 et 50 sites les plus appropriés (figure 6).

Localisation des caméras ANPR

7. Validation des résultats

En l'absence de caméras installées, et ne pouvant accéder aux données détaillées sur les délits résolus, il ne nous a évidemment pas été possible de valider la proposition de localisation des sites des caméras sur base de cas observés. Le seul objectif qui nous était assigné (cf. § 3.1) et qui peut faire l'objet d'une tentative de validation porte sur la couverture du territoire assurée par les sites proposés. Par conséquent, à titre de validation de la procédure, un taux de couverture des *hotspots* selon le temps de parcours depuis un site ANPR a été calculé et classé en différents pas de temps. Ce taux correspond au ratio de la somme des densités couvertes pour le pas de temps donné, divisé par la somme des densités de faits sur la totalité du territoire de la province. On remarque que les couvertures varient substantiellement pour les temps de parcours de 3 à 5 min (figure 5). Avec un pas de temps de 10 minutes, on constate que la totalité de la province est couverte dès l'installation de seulement 20 sites ANPR sur le territoire.

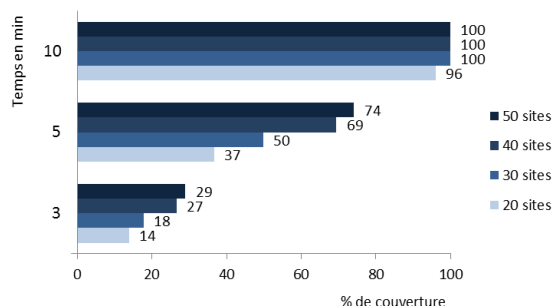


FIGURE 5. Couverture des hotspots selon le temps de trajet depuis un site ANPR

Les distances maximale et moyenne à un site ANPR diminuent progressivement à chaque itération de la procédure de sélection des sites (figure 6) : de respectivement 36 km et 19 km (pour 20 sites) à moins de 20 km et de 10 km (pour 50 sites). Ces distances restent toutefois plus élevées que les distances moyennes parcourues par les malfaiteurs pour les types de délits envisagés ici (cf. figure 2).

SAGEO'2016 – Nice, 6-9 décembre 2016

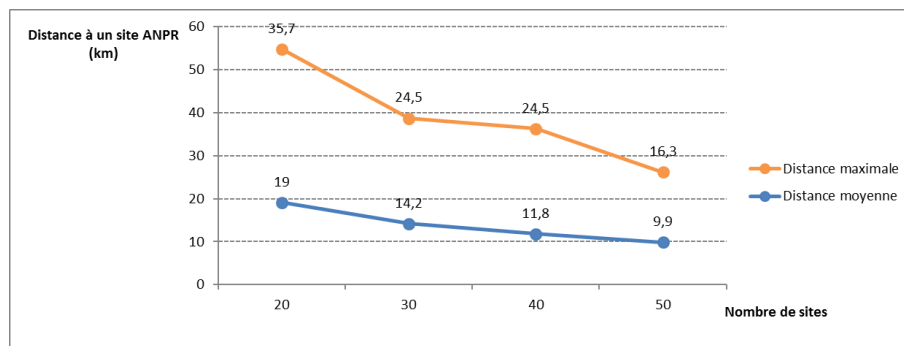


FIGURE 6. Distances maximum et moyenne au site de caméras ANPR le plus proche, selon les scénarios de 20 à 50 sites

8. Conclusions

L'absence de méthode avérée pour localiser les caméras ANPR le long du réseau routier nous a amené à proposer une approche combinant des outils classiques de la cartographie criminelle (KDE), de l'ingénierie du trafic (plus courts chemins) et de l'aide à la décision spatiale (analyse multicritère). Vis-à-vis des données traitées, relatives à différentes catégories de vols, le recours aux caméras ANPR n'est utile que dans le cas de bandes organisées parcourant de grandes distances. Mais le canevas méthodologique reste cohérent, quel que soit la nature des délits étudiés.

Les quelques biais qu'il a fallu consentir, dus principalement aux données et à la part d'arbitraire inhérente aux méthodes, constituent autant de pistes de recherche pour une amélioration de l'approche dans un but opérationnel. Ainsi, des données complémentaires et/ou plus précises devraient permettre d'affiner des paramètres importants, tels que la largeur de la bande passante et la fonction de lissage du KDE, ou les vitesses maximales praticables sur le réseau routier en fonction du trafic et de la période. Elles devraient aussi permettre de valider la méthode sur des cas de délits résolus en comparant les localisations des caméras proposées avec les parcours effectivement empruntés par les criminels. En outre, le modèle multicritère offre une grande souplesse d'adaptation dans la sélection des critères et de la fonction d'agrégation s'il fallait, par exemple, intégrer des délits de natures très différentes. C'est donc aux services de police, à qui la procédure a été présentée, qu'il revient d'affiner la proposition sur base des données détaillées qu'ils possèdent et de procéder à diverses simulations en jouant sur les paramètres du modèle multicritère.

À l'heure où les plans d'équipement de caméras ANPR se multiplient, dans le cadre d'une politique anti-terrorisme, une réflexion qui incorpore la dimension spatiale nous paraît s'imposer.

Localisation des caméras ANPR

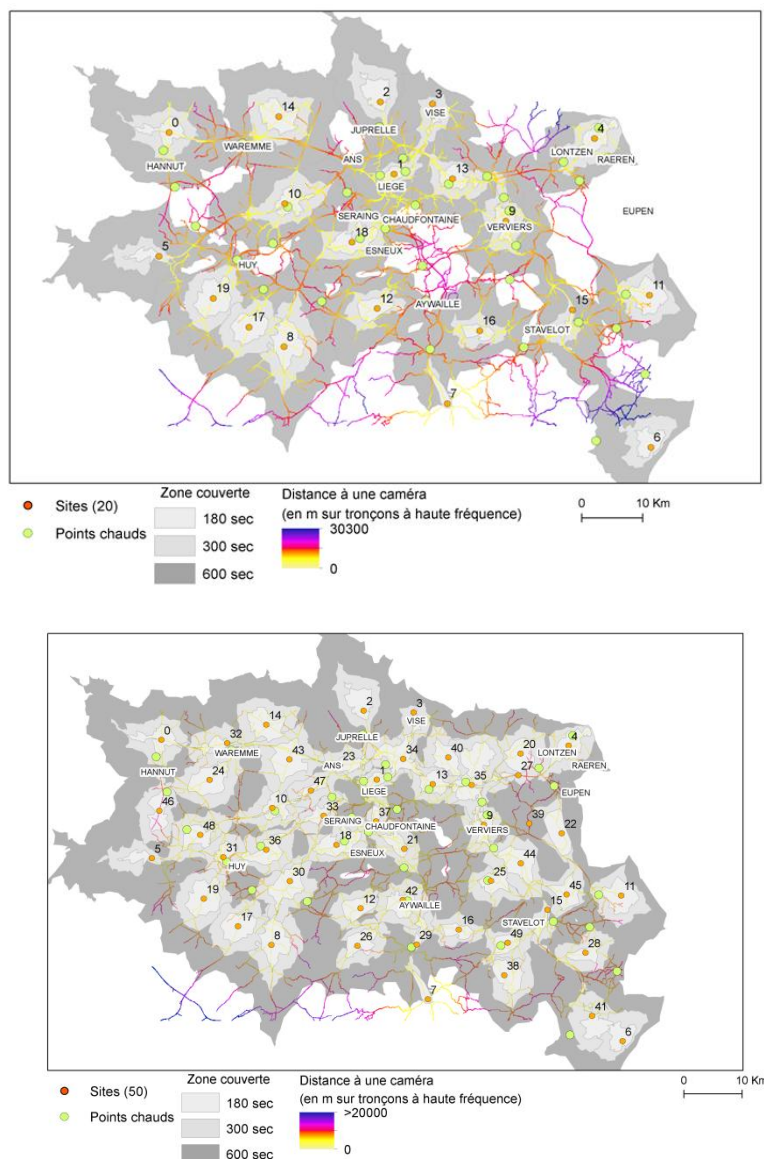


FIGURE7. Positionnement des sites ANPR et couverture par temps de parcours (20 sites en haut, 50 sites en bas)

SAGEO'2016 – Nice, 6-9 décembre 2016

Références

- CCTV Dubai (2015). *ANPR Camera Dubai*. <http://cctvdubai.biz/anpr-camera-dubai/> Consulté le 21/06/2016.
- Chainey S. (2013) Examining the influence of cell size on kernel density estimation crime hotspots maps for predicting spatial patterns of crimes, *Bulletin de la Société géographique de Liège*, 60, p.7-19.
- Chainey S. Tompson L. Uhlig S. (2008) The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime, *Security Journal*, 21 (1-2), p. 4-28.
- College of Policing (2013). *Authorized Professional Practice – Investigations – Investigative Strategies – ANPR*, Londres. <https://www.app.college.police.uk/app-content/investigations/investigative-strategies/investigative-strategiesanpr/> Consulté le 21/06/2016.
- ESRI (2015). *ArcGIS Resources*.
<http://resources.arcgis.com/fr/help/main/10.2/index.html#//004800000028000000>
Consulté le 25 mai 2015.
- Friedrich, M., Jehlicka, P. & Schlaich, J. (2008). Automatic Number Plate Recognition for the Observance of Travel Behaviour, *8th International Conference on Survey Methods in Transport*, p. 17.
- Gurney, R. Rhead, M., Lyons, V. & Ramalingam, S. (2013) The effect of ANPR Camera Settings on System Performance (poster), *IET 5th ICDP Conference Proceedings on Imaging for Crime Prevention and Detection*, Londres.
- Highways England (2016). *Cameras – Fair Processing Notice*.
<http://www.highways.gov.uk/traffic-information/traffic-information-services/traffic-cameras/cameras-fair-processing-notice/> consulté le 20/06/2016.
- Homburger W.S., Hall J.W., Reilly W.R., Sullivan E.C. (2007). *Fundamentals of traffic engineering* (16th edition), Institute of Transportation Studies, University of California, Berkeley.
- Home Office (2011). *Consultation on a code of practice relative to surveillance cameras*, Londres.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/517367/code-surveillance-cameras.pdf Consulté le 21/06/2016.
- Home Office (2014). *National ANPR Standards for Policing: Part 1 – Data Standards ; Part 2 – Infrastructure Standards ; Part 3 – Data Access and Management Standards*, p. 11, 21, 16. <http://www.npcc.police.uk/FreedomofInformation/ANPR.aspx> Consulté le 21/06/2016.
- Hounslow (2016). *List of CCTV locations*. London Borough of Hounslow.
http://www.hounslow.gov.uk/cameralistsnew30_june_2016_.pdf. Consulté le 21/06/2016.
- Johansson E., Gahlin C., Borg A. (2015). Crime Hotspots: An Evaluation of the KDE Spatial Mapping Technique, *Proceedings of the 2015 European Intelligence and Security Informatics Conference (EISIC)*, p. 69-74.
- Matysiak, A., Kruszewski, M., Niezgoda, M. & Kaminski, T. (2015) The analysis of ANPR camera location points in bus lanes monitoring system in the city of Warsaw, *Journal of KONES*, 20(4), p. 269-275.
- Meaney, R. (2004). Commuters and Marauders: An Examination of the Spatial Behaviour of Serial Criminals, Australia, *Journal of Investigative Psychology and Offender Profiling*, p. 121-137.
- PA Consulting (2004) *Driving crime down. Denying criminals the use of the road*, Report study commissioned by the Home Office Police Standards Unit (PSU), p. 170.
- Paulsen D. (2007). Improving Geographic Profiling through Commuter/Marauder Prediction, *Police Practice and Research*, 8(4), p. 347-357.

Localisation des caméras ANPR

- Police.UK (2016). *How police forces and other law enforcement agencies use Automatic Number Plate Recognition (ANPR)*. <https://www.police.uk/information-and-advice/automatic-number-plate-recognition/> Consulté le 21/06/2016.
- Rhead, M., Gurney, R., Ramalingan, S. & Cohen, N. (2012) Accuracy of Automatic Number Plate Recognition (ANPR) and Real World UK Number Plate Problems, *Proceedings 46th IEEE International Conference on Security Technology*, p. 286-291.
- Rossmo (2000). *Geographic Profiling*. Boca Raton, CRC Press.
- SPW (2016). *Géoportail de Wallonie: Projet Informatique de Cartographie Continue (PICC)*, Service Public de Wallonie, <http://geoportail.wallonie.be/home.html>
- Saaty T.L. (1980). *The Analytical Hierarchy Process*. New York :MacGraw Hill.
- Snook, B. (2004). Individual differences in distance travelled by serial burglars, *Journal of Investigative Psychology and Offender Profiling*, 1(1), p. 53-66.
- Sevtsuk, A., Mekonnen, M. (2012). Urban Network Analysis Toolbox, *International Journal of Geomatics and Spatial Analysis*, 22/2, p. 287-305.
- Trotta M., Donnay, J.-P. (2015). *Étude du positionnement des caméras ANPR en Province de Liège – Scénarios 1 et 2*, Rapports de recherche établis à la demande du Gouverneur de la Province de Liège, Unité de Géomatique de l'Université de Liège, p. 26 & 13.
- Vandeviver C., Neutens T., van Daele S., Geurts D., Vander Beken T. (2015). A discrete spatial choice model of burglary target selection at the house-level, *Applied Geography*, 64, p. 24-34.

Modèle sémantique de trajectoire d'objet mobile

Mojdeh Soltan Mohammadi¹, Isabelle Mougenot²,
Thérèse Libourel², Christophe Fagot¹

1. UMR ESPACE-DEV (IRD - UM - UAG - UR)

therese.libourel, isabelle.mougenot@umontpellier.fr

2. Intactile DESIGN

soltan-mohammadi.m, fagot.c@intactile.com

RÉSUMÉ. L'évolution des référentiels spatio-temporels et les dernières avancées des systèmes d'information géographique ont favorisé l'apparition de nouveaux types de services et d'applications liés à la localisation et à la mobilité d'entités d'intérêt. Notre approche propose un modèle de connaissances dédié à la prise en charge de trajectoire d'objet mobile afin d'en extraire des informations relatives à la détection de comportements remarquables. Nous nous appuyons sur différents travaux de recherche existants autour de la modélisation et de l'échange de données spatiales comme temporelles et nous les étendons pour répondre à la spécificité de l'analyse sémantique en temps réel de la mobilité des objets. Le paradigme de représentation retenu est d'envisager chaque objet comme indissociable de sa trajectoire de vie. Nous illustrons notre approche au travers d'un cas d'étude concret dans le domaine des systèmes de supervision et de contrôle d'objets mobiles.

ABSTRACT. The technology evolution around wireless communication and the latest advances in geographic information system (GIS) have promoted new types of services and applications related to location and mobility of prominent entities. In this work, we propose a knowledge-based model dedicated to handle the trajectory of moving objects in order to extract meaningful patterns related to the detection of remarkable behaviors. We rely on different existing pieces of research on modeling and exchange of spatial and temporal data. We extend these approaches to address the specificity of semantic analysis and real time mobility of objects. The paradigm we choose is to consider each object as inseparable from its life trajectory. We illustrate our approach through a concrete use case in domain of supervision and control systems of moving objects.

MOTS-CLÉS : Modèle spatio-temporel, trajectoire, objet mobile, ontologie

KEYWORDS: Spatio-Temporal model, trajectory, moving object, ontology

Modèles émantique de trajectoire d'objet mobile

1. Introduction

L'essor des technologies autour de la communication sans fil, l'évolution des systèmes de référencement spatiaux et temporels et les dernières avancées des systèmes d'information géographique (SIG) ont favorisé l'apparition de nouveaux types de services et d'applications liés à la surveillance d'objets mobiles. Par ailleurs, de larges volumes de données sont souvent acquis en temps réel, ou à intervalles réguliers au travers de nombreux capteurs qui permettent de retranscrire au moins partiellement le positionnement et la dynamique de ces données. Ces données peuvent, en outre, être organisées et présentées en fonction du domaine de spécialité qui en a permis l'acquisition, à l'exemple du domaine de l'urbanisme ou du transport. Ainsi, l'historisation à large échelle de l'évolution d'objets peut s'envisager sous forme de trajectoires et offre de nouvelles perspectives quant à l'analyse des comportements particuliers de ces objets. Notre approche se consacre à la définition d'un modèle conceptuel, simple et optimisé dans son niveau d'expression, qui factorise la représentation des sémantiques des dimensions spatiale, temporelle et contextuelle des objets mobiles. Elle réutilise largement l'existant et en particulier, le patron de conception ontologique proposé par (Hu *et al.*, 2013) modélisant la trajectoire sémantique d'objet mobile, le standard GeoSPARQL¹ qui offre un modèle général pour la représentation de connaissances spatialisées et le modèle OWL-Time², mis à contribution pour tout ce qui concerne les aspects temporels.

L'article est structuré de la manière suivante : la section 2 propose une synthèse de l'existant en matière de modèles de connaissances représentant la trajectoire d'objet mobile et introduit l'ontologie GeoSPARQL que notre modèle de connaissance vient étendre. La section 3 est consacrée à l'approche méthodologique suivie. Nous relatons comment les données sont acquises, puis la construction du modèle sémantique nous sert à intégrer ces données au format unifié RDF en vue d'analyses. La dernière section résume le travail accompli et en dresse quelques perspectives.

2. État de l'art

La gestion en temps réel de l'information géographique est devenue une priorité dans bon nombre de domaines liés au territoire, comme par exemple l'environnement, l'urbanisme, ou le transport. Ce constat nous amène à définir en tout premier lieu les notions essentielles à la prise en charge du temps et de l'espace dans les systèmes. Par la suite, nous présentons les modèles de connaissances spatio-temporels qui nous semblent les plus significatifs au regard des notions de trajectoire et de motifs spatio-temporels remarquables. Nous abordons en dernier lieu le standard GeoSPARQL d'OGC qui offre un cadre à la

1. <http://www.opengeospatial.org/standards/geosparql>

2. OWL-Time a été défini dans le périmètre des web services sémantiques d'OWL-S.

SAGEO'2016 – Nice, 6-9 décembre 2016

représentation des entités géospatiales en permettant notamment de distinguer les entités géographiques de leurs géométries.

2.1. Les notions de base

2.1.1. Entités spatio-temporelles et objet mobile

De nombreux travaux issus de domaines variés (Gayte *et al.*, 1995) ont proposé la définition d'objets à partir de la notion de classe. La classe regroupe les objets de même nature, qui sont alors décrits par des attributs et des comportements similaires. Chaque objet est différencié par un identifiant qui lui est propre et qui est invariant tout au long de son existence. La notion d'entité spatio-temporelle est une représentation d'une entité du monde réel, dotée d'une identité (indépendante de toute valeur) et de propriétés descriptives dites thématiques et spatiales. Le temps va impacter ces divers constituants. Une classification relative à la dynamique de ces entités a été proposée (Lardon *et al.*, 1999) afin de mieux traduire les phénomènes liés à leur évolution (ceux-ci peuvent affecter la dimension spatiale, thématique et leur identité) :

- **Objet modifiable ou changeant**, à l'exemple d'un quartier d'une agglomération qui est soumis à une transformation urbaine et dont la caractéristique thématique évolue,
- **Objet déformable**, à l'exemple, d'un port maritime dont la dimension spatiale évolue en raison de nouvelles politiques territoriales,
- **Objet transformable**, à l'exemple, d'une manifestation humaine se séparant en deux sous-ensembles.
- **Objet mobile ou mouvant**, à l'exemple, d'un piéton ou d'un bateau qui se déplace au fil du temps et à qui il est d'usage d'associer une trajectoire.

Nous nous concentrons dans ce travail sur les seuls objets mobiles qui ont pour qualité de se déplacer dans l'espace qui est alors considéré comme leur environnement. Leur déplacement se traduit par un positionnement géographique qui est amené à changer au cours du temps (Étienne, 2011). Chaque objet mobile dispose en outre de différentes caractéristiques qui lui sont propres telles que la catégorie sémantique à laquelle il est rattaché (par exemple bateau, avion, piéton), ses capacités de déplacement ou encore sa forme géométrique. Nous nous intéressons particulièrement à l'étude du mouvement d'objets mobiles évoluant dans un espace ouvert, peu contraint, et dont la taille de la géométrie est négligeable comparée à l'espace dans lequel l'objet évolue.

2.1.2. Trajectoire d'objet mobile

Le mouvement d'un objet mobile peut s'approcher de différentes façons (Sistla *et al.*, 1997). À ce titre, (Nathan *et al.*, 2008) analyse le chemin de mouvement de l'objet en trois échelles fondamentales spatio-temporelles à partir de l'ensemble des positions successives de l'objet dans l'espace et en fonction

Modèles émantique de trajectoire d'objet mobile

du temps (appelé trajectoire). Pour (Noël *et al.*, 2015), la trajectoire est dite sémantique quand elle est enrichie par différentes informations souvent thématiques qui sont utiles à la description de l'objet dans un contexte d'étude donné. Comme illustré dans la figure 1, une trajectoire consiste donc en une série de points (positions) produits à l'aide d'un référentiel (système de coordonnées) et à partir du mouvement d'un objet mobile. Ces points sont souvent représentés sous la forme d'un triplet $\{x_i; y_i; t_i\}$ (x_i et y_i désignent une position dans le plan géographique 2D, et t_i représente une estampille temporelle) ou bien d'un quadruplet $\{x_i; y_i; z_i; t_i\}$ (z_i donne l'altitude) si la trajectoire est à

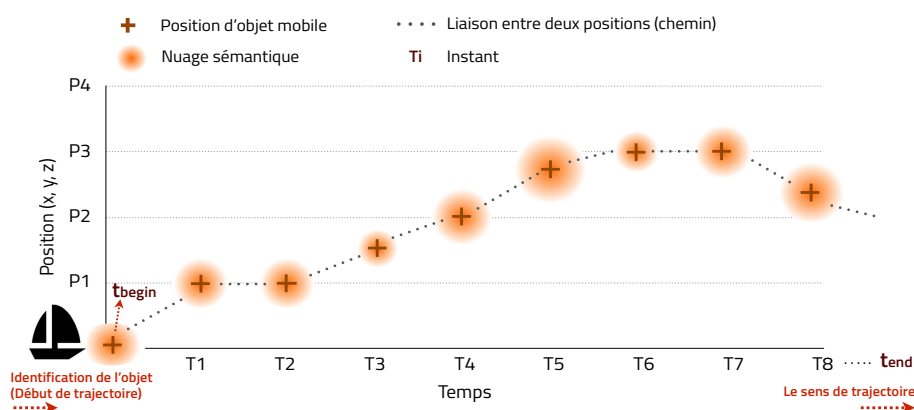


FIGURE 1. Trajectoire d'un objet mobile définie par ses positions successives acquises au cours du temps et les sémantiques additionnelles pouvant y être associées.

Par exemple, dans un contexte de supervision maritime, il est important d'avoir connaissance de la nature du mode de transport, ou encore des lieux traversés par le navire qui est sous surveillance afin d'en analyser le comportement. Ces informations complémentaires peuvent de plus provenir de sources de données tierces.

2.2. Modèles sémantiques de trajectoire

La mobilité d'un objet dans l'espace et au travers du temps peut se représenter naturellement à l'aide d'un graphe ; les phases successives de la trajectoire de l'objet correspondant aux arêtes du graphe (les positions étant les nœuds). Une telle représentation s'accorde bien à une modélisation à l'aide d'ontologies de

SAGEO'2016 – Nice, 6-9 décembre 2016

domaines qui exploitent le plus souvent le paradigme des graphes et qui réutilisent pour un contexte donné, les ontologies dédiées au temps et à l'espace. La difficulté réside dans le niveau d'expressivité de la représentation conceptuelle et formelle des objets, que cela soit au regard de la problématique considérée, ou bien des aspects spatiaux et temporels. Ainsi, la représentation d'objet mobile et de sa trajectoire doit pouvoir être ouverte et évolutive. Il s'agit non seulement, d'expliciter et de conserver l'évolution spatio-temporelle de l'objet mais également de pouvoir enrichir la trajectoire d'objet par des connaissances en fonction des besoins des analyses à mener, en particulier pour tout ce qui concerne la détection d'irrégularités dans les comportements d'objets.

Différents modèles conceptuels ont été proposés dans un passé récent pour organiser au mieux les éléments tangibles constitutifs d'une trajectoire. La notion de trajectoire est exploitable dans différents contextes à l'exemple de la supervision du trafic routier ou bien de l'étude des routes migratoires d'oiseaux et il est attendu à cet effet de disposer d'un modèle suffisamment générique pour s'adapter à toutes ces problématiques.

Le modèle proposé par (Spaccapietra *et al.*, 2008) pensé à l'origine pour le suivi de mouvement d'oiseaux migrateurs tout au long de leur périple, déconstruit les trajectoires au travers des notions d'arrêt et de mouvement. Un point fixe peut en outre marquer le début (*begin*) ou bien la fin (*end*) de la trajectoire. La trajectoire est de plus envisagée comme pourvue d'une **intention**, et à cet effet, comprend une facette sémantique en complément de la facette géométrique avec la possibilité d'attacher des annotations aux éléments de la trajectoire. La dimension temporelle est retranscrite sous la forme d'un intervalle de temps défini comme un simple attribut attaché à chaque objet de la trajectoire. Le modèle décrit dans (Baglioni *et al.*, 2008) étend les arrêts (appelé **Stop**) et les mouvements (appelé **Move**) de modèle précédent, en partant du principe que les trajectoires sémantiques nécessitent la définition d'une ontologie pour pouvoir notamment fixer des interprétations sur les intentions des trajectoires. Toujours sur la base des travaux de (Spaccapietra *et al.*, 2008), (Yan, 2011) propose une ontologie de trajectoire modulaire composée d'un module géométrique, d'un module géographique et d'un module métier. De fait, la modélisation d'une trajectoire pour un domaine d'application donné est associé à un processus de construction sémantique en plusieurs étapes. Ce travail est ensuite réutilisé dans (Vandecasteele, 2012) dans le contexte du contrôle maritime.

Les travaux présentés par (Hu *et al.*, 2013) partent du constat déjà livré dans (Spaccapietra *et al.*, 2008), à savoir que la notion de trajectoire est une notion très générale et retrouvée dans de nombreuses disciplines. Ils proposent à cet effet de modéliser cette notion définitivement au travers d'un patron de conception ontologique. Ce patron ontologique trajectoire est comme tous les patrons, non seulement réutilisable mais aussi extensible, et peut être exploité en association avec d'autres modules ontologiques. L'approche proposée par (Janowicz *et al.*, 2015) réutilise le patron précédent, en particulier en association avec le patron de mouvement de (Narock, Fox, 2015) qui prend en charge

Modèles émantique de trajectoire d'objet mobile

la représentation des objets mobiles. Dans ce contexte, un objet mobile peut avoir plusieurs trajectoires et a contrario une trajectoire peut être partagée par différents objets mobiles.

2.3. Standard GeoSPARQL

Le standard GeoSPARQL s'appuie sur les langages du web sémantique et en particulier sur SPARQL proposant différentes spécifications qui facilitent les usages pouvant être faits des données géographiques. GeoSPARQL exploite à cet effet des modèles antérieurs à l'exemple de *General Feature Model* ainsi que de *Simple Feature Model* de l'ISO 19101 (Herring, 2011) et propose des représentations adaptées aux entités spatiales, à leurs propriétés et aux fonctions de calcul (par exemple topologiques ou métriques) qui s'y appliquent. GeoSPARQL est construit de manière modulaire et comprend différents composants parmi lesquels (Lupp, 2008) :

- **Core** qui définit le vocabulaire lié aux entités spatiales de haut niveau³.
- **Extension géométrique** qui définit le vocabulaire relatif à la diversité des types de géométrie et aux fonctions non-topologiques.
- **Extension topologique** définit les relations topologiques.
- **Topologie géométrie** qui définit les fonctions SPARQL étendues à la prise en charge de calculs spatiaux, concernant les relations topologiques.

La figure 2 illustre brièvement le module Core de GeoSparql sous la forme d'un diagramme de classes UML sur lequel nous nous sommes appuyés pour notre proposition. La superclasse `SpatialObject` est spécialisée par `Feature` et `Geometry`. La classe `Feature` correspond à la classe `FeatureType` du standard ISO 19101 intitulée *Spatial schéma* et la classe `Geometry` correspond à la classe racine de la hiérarchie des types géométriques en relation avec les systèmes de référence de coordonnées spatiales. Les classes `Feature` et `Geometry` sont disjointes. Le choix est fait de distinguer une entité de ses géométries, de manière à donner de la flexibilité dans le calcul, par exemple des relations topologiques⁴ ou métriques, entre objets géométriques, permettant de percevoir et situer les objets les uns par rapport aux autres. De plus, l'entité va pouvoir être caractérisée d'un point de vue thématique. Ainsi, une **feature** est tout simplement une entité du monde réel, qui pourrait être un parc, un navire ou encore un restaurant. Une **geometry** représente toute forme géométrique, par exemple un point, un polygone, ou une ligne, et est une représentation de la position spatiale d'une **feature**, liée à au moins un référentiel géographique instance de la classe `CoordinateReferenceSystem`. À cet effet, **feature** et **geometry** sont

3. Core constitue une ontologie à part entière décrite en OWL 2

4. Elles sont prises en charge par trois vocabulaires différents nommés le *Modèle Simple Feature de l'OGC* (Herring, 2011), le *Modèle d'Egenhofer* dit des *9-intersections (DE-9IM)* et le *Modèle RCC8* (Region Connection Calculus).

SAGEO'2016 – Nice, 6-9 décembre 2016

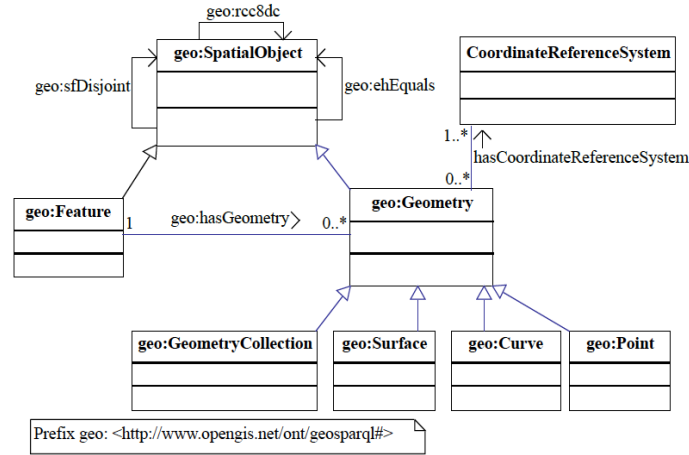


FIGURE 2. Diagramme de classes UML illustrant le modèle Core de GeoSPARQL

associées au travers de la relation `hasGeometry`. En termes de multiplicité, une `feature` peut être associée de 0 à n objet de type `geometry` et une `geometry` est à l'inverse associée à une seule objet de type `feature`. La classe `Geometry` possède différents attributs qui la qualifient. Ainsi, les deux propriétés `asWKT` et `asGML` font appel aux formats textuels `wktLiteral` et `gmlLiteral` pour faciliter le partage et l'échange des valeurs liées au type de géométrie.

Les associations réflexives `sfDisjoint`, `rcc8dc` et `ehEquals` de la figure 2, sont des exemples de relations spatiales qui permettent de relier au moins deux `SpatialObject` de type `Feature` ou de type `Geometry`. Elles vont par la suite, faciliter les comparaisons et les calculs entre ces objets. (Clementini, Laurini, 2008) a classé ces relations spatiales selon trois catégories : les relations topologiques, les relations cardinales et les relations métriques.

3. Approche méthodologique

3.1. Extraction du jeu de données

La société Intactile DESIGN dispose de plusieurs composants logiciels associés à des scénarios d'études liés à des activités humaines et notamment celle des transports. Nous nous positionnons d'emblée dans un contexte de grandes masses de données car la société dispose de différents jeux de données retranscrivant les comportements d'objets mobiles. Les données sont acquises à intervalles de quelques secondes et sous forme du flux. L'idée est d'aider à la décision relative à la supervision temps réel du trafic de véhicules. Pour ce faire, et comme

Modèles émantique de trajectoire d'objet mobile

illustré dans la figure 3, le modèle de trajectoire sémantique que nous proposons permettra l'intégration de diverses sources de données spatio-temporelles et sémantiques et ceci dans un format unifié RDF. Ces données intégrées ainsi élicitées seront ensuite exploitées pour l'extraction de motifs dans le cadre d'un système d'aide à la décision de la société.

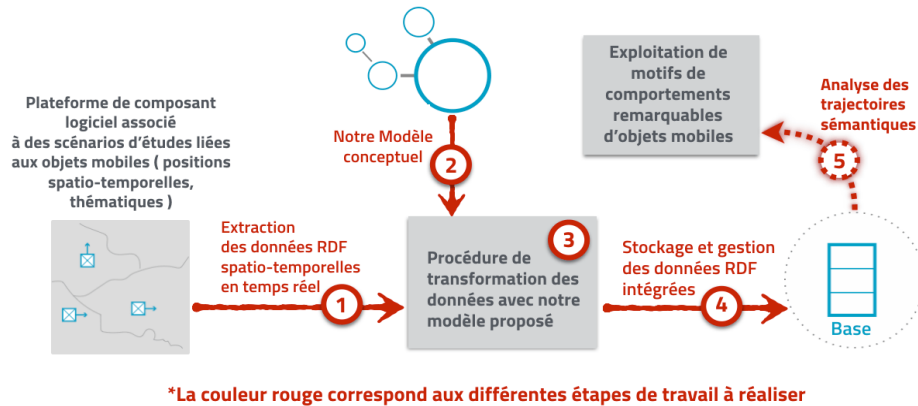


FIGURE 3. Extraction des données spatio-temporelles en temps réel

3.2. Construction du modèle intégré

Nous partons du postulat qu'un objet mobile ne possède qu'une seule et même trajectoire tout au long de sa vie (cf.figure 4). Nous nous intéressons à un modèle intégrant sémantique de l'objet et de sa trajectoire, à savoir qu'il ne s'agit pas de modéliser seulement la position de l'objet mais de prendre aussi en charge différentes informations complémentaires pouvant être liées à l'objet et à ses mouvements : conditions météorologiques, autres objets à proximité. Cette trajectoire ne peut cependant être renseignée que partiellement sur des segments spatio-temporels bien identifiés. Elle est de plus en constante évolution au fur à mesure que le temps s'écoule. Cette vision s'applique bien à notre cas d'étude temps réel dans le contexte entrepreneurial.

La plupart de modèles présentés précédemment dans l'état de l'art, dissocient les trajectoires des objets ainsi qu'ils ne prennent pas en compte la sémantique relative à l'objet correspondant à la trajectoire. L'idée est donc de disposer d'une représentation unifiée de l'objet et de sa trajectoire, de manière à en faciliter par la suite l'interrogation et l'analyse spatio-temporelle en particulier. La figure 5 donne une vue schématique de notre modèle. Nous proposons à cet effet de réutiliser l'existant et d'en décliner une extension qui va se conformer à notre vision. Ainsi nous adaptons le patron générique de trajectoire

SAGEO'2016 – Nice, 6-9 décembre 2016

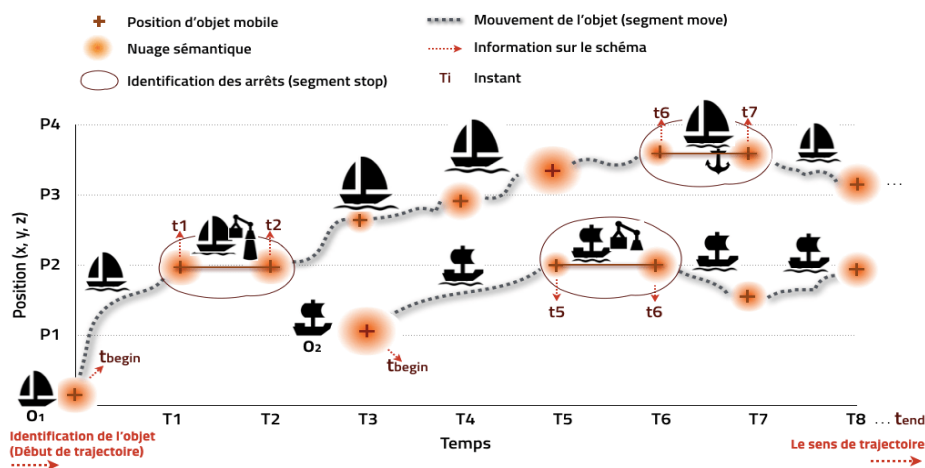


FIGURE 4. Trajectoires de deux objets mobiles définissent par des positions et des sémantiques acquises à différents intervalles de temps

sémantique, défini par (Hu *et al.*, 2013) qui réutilise par ailleurs l'ontologie pour le temporel OWL-Time. La modélisation des entités spatialisées est guidée par le standard OGC GeoSPARQL. Le patron de conception et les deux ontologies cadres nous fournissent les concepts primitifs des grands domaines investis, à savoir le spatial, le temporel et la mobilité. L'idée est ensuite de raffiner les concepts primitifs pour en produire des concepts définis qui permettent de modéliser notre domaine de manière précise. Les concepts ainsi dérivés sont munis de relations qui trouvent des prolongements soit dans l'un, soit dans l'autre des composants ontologiques de haut niveau, faisant ainsi émerger un modèle d'objet mobile et trajectoire intégré. La mise en relation des modèles cadres permet de fixer des interprétations inexploitées jusqu'alors et de faire émerger de nouvelles connaissances notamment en matière de comportements fréquents ou au contraire remarquables.

Le modèle construit prend en charge pour chaque objet les informations acquises en temps réel et l'historique des mises à jour de ses caractéristiques qui seront ensuite utilisées lors des phases d'analyse. Bien que conçu à partir de modèles préexistants, le modèle proposé intégré précise les divers concepts pris en compte. Nous étendons la classe `Feature` du modèle Core de GeoSPARQL pour prendre en charge l'entité mobile (`MovingObject`). Un `MovingObject` possède des attributs comme un nom et des caractéristiques thématiques. À l'aide du lien `hasGeometry`, les objets sont de plus liés à la géométrie de type `Point` qui en fournit les attributs spatiaux. La seule trajectoire du `MovingObject` nommée `Semantic Trajectory` fournit l'ensemble des mises

Modèles émantique de trajectoire d'objet mobile

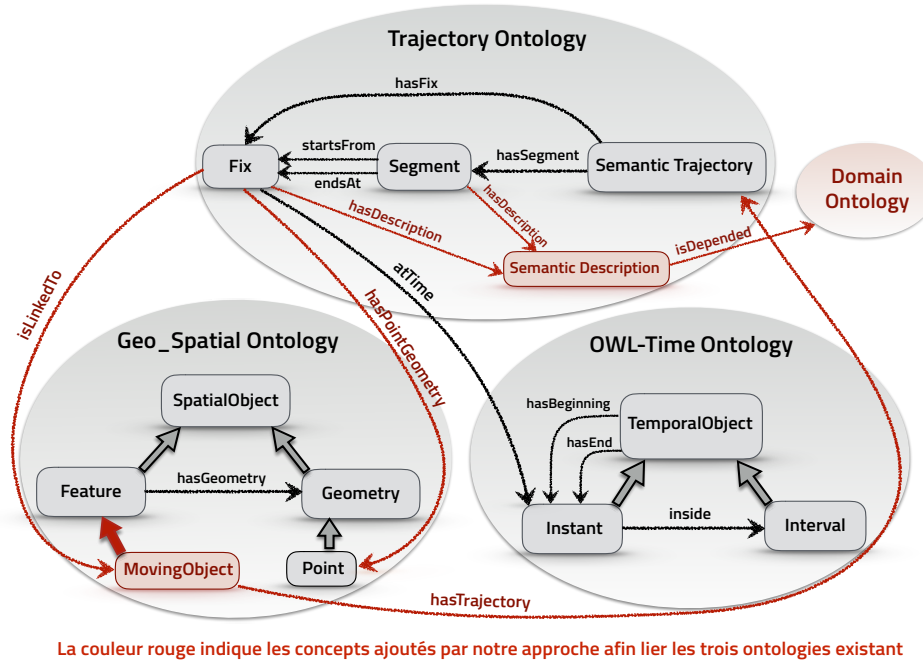


FIGURE 5. Modèle spatio-temporel sémantique

à jour des propriétés de l'objet mobile et les informations liées à ces dernières tout au long de la vie de l'objet. Le lien **hasTrajectory** associe un objet mobile à sa trajectoire. Chaque mise à jour nommée **Fix** est valide à un instant donné défini par le lien **atTime** et si il est effectué sur la position (représentation spatiale) de l'entité mobile donc **Fix** sera lié à une position introduite par le lien **hasPointGeometry**. Par analogie à l'ontologie de trajectoire (Hu *et al.*, 2013), nous décomposons la trajectoire en phases ou segments nommés **Segment** à l'aide du lien **hasSegment**. Ces phases (correspondant à l'acquisition des mises à jour) sont délimitées par deux **Fix** successifs définis par les relations **startsFrom** et **endsAt** afin de pouvoir analyser le comportement d'une entité durant cet intervalle. Par conséquent, la représentation complète d'une entité est le résultat de l'agrégation de tous les **Fix** avec le lien **isLinkedTo** au cours du temps. Ainsi, nous pouvons obtenir les traces de déplacements d'un navire (une partie de sa trajectoire) dans une durée de temps à l'aide du lien **hasFix** et l'ensemble des **Fix** de ce navire.

Grâce aux liens **isLinkedTo** et **hasPointGeometry**, le modèle tire avantage des entités spatiales du module ontologique Geo-Spatial et donc peut disposer des relations spatiales liées au concept **SpatialObject**. Ainsi, avec le lien **atTime**,

SAGEO'2016 – Nice, 6-9 décembre 2016

il possède des avantages des entités temporelles du module OWL-Time qui explicite les relations temporelles définies entre les durées du temps.

Cette approche a pour but premier d'être applicable à divers ensembles de données liées au mouvement d'objet et, par conséquence, faciliter l'analyse et le traitement à large échelle. Deuxièmement, elle doit être facilement extensible, en s'alignant sur des ontologies cadres ou des patrons de conceptions, ainsi que sur d'autres ontologies de domaine enrichissant la trajectoire d'objet mobile. Ce travail permet de relier sémantiquement les informations provenant des diverses sources avec les liens modélisés par le modèle et d'aboutir à une véritable trajectoire sémantique. C'est cette trajectoire qui sera ensuite utilisée lors des phases d'analyse et d'étude des comportements remarquables d'objets.

3.3. *Exploitation du modèle*

3.3.1. *Représentation de données avec notre approche*

La figure 6 est une représentation RDF qui résulte de l'extraction de données (valeurs brutes de propriétés) recueillies toutes les 10 secondes à partir d'un objet mobile désigné par "mObj-231". Cette collection de données est conforme au modèle intégré proposé dans notre contribution. À l'exception de son identifiant, un objet peut voir la valeur de ses propriétés varier au cours du temps. À cet effet, nous créons à la volée des instances de la classe "Fix", qui sont liées à l'objet mobile d'origine à l'aide de liens de type "isLinkedTo" et qui sont par ailleurs caractérisées par les valeurs des propriétés mises à jour à chaque intervalle du temps.

3.3.2. *Interrogation des données modélisées*

Le modèle proposé permet à un utilisateur de poser une requête qui va tirer parti de tous les modèles sous-jacents et ainsi permettre d'interroger les données intégrées afin de détecter des motifs d'intérêt. Imaginons qu'un utilisateur cherche à savoir à quel moment l'objet "mObj-120" a pu rencontrer de très près un autre objet mobile. Cette requête est traduite en SPARQL dans la figure 7. Les positions exprimées au format WKT des `Fix`s de l'objet de l'étude sont tout d'abord recherchées. Ces positions sont ensuite comparées avec toutes les positions des autres objets présents dans le triplestore. Les objets situés dans un rayon de moins de 0.001 km sont les seuls sélectionnés à l'aide d'un filtre exploitant la fonction `nearby()` définie dans la librairie `Jena Spatial`. Enfin, une sélection est également opérée sur le temps grâce à une nouvelle condition de filtrage. Nous pourrions aussi interroger les sémantiques liées à ces positions comme par exemple la vitesse de l'objet, sa direction, ainsi que sa trajectoire à chacune de ses positions. Le résultat de la requête est présenté dans la figure 8 et indique que l'objet "mObj-120" a rencontré l'objet "mObj-230" à l'instant "2009-11-26T16:00:18Z". La localisation géographique exacte de chaque objet est également retournée.

SAGEO'2016 – Nice, 6-9 décembre 2016

```

PREFIX sttraj: <http://www.intactile.com/ontologies/2016/TrajOntology#>
PREFIX geosparql: <http://www.opengis.net/ont/geosparql#>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX spatial: <http://jena.apache.org/spatial#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?time ?movingObj ?position ?movingObj2 ?position2 WHERE {

    ?fix rdf:type stTraj:Fix ;
    ?fix sttraj:isLinkedTo ?movingObj.
    ?movingObj rdf:type sttraj:MovingObject ;
        sttraj:object_id 120 .

    ?fix2 rdf:type sttraj:Fix ;
    ?fix2 sttraj:isLinkedTo ?movingObj2 .
    ?movingObj2 rdf:type stTraj:MovingObject .

    FILTER ( ?movingObj != ?movingObj2 ) .

    ?fix sttraj:atTime ?instant ;
        sttraj:hasPointGeometry ?geoPoint .
    ?geoPoint geosparql:asWKT ?position .
    ?instant time:inXSDDateTime ?datetime .

    ?fix2 sttraj:atTime ?instant2 ;
        sttraj:hasPointGeometry ?geoPoint2 .
    ?geoPoint2 geosparql:asWKT ?position2 .
    ?instant2 time:inXSDDateTime ?datetime2 .

    ?position2 spatial:nearby(position 0.001) .

    FILTER ( ?datetime = ?datetime2 )
}

```

FIGURE 7. Requête exemple

time	movingObj	position	movingObj2	position2
2009-11-26T16:00:18Z	mObj_120	30.800934509220667, 28.93014485584921	mObj_230	30.801829078219473, 28.893068839566347

FIGURE 8. Résultats de la requête exemple

La première validation opérationnelle du modèle se fait via l'insertion et l'interrogation d'un jeu de données spatio-temporelles lié aux mouvements d'objets mobiles au format RDF. Les données sont stockées dans un triplestore Jena-TDB⁵ et elles sont interrogées grâce à un jeu de requêtes SPARQL (langage de requêtage Jena-TDB).

5. <https://jena.apache.org/documentation/tdb/>

Modèles émantique de trajectoire d'objet mobile

Les prochaines étapes vont consister à exploiter ces données par des calculs et raisonnements spatio-temporels afin d'en extraire les informations relatives aux comportements des objets si possible sous forme de patrons de comportements remarquables ou inhabituels. D'un point de vue académique, nous songeons à évaluer l'apport de cette approche usage des modèles ontologiques et d'analyse spatio-temporelle par inférence face aux systèmes d'information géographique qui intègrent bon nombre des fonctions de calcul spatial.

Bibliographie

- Baglioni M., Macedo J., Renso C., Wachowicz M. (2008). An ontology-based approach for the semantic modelling and reasoning on trajectories. In *Advances in conceptual modeling—challenges and opportunities*, p. 344–353. Springer.
- Clementini E., Laurini R. (2008). Un cadre conceptuel pour modéliser les relations spatiales. *Revue des Nouvelles Technologies de l'É*, vol. 8217, p. 1–17.
- Etienne L. (2011). *Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels, application au trafic maritime*. Thèse de doctorat non publiée, Université de Bretagne occidentale-Brest.
- Gayte O., Libourel T., Cheylan J., Lardon S. (1995). Modélisation objet des entités spatio-temporelles. In *Actes conférence cassini*, p. 1–17.
- Herring J. (2011). Opendis implementation standard for geographic information-simple feature access-part 1: Common architecture. *OGC Document*, vol. 4, n° 21, p. 122–127.
- Hu Y., Janowicz K., Carral D., Scheider S., Kuhn W., Berg-Cross G. *et al.* (2013). A geo-ontology design pattern for semantic trajectories. In *Spatial information theory*, p. 438–456. Springer.
- Janowicz K., Narock T., Raymond L., Shepherd A. (2015). Ontology pattern modeling for cross-repository data integration in the ocean sciences: The oceanographic cruise example. *The Semantic Web in Earth and Space Science. Current Status and Future Directions*, vol. 20, p. 185.
- Lardon S., Libourel T., Cheylan J. (1999). Concevoir la dynamique des entités spatio-temporelles. *Représentation de l'espace et du temps dans les SIG, Revue internationale de géomatique, No9*, 45-65.
- Lupp M. (2008). Open geospatial consortium. In *Encyclopedia of gis*. Springer.
- Narock T., Fox P. (2015). The geovocamp workshop experience and ontology design pattern development. *The Semantic Web in Earth and Space Science. Current Status and Future Directions*, vol. 20, p. 171.
- Nathan R., Getz W. M., Revilla E., Holyoak M., Kadmon R., Saltz D. *et al.* (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, vol. 105, n° 49, p. 19052–19059.
- Noël D., Villanova-Oliver M., Gensel J., Le Quéau P. (2015). Modélisation de trajectoires sémantiques intégrant perspectives multiples et facteurs explicatifs. In *Sageo spatial analysis and geomatics*.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Sistla A. P., Wolfson O., Chamberlain S., Dao S. (1997). Modeling and querying moving objects. In *icde*, p. 422.
- Spaccapietra S., Parent C., Damiani M. L., Macedo J. A. de, Porto F., Vangenot C. (2008). A conceptual view on trajectories. *Data & knowledge engineering*, vol. 65, n° 1, p. 126–146.
- Vandecasteele A. (2012). *Modélisation ontologique des connaissances expertes pour l'analyse de comportements à risque: application à la surveillance maritime*. Thèse de doctorat non publiée, Ecole Nationale Supérieure des Mines de Paris.
- Yan Z. (2011). *Semantic trajectories: computing and understanding mobility data*. Thèse de doctorat non publiée, EPFL.

Interrogation de données spatio-temporelles de géolocalisation indoor à partir de trajectoires hybrides

Franck Jeveme Panta, Florence Sèdes

IRIT, Université Paul Sabatier Toulouse 3
franck.panta@irit.fr, florence.sedes@irit.fr

RESUME. Le GPS (Global Positioning System) basé sur la visibilité directe entre le satellite et le récepteur, s'est imposé pour la localisation outdoor. En l'absence de signal GPS, la localisation d'une cible dans les environnements indoor peut se faire avec un réseau de capteurs. Ces réseaux déployés dans les bâtiments sont de plus en plus nombreux et servent à plusieurs applications basées sur la localisation : surveillance, détection, navigation, etc. Ces capteurs de localisation indoor génèrent une quantité importante d'informations de tracking. Le contexte d'application de ce travail est celui des systèmes de localisation indoor basés sur les cellules Wifi, sur les capteurs ICCARD et sur les caméras de vidéosurveillance. Dans un tel contexte, il n'existe aucun système global de référence similaire au GPS : les informations de localisation sont hétérogènes (positions géométriques et symboliques / multitude de systèmes de référence). Cet article présente un ensemble de démarches qui ont permis de concevoir et implémenter un framework utilisant les informations générées par les réseaux de capteurs de localisation déployés dans un environnement indoor, illustré dans le cadre du forensic.

ABSTRACT. GPS has become the reference for Outdoor positioning, where a direct uplink between the GPS satellite and a receiver device can be established. In the absence of GPS signal for Indoor positioning, the location of a target device requires the use of wireless sensors network. Nowadays sensor networks deployed in buildings are becoming more and more used for many applications based on the location: surveillance, detection, navigation, etc. These indoor locating sensors generate an important number of tracking information. The application context of this paper is indoor locations systems based on wireless cell, ICCARD sensors and video surveillance. In this context, there is no global reference system similar to GPS, the location information is heterogeneous (geometric and symbolic positions / various references systems). This paper presents sequences of steps which help to design and implement a framework using the information generated by the location sensor networks deployed in an indoor environment, , illustrated in the forensic.

MOTS-CLES : modèle de localisation indoor, dataset, requêtes spatio-temporelles, requête hybride.

KEYWORDS: indoor location model, dataset, spatio-temporal queries, hybrid query.

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

La mise à disposition de l'utilisateur, via des requêtes spatio-temporelles, de données géo-localisées, basées sur la captation via des réseaux de capteurs, vise à localiser, identifier, extraire les métadonnées des capteurs, dispositifs ou objets connectés situés dans une région donnée, à une certaine distance d'autres objets ou capteurs, à l'intersection de trajectoires, etc. de telles applications sont basées sur la position de l'objet, qui peut être exprimée de deux façons, selon différents systèmes de référence : géométrique (coordonnées par rapport à un système de référence comme le système géodésique ou le plan d'un bâtiment) et symbolique (description plus sémantique par rapport à des points d'intérêt, adresse postale, réseaux de transport, pièces d'un bâtiment, etc.). Au moment du traitement de ces positions, traitement qui consiste à exprimer les prédicats par rapport à une référence commune, la principale différence entre les deux types d'expression des positions est la précision, une position géométrique pouvant être exprimée par un point tandis qu'une position symbolique est le plus souvent représentée par une géométrie (d'une pièce, d'un bâtiment ou d'une cellule RFID).

Dans les environnements outdoor, les applications utilisent dans la plupart des cas des modèles de localisation basés sur des capteurs GPS qui sont largement présents dans les systèmes embarqués (voitures, smartphones, etc.). Les environnements indoor sont de plus en plus complexes et il n'existe pas de système de référence unique auquel toute mesure peut se rapporter : les bâtiments des entreprises, infrastructures des métros, des aéroports, etc., génèrent un besoin important de gérer finement la localisation indoor en termes d'applications, sans qu'il n'existe encore de réponse satisfaisante et générique.

Les travaux présentés dans ce papier s'appuient sur les recherches menées dans notre équipe sur les problèmes de localisation outdoor, dans le cadre du forensic (Sedes et al. 2012, Codreanu 2015) et proposent une extension dans les environnements indoor avec des capteurs (appropriés) tels que bornes WiFi, lecteurs ICCARD et caméras de vidéosurveillance. Le problème principal dans le contexte indoor est l'hétérogénéité de l'information de localisation donnée par les différents capteurs. Cette hétérogénéité concerne les cas suivants :

- les positions sont soit géométriques soit symboliques: tandis que la localisation basée sur le wifi donne des positions géométriques par rapport au système de coordonnées à 2 dimensions, la localisation basée sur les cellules dans le cas du réseau cellulaire donne des positions symboliques, ce qui est aussi le cas avec les technologies telles que l'infrarouge (IR) et le RFID (Radio Frequency Identification) ;
- les positions sont exprimées par rapport à différents systèmes de coordonnées ou systèmes de référence.

Interrogation de données spatio-temporelles

Notre travail porte principalement sur ces deux points et consiste à développer une solution générique de modélisation et un framework de requête de trajectoire d'objets en environnement indoor.

La suite de ce papier est structurée comme suit : après un examen des travaux connexes concernant les modèles de localisation en indoor, les trajectoires et les types de requêtes spatiales dans la section 2, la section 3 présente la définition d'un modèle de requête trajectoire hybride en environnement indoor. Les différents algorithmes de recherche sont présentés à la section 4. A la section 5, l'architecture proposée pour le framework est présentée et enfin, la section 6 conclut et discute les recherches futures.

2. Etat de l'art

2.1. Les modèles de localisation en indoor

La localisation indoor vise à définir l'emplacement d'un utilisateur/objet dans un environnement intérieur. Un environnement indoor est en général constitué par un espace en trois dimensions (2D + étage par exemple) : ainsi, un édifice dans lequel vivent, travaillent, transitent, etc. des personnes passent habituellement leur temps (maisons, centres commerciaux, entreprises, gares, etc.) (Li et al., 2008). Pour représenter cet espace, les modèles spatiaux reposent sur la définition d'un modèle de topologie qui traduit l'espace, la définition de l'emplacement des objets et les trajectoires. Cette représentation est appelée modèle de localisation. (Leonhardt et U., 1998) définissent un modèle de localisation comme une représentation expressive, souple et efficace des informations de localisation.

Il existe deux classes de base de modèles de localisation :

- les modèles de localisation géométriques, dans lesquels les positions sont représentées par des points (ex : les points $P_1(x_1, y_1)$ et $P_2(x_2, y_2)$) (Parent et al., 2013) ;
- les modèles de localisations symboliques : dans ce cas les positions sont désignées par des noms ou symboles abstraits (ex: "Salle C1", "Salle C3") (Jensen et al., 2009).

Les modèles hybrides sont des combinaisons de modèles géométriques et symboliques (C. Becker et F. Darr, 2005). Nous nous intéressons aussi à la variation des positions dans le temps, C'est dans ce même ordre idée que nous présentons dans ce qui suit une revue de la littérature sur les trajectoires.

2.2. Définition des trajectoires

Une trajectoire est définie comme le changement de la position d'un objet mobile dans le temps. Une trajectoire peut être segmentée en fonction de plusieurs critères. Le segment d'une trajectoire est défini comme une sous-séquence maximale

SAGEO'2016 – Nice, 6-9 décembre 2016

de la trajectoire de telle façon que toutes les positions spatio-temporelles soient conformes à un prédicat donné (ex : direction constante) (Mountain et Raper, 2001).

(D. Codreanu, 2015) donne une définition formelle et générique d'une trajectoire et d'un segment de trajectoire. Une trajectoire (voir l'équation 1) est définie comme un tuple composé d'un identifiant de trajectoire, un identifiant d'objet, un ensemble de segments de trajectoires et un intervalle de temps global qui marque le "start" et le "end" de la trajectoire.

$$Tr = (tr_{id}, obj_{id}, \{u_k\}, [t_{start}, t_{end}]) \quad (1)$$

$$u_k = (u_{id}, \{position_i\} / 1 \leq i, [t_{kstart}, t_{kend}]) \quad (2)$$

Un segment (voir l'équation 2) de trajectoire est un tuple qui contient un identifiant de segment, un ensemble de positions homogènes et un intervalle de temps qui est optionnel.

Le mouvement des objets peut être effectué de façon libre (ex : le déplacement de l'aire) (Parent et al., 2006) ou contraint par exemple par l'architecture d'un bâtiment (Jensen et al., 2009) (ex : le mouvement d'une personne à l'intérieur d'une station de métro est contraint par la disposition des différents espaces et leurs connexions).

Dans ce qui suit, nous présentons les types de requêtes spatiales envisageables sur une trajectoire en environnement indoor.

2.3. Requêtes spatiales

Une requête spatiale est toute interrogation de l'utilisateur comportant au moins un élément de géolocalisation ou de situation spatiale (exprimée en termes de coordonnées géométriques ou de noms symboliques) et/ou une référence temporelle (date, heure, intervalle) (Ilarri et al., 2010). L'utilité de requêtes spatiales concernant la trajectoire passée des objets et les requêtes prédictives a été prouvée par une grande quantité de recherches et travaux commerciaux. Il existe plusieurs travaux concernant les requêtes spatiales comme : les requêtes de position (Becker et Durr, 2005) : retournant la position d'un objet mobile ou statique (position géométrique ou symbolique), les requêtes de région (Gu et al., 2009 ; Ay et al., 2008 ; Scuturici et Ejigu, 2006 ; Yuan et Schneider, 2010) : «quelles sont les pharmacies situées dans cette zone ?», les requêtes des voisins les plus proches (Zhang et al., 2003) : «quelles sont les maisons les plus proches du carrefour ?», les requêtes de navigation ou de trajectoire (Afyouni et al., 2012b) : « quel est le plus court chemin pour arriver à la maison ? ».

La plupart des ouvrages traitant les trajectoires d'objets visent à utiliser les anciennes trajectoires pour déceler les déplacements habituels et donc à prédire à l'avenir la position des objets. Moins de travaux considèrent la trajectoire orientée vers les requêtes de reconstitution et à notre connaissance, il n'y a pas de framework

Interrogation de données spatio-temporelles

de requête traitant les trajectoires hétérogènes (contenant à la fois des segments symboliques et géométriques) en environnement indoor. Dans ce qui suit, nous allons présenter un cas d'utilisation qui montre l'utilité d'une telle requête.

Supposons un bâtiment "intelligent" dans lequel chaque employé dispose d'un smartphone suivi par des systèmes de positionnement basés sur différents capteurs de localisation. D'autres types de capteurs surveillent la consommation d'énergie et la température partout dans le bâtiment.

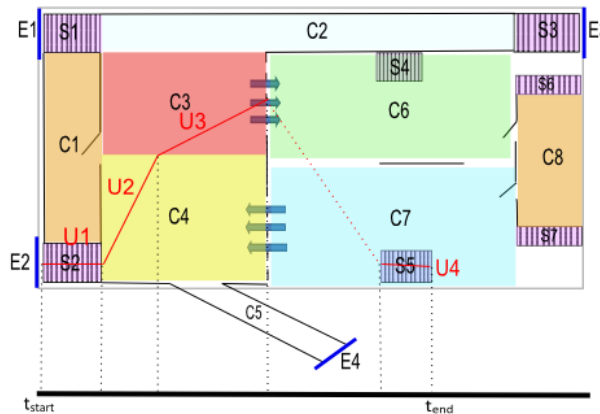


FIGURE 1. Exemple de trajectoire 1

La Figure 1 illustre la trajectoire d'une personne dans le bâtiment. La trajectoire est composée de quatre segments (u1, u2, u3 et u4). Il pourrait y avoir un manque d'information de position pour une certaine période de temps (une personne a fermé son smartphone et aucune information n'existe sur sa position). La ligne pointillée représente une lacune dans la définition de la trajectoire (appelé «trou») qui peut être "remplie" à l'aide des fonctions d'interpolation (Vazirgiannis et Wolfson, 2001). Sur la base de cette trajectoire, on veut trouver les différents opérateurs comme la moyenne de la consommation d'énergie, l'identification des capteurs situés sur la trajectoire ou à une certaine distance de la trajectoire.

Considérons un autre cas, tel un incident dans une zone de l'édifice. Une personne se déplace à l'intérieur d'une zone interdite. La génération de ses positions est basée sur la détection de son smartphone par les cellules WiFi environnantes. Ceci génère des coordonnées (x, y) au regard du plan du bâtiment. À un certain point, le signal est perdu et la seule information que nous avons est qu'il est entré dans une salle sécurisée (son badge a été validé à l'entrée de la salle). Dans ce cas, la trajectoire de la personne ressemble à celle de la Figure 2. Les deux premiers segments de la trajectoire sont u1 et u2, il y a une période où l'objet a été perdu et la dernière partie de la trajectoire est u3 qui est la chambre C1. Dans ce cas, nous

SAGEO'2016 – Nice, 6-9 décembre 2016

voulons localiser et identifier les caméras vidéo situées sur le parcours donné afin de visualiser les enregistrements qu'elles ont pu réaliser.

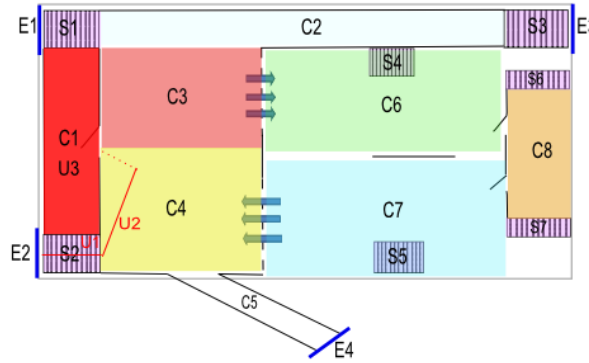


FIGURE 2. Exemple de trajectoire 2

Nous constatons dans ces deux cas que nous avons à faire à des trajectoires hétérogènes c'est-à-dire des trajectoires constituées de multiples positions (géométriques et symboliques) ayant des systèmes de coordonnées différents.

3. Définition d'un modèle de requête trajectoire hybride en environnement indoor

Selon les étapes de notre stratégie, nous allons commencer par définir un modèle de requête que le système va pouvoir supporter en se basant sur les deux exemples de cas d'utilisation présentés dans la section 2.3. L'innovation que nous introduisons dans la définition d'une trajectoire est la prise en compte du fait que chaque position est exprimée par rapport à un système de référence (ex : plan du bâtiment, réseau de lecteur ICCARD). Les positions sont homogènes à l'intérieur du segment (exprimées par rapport à un seul système de référence) mais les différents segments peuvent contenir des positions hétérogènes comme dans les exemples présentés à la section 2.3.

On définit une requête trajectoire hybride comme un tuple illustré ci-dessous :

$$T_r = (tr_{id}, \{u_k\}, [t_{start}, t_{end}]) \quad (3)$$

où tr_{id} est l'identifiant de la trajectoire, $\{u_k\}$, est l'ensemble des segments spatiaux constituant la trajectoire et $[t_{start}, t_{end}]$ représente l'intervalle de temps de la requête trajectoire.

La définition d'un segment de trajectoire est la suivante :

Interrogation de données spatio-temporelles

$$u_k = (u_{id}, refS_{id}, \{position_i\}) \quad (4)$$

où u_{id} est l'identifiant du segment, $refS_{id}$ est l'identifiant du système de référence par rapport auquel les positions sont définies et $position_i$ est l'ensemble des positions qui composent le segment de trajectoire.

Pour pouvoir gérer l'hétérogénéité des segments de trajectoire nous définissons la position d'un objet comme étant composée de deux grandes parties : une partie désignant le système de référence et une autre partie qui désigne la localisation par rapport au système de référence. L'idée va dans le même sens que celle proposée par (Jiang et Steenkiste, 2002) qui définissent la syntaxe d'une localisation indoor hybride basée sur la syntaxe d'un URI (Universal Resource Identifier).

Une requête trajectoire hybride est composée d'une partie spatiale et d'une partie temporelle. La partie spatiale est constituée d'une séquence de segments, chaque segment étant constitué d'une partie désignant le système de référence et une séquence de positions (géométriques ou symboliques) exprimées par rapport au système de référence correspondant. La Figure 3 présente un exemple de trajectoire hybride avec les différents types de position, de systèmes de référence et l'intervalle de temps de la requête.

```

{
  "query": {
    "spatial": [
      {
        "Floor8": [
          [128.53813, 368.34243],
          [190.89598, 357.04869],
          [140.32508, 345.56478],
          [136.62363, 322.94129],
          [153.10012, 305.00549],
          [172.11112, 288.10549]
        ]
      },
      {
        "ICCARD": ["0000110015010003"]
      }
    ],
    "temporal": {
      "start": "14-02-14 04:05:34",
      "end": "14-02-2014 05:50:10"
    }
  }
}

```

FIGURE 3. Exemple de requête trajectoire hybride

C'est cette requête qui donne la trajectoire suivie par le suspect. Les informations spatio-temporelles générées par les différents capteurs et sauvegardées seront interrogées par cette trajectoire pour déterminer les objets (personnes, dispositifs,

SAGEO'2016 – Nice, 6-9 décembre 2016

etc.) se trouvant sur celle-ci, ainsi que les caméras de vidéosurveillance susceptibles d'avoir filmé cette trajectoire pendant l'intervalle de temps de la requête.

Dans la section suivante, nous présentons les différents algorithmes de recherche des positions et des caméras.

4. Algorithmes de recherche

Le point d'entrée de notre système est une requête trajectoire hybride, constituée des positions géométriques et symboliques exprimées par rapport à différents systèmes de référence et d'un intervalle de temps. Une fois que cette requête est interprétée par un module interpréteur de requête, elle est exprimée comme un ensemble de points (p_1, p_2, \dots, p_n) et un intervalle de temps $[t_1, t_2]$. Ces points forment une séquence de segments spatiaux (u_1, u_2, \dots, u_n) pour ce qui concerne les positions géométriques et des cercles de centre (c_1, c_2, \dots, c_n) pour les différentes positions symboliques. Cette séquence de points avec l'information temporelle constitue l'entrée des algorithmes de recherche. Le but est :

- de proposer un ensemble d'objets (personne, dispositif, etc.) dont les positions ont «intersecté» la trajectoire donnée par la requête dans l'intervalle de temps $[t_1, t_2]$;
- de proposer un ensemble d'objets se trouvant dans la zone de couverture (portée) de chaque lecteur ICCARD de la requête trajectoire dans l'intervalle de temps $[t_1, t_2]$;
- de proposer une liste des caméras dont le champ de vue (qui peut être variable) a «intersecté» la trajectoire de la requête dans l'intervalle de temps.

4.1. Recherche des positions des objets liées aux bornes WiFi

A partir d'une trajectoire (séquence de segments u_k) donnée comme requête, il s'agit de retrouver les objets qui se trouvent « sur » cette trajectoire. Nous allons prendre en compte une incertitude car les points donnés dans la requête trajectoire peuvent ne pas être exacts (Alboody et al. 2009). Pour cela nous allons tracer la trajectoire (Poly-ligne) avec les points donnés dans la requête, puis effectuer une translation d'un côté comme de l'autre de cette trajectoire dans le but de construire une géométrie. Par la suite nous allons « intersecter » cette géométrie par l'ensemble de données (positions) WiFi enregistrées, qui va identifier l'ensemble des objets qui se trouvent sur la trajectoire. Rappelons qu'il faut «intersecter» la géométrie avec des données WiFi (positions des objets) dont le temps (DATETIME) est compris dans l'intervalle $[t_1, t_2]$ donné dans la requête.

Opérateurs :

Le résultat est l'ensemble des doublets: $R=\{r=(obj, pos)\}$; la fonction *construire_geometry*(u_k) permet de construire une géométrie à partir de la poly-ligne formée par les segments u_k ; la fonction *trouverObjet*($[t_1, t_2]$) renvoie la liste

Interrogation de données spatio-temporelles

des objets dont le temps (DATETIME) est compris dans l'intervalle $[t_1, t_2]$ ainsi que les positions correspondantes.

Algorithme :

Algorithme 1 : Algorithme de recherche des objets dont les positions ont intersecté la trajectoire de la requête

Entrées : une suite de segments de trajectoire : u_k et un intervalle de temps : $[t_1, t_2]$.
Sorties : la liste des objets qui ont intersecté la trajectoire et les positions correspondantes.

```

1:  $geometry \leftarrow construire\_geometry(u_k)$ 
2:  $listeObjet(obj_i, pos_i) \leftarrow trouverObjet([t_1, t_2])$ 
3: pour chaque  $obj_i$  de  $listeObjet(obj_i, pos_i)$  faire
4:   si  $pos_i$  intersecte  $geometry$  alors
5:      $ajouterResultat(obj_i, pos_i)$ 
6:   fin condition
7: fin boucle

```

4.2. Recherche des positions des objets liées aux lecteurs ICCARD

Dans ce cas, nous avons à faire à des positions symboliques qui sont des surfaces précises dans le bâtiment. Chaque lecteur ICCARD a un champ d'action ou bien une portée. Une fois la requête trajectoire passée dans le module interpréteur de requête, les noms des lecteurs sont représentés sur le plan du bâtiment par des points (entrées de l'algorithme) qui sont les positions affectées aux lecteurs ; à partir de ces points, sont construites des géométries. Ces géométries sont des cercles dont le diamètre est la portée de chaque lecteur. Le rayon de ces cercles est paramétrable. Les objets recherchés se trouvent donc dans la surface de ces cercles. On recherche donc ici les objets contenus dans les géométries construites.

Opérateurs :

Le résultat pour les lecteurs ICCARD est l'ensemble des doublets : $R=\{r(obj_i, pos_i)\}$; la fonction *construire_cercle* ($p_n, rayon_i$) permet de construire des cercles dont les centres sont les p_n et les rayons sont les $rayon_i$; la fonction *trouverObjet* ($[t_1, t_2]$) renvoie la liste des objets se trouvant dans la base de données dont le temps (DATETIME) est compris dans l'intervalle $[t_1, t_2]$ ainsi que les positions correspondantes.

Algorithme :

Algorithme 2 : Algorithme de recherche des objets contenus dans les géométries

Entrées : une suite de points : p_n et un intervalle de temps : $[t_1, t_2]$.
Sorties : la liste des objets contenus dans les géométries et les positions correspondantes.

```

1:  $listeObjet(obj_i, pos_i) \leftarrow trouverObjet([t_1, t_2])$ 
2: pour chaque  $p_n$  de la requête faire
3:    $geometry_i \leftarrow construire\_cercle(p_n, rayon_i)$ 
4:   pour chaque  $obj_i$  de  $listeObjet(obj_i, pos_i)$  faire
5:     si  $obj_i$  contenu dans  $geometry_i$  alors
6:        $ajouterResultat(obj_i, pos_i)$ 

```

SAGEO'2016 – Nice, 6-9 décembre 2016

7: **fin condition**
 8: **fin boucle**
 9: **fin boucle**

4.3. Algorithme de sélection des caméras

(D. Codreanu, 2015) propose un algorithme de sélection des caméras fixes. Notre travail étant une extension en environnement indoor du sien, nous allons adapter cet algorithme dans ce cas.

Le but est de proposer à l'opérateur une liste de caméras susceptibles d'avoir filmé les segments de la trajectoire donnée comme requête. Pour chaque caméra on doit donner l'intervalle de temps pendant lequel elle a filmé la trajectoire.

L'algorithme se déroule en deux étapes, l'étape de filtrage et l'étape d'affinage des résultats :

- filtrage : on sélectionne les caméras qui se trouvent à une distance inférieure à la distance de visibilité maximale des segments de la trajectoire ;
- affinage : pour les caméras sélectionnées dans l'étape de filtrage, on calcule la géométrie des champs de vue, et on choisit celles dont les géométries touchent les segments de la trajectoire.

Le résultat est l'ensemble des triplets $R=\{r=(c_i, u_k, [t_a, t_b])\}$ où c_i est la caméra sélectionnée, u_k est le segment filmé et $[t_a, t_b]$ l'intervalle de temps pendant lequel la caméra a filmé le segment de la trajectoire.

Ces algorithmes sont implémentés dans le module correspondant au moteur de recherche de l'architecture du framework développé.

Dans la section suivante nous présentons cette architecture.

5. Architecture du framework et mise-en-oeuvre

Le framework développé implémente le modèle de requête trajectoire hybride et les algorithmes de recherche présentés dans les sections précédentes. L'implémentation du prototype a été réalisée en utilisant JDK 1.7 avec les drivers JDBC pour la connexion avec la base de données Oracle.

Les principaux modules constituant cette architecture sont :

- **Interface utilisateur** : constituée des différents blocs qui permettent l'interaction avec l'utilisateur, la visualisation des données et des résultats ;
- **Interpréteur de requête** : effectue la transformation de la requête trajectoire hybride soumise par l'utilisateur en requête spatio-temporelle (un ensemble de points 2D exprimés par rapport au plan du bâtiment formant des séquences de segments et/ou des géométries et un intervalle de temps) ;
- **Moteur de recherche** : implémente les algorithmes de recherche définis.

Interrogation de données spatio-temporelles

- **Stockage** : contient la base de données spatiale, la collection de données ainsi que des modules qui permettent l'interaction avec celles-ci ;
- **Générateur de requête SQL** : réalise la communication entre le moteur de recherche et le module de stockage ;
- **Collecte des données** : collecte les données générées par les différents capteurs de localisation et les organise dans un format donné.

Les relations entre les différents modules sont présentées à la Figure 4.

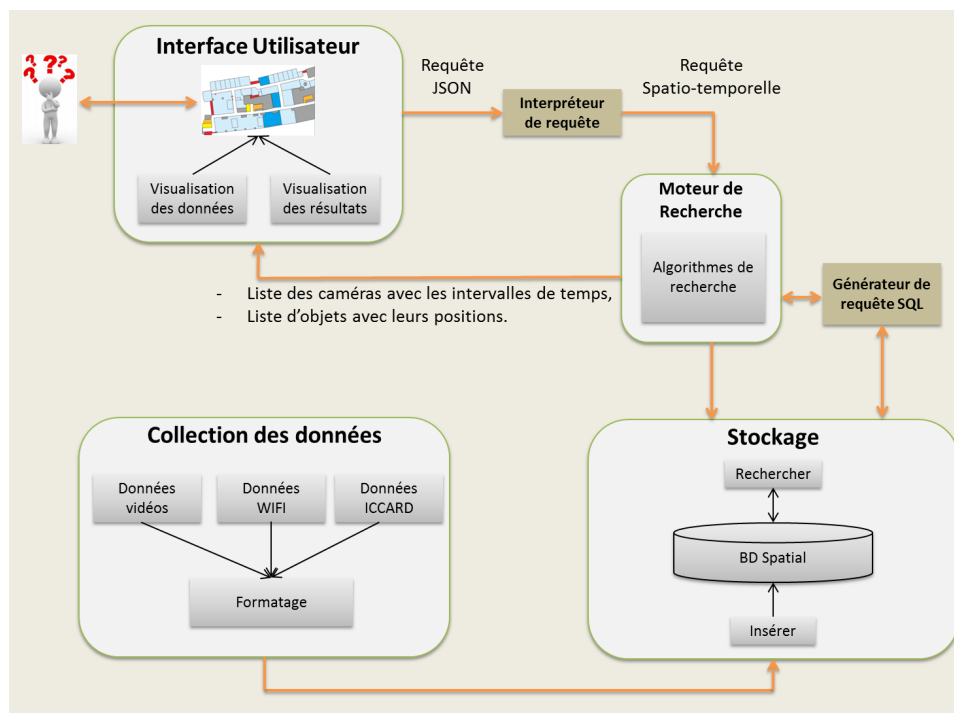


FIGURE 4. Architecture du framework

L'objectif est de mettre en œuvre la recherche d'objets via requête : recherche d'objets se trouvant sur une trajectoire donnée comme requête, d'objets se trouvant à la portée des lecteurs ICCARD, de caméras susceptibles d'avoir filmé la trajectoire pendant l'intervalle de temps donné par la requête.

SAGEO'2016 – Nice, 6-9 décembre 2016

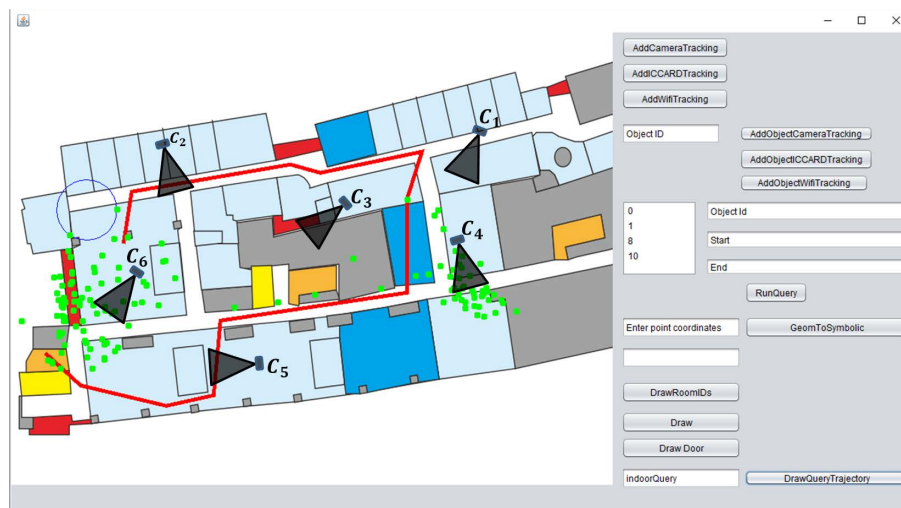


FIGURE 6. Application

L'interprétation de la requête de la Figure 3 est détaillée Figure 6 : la trajectoire (poly-ligne en rouge) représente le premier segment de la requête qui est constitué de l'ensemble des positions géométriques ([45;425], [92;468], [206;495], [269;481], [278;382], [526;345], [526;221], [546;158], [411;186], [371;175]) ; le cercle en bleu représente la portée du seul lecteur ICCARD ("0000110015010003") présent dans la requête. Les points en vert sont les positions des personnes générées par les bornes wifi et sauvegardées dans la base de données Oracle.

Les personnes qui nous intéressent sont celles qui se trouvent sur la trajectoire qui est la poly-ligne en rouge (en l'occurrence 6) et celles qui se trouvent à l'intérieur ou sur le cercle en bleu (en l'occurrence 2). Les caméras qui nous intéressent sont celles dont le champ de vue a «intersecté» la trajectoire (en l'occurrence C_2 et C_5).

L'implémentation des requêtes dans Oracle Spatial permet de visualiser ces résultats en retournant la liste des identifiants des personnes et leurs positions, la sélection des caméras qui ont «intersecté» les segments de la trajectoire ainsi que les intervalles de temps pendant lesquels ces champs de vue ont « intersecté » la trajectoire.

6. Conclusion

Nous avons présenté dans cet article notre contribution dans le domaine de la localisation indoor, comme une extension des précédents travaux sur la construction de trajectoires. Nous définissons un système de coordonnées en indoor qui peut être associé au plan du bâtiment, à un étage ou à chaque espace. Les positions sont

Interrogation de données spatio-temporelles

exprimées en utilisant les coordonnées 2D au regard du plan de l'étage, une troisième coordonnée est associée à chaque position dans le but d'indiquer le numéro de l'étage par exemple. Nous avons proposé un modèle de requête trajectoire hybride constitué des positions géométriques et symboliques exprimées par rapport à différents systèmes de référence. Les algorithmes développés permettent de rechercher les objets (personnes, dispositifs, etc.) se trouvant sur une trajectoire donnée comme requête sur un intervalle temporel donné, et de rechercher les caméras susceptibles d'avoir filmé cette trajectoire.

Notre travail peut s'instancier dans le contexte du forensic, par ex. lors d'une enquête (Codreanu et al., 2015). Nous envisageons d'étendre ce contexte afin d'être en mesure de traiter les requêtes en temps réel ou de prédire les trajectoires sur la base des statistiques réalisées sur des données stockées. Pour les systèmes de localisation indoor basés sur les caméras de vidéosurveillance, le résultat de l'opérateur proposé est une liste de caméras dont les intervalles de temps peuvent se recouvrir. Une perspective concerne donc l'ordonnement des résultats en fonction de la distance des caméras par rapport aux segments de la requête par exemple. Dans nos algorithmes, il est question de rechercher des relations topologiques entre les objets et les géométries. La question de la recherche de relations topologiques entre les régions de l'espace est d'actualité, surtout entre les régions de l'espace avec des incertitudes. Dans le traitement d'une requête, il faudra prendre en compte l'incertitude (fuzzy temporel et fuzzy spatial) (Sèdes et al., 2010).

Ces contributions ont permis de mettre sur pied les bases du développement d'un framework dont l'architecture a été présentée et qui peut être considéré comme un outil d'aide au requêtage hybride.

Bibliographie et références

- Afyouni, I., Ray, C., Ilarri, S., and Claramunt, C. (2012b). *Algorithms for continuous location-dependent and context-aware queries in indoor environments*. 20th Intl Conf on Advances in Geographic Information Systems, ACM SIGSPATIAL '12, pp 329-338.
- Ahed Alboody, Jordi Inglada, Florence Sèdes (2009) *Enriching the spatial reasoning system RCC8*. SIGSPATIAL Special 1(1): 14-20.
- Arslan Ay, S., Zimmermann, R., and Kim, S. H. (2008). *Viewable scene modeling for geospatial video search*. 16th ACM Intl Conf on Multimedia, MM '08, pp 309-318.
- Becker, C. and Durr, F. (2005). *On location models for ubiquitous computing*. Personal Ubiquitous Comput. 9(1) :20-31.
- C. Becker and F. Durr (2005). "On location models for ubiquitous computing," Personal and Ubiquitous Computing, vol. 9, pp. 20. [Online]. Available: <http://dx.doi.org/10.1007/s00779-004-0270-2>.
- D. Codreanu (2015). *Modélisation des métadonnées spatio-temporelles associées aux contenus vidéo et interrogation de ces métadonnées à partir des trajectoires hybrides : Application dans le contexte de la vidéosurveillance*. Thèse de doctorat, Université Paul Sabatier, mai 2015.

SAGEO'2016 – Nice, 6-9 décembre 2016

Dana Codreanu, André Péninou, Florence Sèdes, *Video Spatio-Temporal Filtering Based on Cameras and Target Objects Trajectories - Videosurveillance Forensic Framework*. ARES 2015: 611-617.

Gu, Y., Yu, G., Guo, N., and Chen, Y. (2009). *Probabilistic moving range query over rfid spatio-temporal data streams*. ACM CIKM '09, pp 1413-1416.

Ilarri, S., Mena, E., and Illarramendi, A. (2010). *Location dependent query processing : Where we are and where we are heading*. ACM Comput. Surv., 42(3) :12 :1–12 :73.

Jensen, C. S., Lu, H., and Yang, B. (2009). *Graph model based indoor tracking*. In Mobile Data Management: Systems, Services and Middleware, 2009. MDM '09. pp 122-131.

Jiang, C. and Steenkiste, P. (2002). *A hybrid location model with a computable location identifier for ubiquitous computing*. In Proceedings of the 4th International Conference on Ubiquitous Computing, UbiComp'02, pages 246–263, London, UK, UK. Springer-Verlag.

Leonhardt, U. (2008). *Supporting Location-Awareness in Open Distributed Systems*. PhD thesis, at Department of Computing, Imperial College of Science, Technology and Medicine, University of London (1998). pp 43-65.

Li, K.-J (2008). *Indoor space: A new notion of space*. In Proc. 8th International Symposium on Web and Wireless Geographical Information Systems (W2GIS) (2008), pp. 1–3. LoI: 10.1007/978-3-540-89903-71.

Mountain, D. and Raper, J. (2001). *Modelling human spatio-temporal behaviour: a challenge for location based services*. 6th International Conference on GeoComputation, pp. 2-9.

Parent, C., Spaccapietra, S., and Zim'anyi, E. (2006). *Conceptual Modeling for Traditional and Spatio-Temporal Applications : The MADS Approach*. Springer-Verlag.

Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., Theodoridis, Y., and Yan, Z. (2013). *Semantic trajectories modeling and analysis*. ACM Comput. Surv., 45(4):42:1-42:32.

Scuturici, V.-M. and Ejigu, D. (2006). *Positioning support in pervasive environments*. In 2006 ACS/IEEE International Conference on Pervasive Services, pp. 19–26.

Florence Sèdes, Ahed Alboody, Jordi Inglada (2010). *Modeling Topological Relations between Uncertain Spatial Regions in Geo-spatial Databases: Uncertain Intersection and Difference Topological Model*. DBKDA 2010, IEEE Computer Society, pp. 56-68.

Sèdes, F., Sulzer, J., Marraud, D., Mulat, C., and Cepas, B. (2012). *Intelligent Video Surveillance Systems, chapter A Posteriori Analysis for Investigative Purposes*. Wiley.

Vazirgiannis, M. and Wolfson, O. (2001). *A spatiotemporal model and language for moving objects on road networks*. In Proceedings of the 7th International Symposium on Advances in Spatial and Temporal Databases, SSTD'01, pages 20–35, London, UK, UK. Springer-Verlag.

Yuan, W. and Schneider, M. (2010). *Supporting continuous range queries in indoor space*. 11th Intl Conf on Mobile Data Management (MDM), pp 209–214. IEEE.

Zhang, J., Zhu, M., Papadias, D., Tao, Y., and Lee, D. L. (2003). *Location-based spatial queries*. 2003 ACM SIGMOD Intl Conf on Management of Data, SIGMOD'03, pp 443–454.

[ISBN : 2-910545-11-3]



Session 5

Risques, changement climatique

Un indice pour caractériser et représenter la dangerosité des canicules en zone urbaine

Pinson L¹, Pham H¹, Ruas A¹

1. Département COSYS, Laboratoire LISIS, Cité Descartes
14-20 bd Newton, 77447 Marne la Vallée Cedex2, France
laura.pinson@ifsttar.fr ; ha.pham@ifsttar.fr ; anne.ruas@ifsttar.fr

RESUME. Il est maintenant irréfutable que le climat de la terre se modifie à un rythme croissant dû, pour l'essentiel, à l'action anthropique. Le GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) estime que cette transformation s'accroîtra dans les décennies à venir. Parmi les évolutions prévues, les extrêmes, à l'exemple de la canicule, posent, la question des impacts sur la santé. En effet, les canicules de 2003, 2006, 2010 ou bien même de 2015 ont causé la mort de plusieurs centaines à plusieurs milliers de personnes. Si en France, les mesures et plans de prévention existent depuis 2003, on montrera qu'il est important de prendre en compte autant les températures extérieures qu'intérieures lors de la création de ces plans. On montrera aussi que les différences thermiques entre bâtiments permettent d'avoir une meilleure connaissance de ces phénomènes. L'article présentera un indice de dangerosité prenant en compte les spécificités thermiques des bâtiments et la durée du phénomène. Une première représentation graphique sera aussi proposée pour conclure sur les différents moyens mis en place afin d'améliorer la communication sur la canicule.

ABSTRACT. It is now undeniable that the earth's climate is changing in an increasing pace mainly due to anthropological actions. The IPCC (Intergovernmental Panel on Climate Change) considers that this shift will intensify in the upcoming decades. Amongst the expected consequences, the most extreme phenomena, such as heat wave, will emphasize the question of the impacts on health. Indeed, the heat waves of 2003, 2006, 2010 or 2015 caused many deaths (from several hundred to several thousand people). If measures and prevention plans exist in France since 2003, we shall show that it is important to equally take into account both indoor and outdoor temperatures during the creation of such plans. Furthermore, we shall demonstrate that the thermal differences between buildings allow to have a deeper knowledge of these phenomena. This article will present a dangerousness indexes creation by considering the thermal specificities of buildings but also the duration of the phenomenon. A sketch of graphical representation will also be proposed to conclude on the actions already implemented to improve the communication on heat wave.

MOTS-CLES : Canicule, indice de dangerosité, représentations graphiques

KEYWORDS: Heat wave, dangerousness index, graphial representations

SAGEO'2016 – Nice, 6-9 décembre 2016

Caractériser et représenter la dangerosité des canicules

1. Introduction

Aux différents types d'événements climatiques sont associés des écarts à la normale climatologique qui sont qualifiés d'extrêmes. Ces événements sont fortement étudiés ces dernières décennies, dus, notamment, à l'augmentation des pertes économiques et aux nombreuses victimes (Munich Re, 2006). Les canicules sont classées comme des épisodes extrêmes. Pour parler de canicule, en France, des seuils d'intensité et de durée ont été mis en place par les services météorologiques et sanitaires. Ces seuils de températures sont définis par département. A Paris, les seuils sont fixés à 21°C la nuit et 31°C en journée. Au niveau de la durée, il faut observer ces températures élevées et maintenues au moins 3 jours et 3 nuits consécutives. Les risques sanitaires de ces phénomènes extrêmes dépendent principalement de l'intensité de la chaleur et de l'humidité relative de l'air, de la durée de l'exposition et de la vulnérabilité des personnes.

La prise de conscience des impacts sanitaires sur ces épisodes a amené Météo-France et l'InVS (Institut de Veille Sanitaire) à concevoir un Plan National Canicule (PNC) basé sur des seuils biométéorologiques. Ces derniers ont été conçus en prenant en compte des seuils météorologiques, des analyses et statistiques sur la surmortalité. Ces canicules sont particulièrement étudiées en ville parce qu'elles sont souvent plus sévères en milieu urbain en raison du phénomène d'îlot de chaleur urbain (ICU) (Oke, 1987). Ainsi, les villes doivent se préparer à ce changement et élaborer des stratégies d'adaptation. De plus, d'après le GIEC (5^{ème} rapport), ce phénomène caniculaire pourrait concerner l'Europe de l'Ouest un été sur deux à partir de 2070.

Les travaux que nous présentons sont menés dans le cadre du projet RepExtrem financé par le programme RDT (Risque Décision Territoire) du MEEM (Ministère de l'Environnement, de l'Énergie et de la Mer), conjointement avec Météo-France et l'IFSTTAR. Ils visent à apporter une analyse et une représentation des phénomènes caniculaires fondées sur la différenciation entre les températures intérieures et extérieures dans les bâtiments. En effet, les premiers résultats du projet ont montré que les températures intérieures sont sensiblement différentes des températures extérieures (Pinson et al, 2015). Ainsi, il nous a semblé pertinent de les intégrer pour mieux comprendre la dangerosité face au risque de canicule, d'autant plus que les personnes particulièrement fragiles passent la plupart de leur temps à l'intérieur de leur appartement. Nous avons ainsi développé une méthodologie permettant de construire des cartes de dangerosité de la canicule, pendant son occurrence, pour différentes configurations urbaines denses. Pour cela nous proposons des méthodes pour estimer la température intérieure et extérieure, construire un indice de dangerosité puis proposer des représentations pertinentes du phénomène via un service web adhoc. Pour l'estimation de la température nous utilisons un modèle SURFEX-TEB et un jeu de mesures de températures provenant de capteurs Tynitag (section 2). Nous avons conçu un indice de dangerosité à l'échelle du quartier qui intègre la durée et l'intensité du phénomène ainsi que le temps de récupération et les effets de différents types d'appartements (section 3). Enfin nous avons conçu un

SAGEO'2016 – Nice, 6-9 décembre 2016

premier service web pour l'exploration de la dangerosité de la canicule de 2015 sur Paris (section 4).

2. Modèle et mesures pour estimer la température pendant la canicule

2.1. Le modèle SURFEX couplé au modèle TEB

La modélisation du climat urbain demande de coupler un modèle météorologique avec un modèle de surfaces. Le modèle météorologique (ici AROME : Application de la Recherche à l'Opérationnel à Méso-Echelle ; Bouttier, 2007) permet d'avoir les prévisions météorologiques des hautes couches. Le modèle de surfaces (ici SURFEX)¹ permet d'ajuster ces températures pour les couches plus basses en fonction d'échanges thermiques spécifiques entre différentes surfaces. SURFEX décrit la surface comme une mosaïque composée de quatre principaux types de couverts (mers/océans, lacs/rivières, sols naturels et la végétation ainsi que les villes). A chaque point de grille du domaine va correspondre une fraction de couverture spatiale pour chacun des types. TEB (Town Energy Balance), sera alors appliqué aux surfaces minéralisées pour simuler les échanges entre les surfaces bâties et l'atmosphère. Ainsi à partir du système de modélisation des surfaces continentales SURFEX, couplé avec le modèle thermique TEB (Masson, 2000) développé par le Centre National de Recherches Météorologiques (CNRM), il est possible de calculer les températures estimées à l'intérieur et à l'extérieur des bâtiments, sur des mailles de 250*250m² (même si les calculs se font généralement à des mailles supérieures). Pour nos expérimentations, le modèle est appliqué à juillet 2015 sur la ville de Paris. Il fournit une série de grandeurs (dont les températures intérieures et extérieures) toutes les demi-heures pour chaque maille.

2.2. Mesures de températures dans différents types d'appartement

Afin d'obtenir des données climatiques plus fines et différencier les configurations, un réseau de capteurs Tynitag a été mis en place à l'intérieur et à l'extérieur de plusieurs appartements. Une campagne a été faite en 2014 puis en 2015. Dans ce papier nous traitons des données issues de la campagne 2015, plus chaude, où nous avons placé 21 capteurs dans 5 appartements du centre urbain de Paris (Figure 1).

On rappelle que la mesure de température est complexe en météorologie puisqu'on observe de fortes variations spatiales selon la topographie, l'éloignement d'une ville, la présence de végétation, les obstacles géographiques locaux, etc. Nous avons donc mis plusieurs capteurs dans des appartements non climatisés pour étudier les variations locales et nous avons étalonné les mesures en début et en fin de campagne.

¹ <http://www.cnrm-game-meteo.fr/surfex/>

Caractériser et représenter la dangerosité des canicules

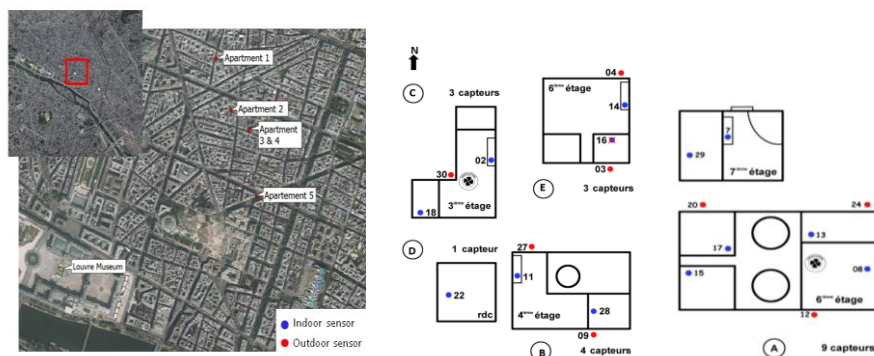


FIGURE 1. Localisation des différents appartements et des capteurs

A partir d’analyses bibliographiques et d’analyses statistiques, 5 types d’appartements, adaptés au centre de Paris, ont été créés en fonction de leurs comportements thermiques. Le 6^{ième} type est le type utilisé par le modèle SURFEX (Figure 2).

Appart1 : Appart-type	Appart2 : Appart-type	Appart3 : Appart-type	Appart4 : Appart-type	Appart5 : Appart-type
Bâti <1850 1 seule expo (rue) Isolé Très peu ventilé Rez-de-chaussée Orientation : E	Bâti haussmannien Double expo Bien isolé Ventilé 6 ^{ème} étage Orientation : N/S	Bâti <1850 1 seule expo (rue) Bien isolé Peu ventilé Sous les toits Orientation : S	Bâti <1850 1 seule expo (cour) Très bien isolé Très peu ventilé 3 ^{ème} étage Orientation : N	Bâti <1850 Double exposition Bien isolé Ventilé Etage élevé Orientation : N/S

FIGURE 2. Diagramme objet des appartements types

Les données couplées au modèle SURFEX par fusion de données ont permis de différencier les températures entre les bâtiments.

Les mesures ont été recueillies, structurées et analysées. Elles couvrent la période du 1^{er} juillet 2015 jusqu’au 1^{er} août minuit 2015 avec un pas de temps de 30min. Pour interpréter les résultats en 3.2 et 3.3 on résume la situation climatique du mois de juillet par 5 périodes : de J1 à J6 correspond à un épisode de forte chaleur, période la plus chaude (la première petite canicule), de J7 à J15, on retrouve un léger rafraichissement, de J16 à J19 est caractérisé une nouvelle période chaude, de J20 à J25 un léger rafraichissement qui s’accroît après J25.

SAGEO'2016 – Nice, 6-9 décembre 2016

2.3. Estimation des températures par fusion de données du modèle et des mesures

Pour mieux estimer les températures intérieures et extérieures, nous partons de deux séries de données présentées aux sections 2.1 et 2.2 : Les données SURFEX calculées toutes les 30mn pendant 1 mois et des mesures faites au centre de Paris (donc compris dans une maille du modèle) mais dans différents appartements. Les températures issues de SURFEX couvrent tout Paris mais sont calculées pour un type d'appartement standard, alors que les observations sont réalisées de manière ponctuelle dans un appartement donné correspondant à des caractéristiques thermiques et des configurations bien spécifiques (2.2). Nous cherchons ici à obtenir la couverture totale de Paris selon le quadrillage de SURFEX mais pour différents type d'appartements. L'objectif est ici d'intégrer les températures mesurées de jour et de nuit et de les recalcr dans un contexte spatial spécifique donné par le modèle SURFEX qui différencie localement les températures (intérieures comme extérieures) par fusion de données. Dans notre cas, nous avons utilisé une méthode très simple de fusion permettant de recalcr pour chaque pas de temps les températures d'une maille pour un type d'appartement en fonction de la température de cette maille pour l'appartement standard et de l'écart de températures entre appartements dans les mailles mesurées. Bien que très simple, cette méthode donne une estimation différenciée des températures plus pertinente que l'estimation initiale pour un type standard d'appartement (figure 3).

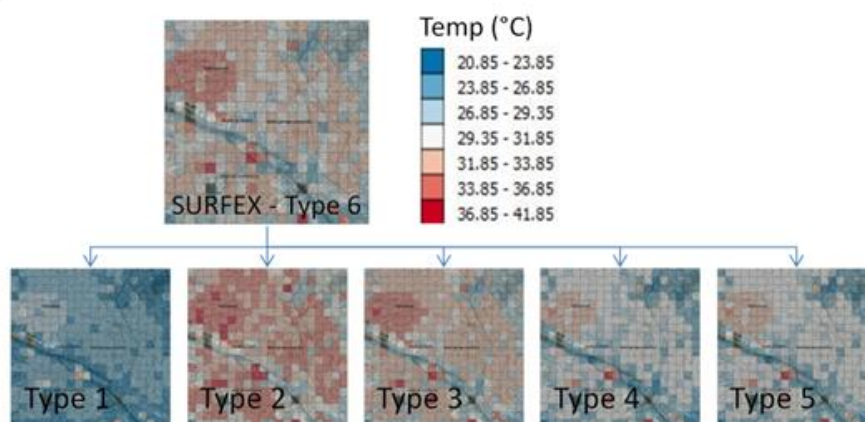


FIGURE 3. Estimation des températures intérieures de jour le 1^{er} juillet 2015 pour les 5 types d'appartements observés

Nous disposons pour chaque type d'appartement d'un tableau de 400 lignes (20*20 mailles correspondant à notre zone d'étude) et de 61 colonnes (correspondant à la période étudiée) que nous importons de R dans QGIS en .csv. Avec l'aide de la fonction *Create-a-Layer-from-a-Delimited* de QGIS, la visualisation des cartes est instantanée. Chaque maille est géoréférencée et comprend une valeur de température pour chaque jour et nuit de l'étude. Nous nous sommes focalisés sur les heures représentant au mieux la canicule (les plus chaudes et les plus fraîches) soit de 15h-

Caractériser et représenter la dangerosité des canicules

22h pour la journée et de 3h-10h pour la nuit (soit de 13h-20h et 1h-8h en heure solaire). Ces données ont été moyennées sur ces plages horaires (durée de 8h). On obtient ainsi une visualisation globale des températures en fonction des différents types d'appartements.

Ces estimations (figure 3) permettent de mettre en évidence le fait que certains types d'appartements ont des températures bien inférieures à celles estimées par SURFEX et d'autres, au contraire, ont de fortes similitudes avec le modèle, sur certains jours spécifiques. Les configurations spatiales et thermiques des bâtiments jouent un rôle crucial dans le comportement des températures et sont à prendre en compte lorsqu'on veut estimer la dangerosité de la canicule.

Un modèle conceptuel spécifique est développé pour représenter la durée de l'événement caniculaire et sa dangerosité à partir des mesures de campagnes *observation ponctuelle* et des sorties du modèles SURFEX-TEB *maille météo* (figure 4).

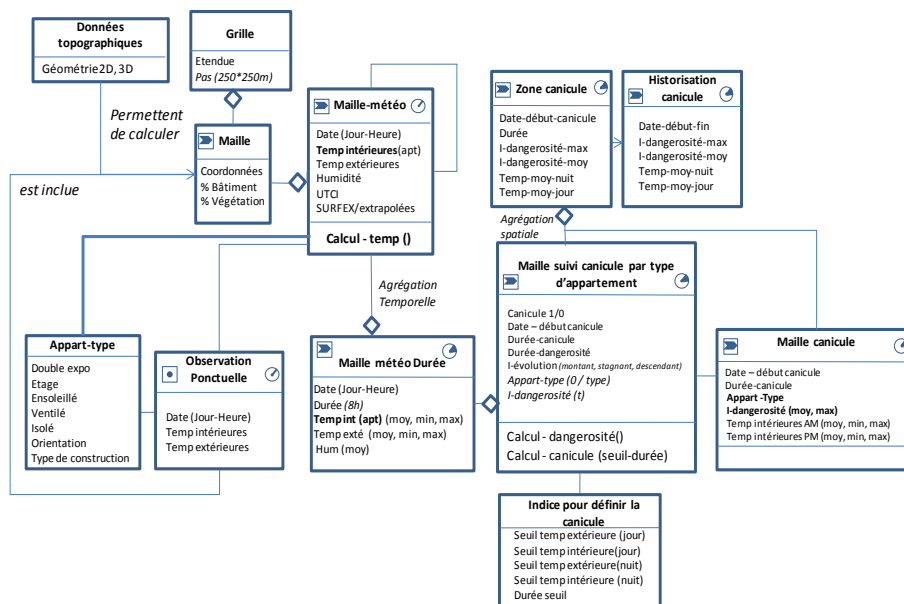


FIGURE 4. Schéma conceptuel de données

Une période de fortes chaleurs ne veut pas dire canicule. La question est donc de savoir si elle va se transformer en canicule. Pour assurer le suivi des températures, la maille suivi canicule est créée et mise à jour toutes les 12h. Cette maille va permettre de calculer la canicule et un indice de dangerosité. Pour la maille suivi canicule par type d'appartement, trois possibilités sont prises en compte :

- Soit les dernières heures ne répondent pas aux critères de canicule et la valeur de l'objet est égale à 0

SAGEO'2016 – Nice, 6-9 décembre 2016

- Soit les températures dépassent le seuil, l'objet enregistre et stocke l'information
- Soit la maille est déjà en canicule, elle met à jour ses attributs, dont ses valeurs thermiques, et la valeur de l'objet est égale à 1

Pour la présentation de dangerosité, un indice est développé. Cet indice indique le début d'une canicule, et déclenche l'alerte. Il permet ainsi de mieux vivre les canicules en ville.

3. Indice de dangerosité

3.1. Construction d'un indice

La prise en compte du temps est cruciale dans l'étude du phénomène caniculaire puisque la canicule devient préjudiciable à partir d'une certaine durée. Nous utilisons le terme de *dangerosité* pour souligner le fait que nous cherchons à identifier le moment à partir duquel certaines personnes peuvent être en danger, il s'agit donc d'une dangerosité potentielle. Le danger réel est fonction de la personne (état de santé général, âge, maladies spécifiques) et de son isolement. L'indice présenté est imparfait parce qu'il n'intègre pas toutes les informations, mais il doit permettre d'alerter sur une dangerosité potentielle.

L'indice que nous souhaitons construire doit intégrer la durée de l'événement, son intensité (en se basant sur les températures moyennes) mais aussi l'accumulation et la récupération d'une personne. En effet, si deux canicules s'enchaînent à peu de jours d'intervalle, le corps sera plus fragilisé à la deuxième canicule qu'à la première. Ceci est important parce qu'on peut avoir, certains étés, une série de canicules qui au final sont préjudiciables à la santé des plus fragiles. L'indice de dangerosité proposé permet d'évaluer le niveau de risque pour les habitants dans différents types d'appartements et de configuration, dans un centre urbain lors d'épisodes de canicule. Plusieurs fonctions ont été testées et analysées. Les fonctions sont du type :

Si température-jour > seuil-jour ou température-nuit > seuil-nuit alors

$$\text{Indice}_{ti} = \text{Indice}_{ti-1} + p1 + \text{Entier-sup}(\text{température}_{\text{Kelvin}} - \text{seuil}(\text{jour, nuit})) * p2$$

Sinon :

$$\text{Indice}_{ti} = \max(0, \text{Indice}_{ti-1} - p3(\text{profil}))$$

Avec

- p1 : si la température continue à être au dessus du seuil, alors la situation s'aggrave. Expérimentalement nous avons choisi p1 = 0.45.
- L'expression suivante exprime le fait que plus la température est élevée, plus l'indice augmente. Expérimentalement nous avons choisi p2 = 0.05 ce qui fait pour un degré au dessus du seuil, l'indice s'incrémente de 0.5 point (p1 + 0.05) toutes les 12h
- p3 est la récupération. Si la température est inférieure au seuil alors formellement l'épisode de canicule est fini. Par contre les personnes récupèrent progressivement. Dans une première série d'expérimentations

Caractériser et représenter la dangerosité des canicules

nous avons distingué deux profils. Pour une personne fragilisée (âgée, très jeune, malade) nous avons pris $p_3 = 0.5$, sinon $p_3 = 1$.

Cette fonction a été créée en prenant en compte les seuils de températures de jours (c'est-à-dire 31°C) et de nuit (21°C). Ainsi, si l'on prend un seuil-nuit de 21°C, le premier jour, son indice sera de 0,50 et évoluera de 0.05 à chaque degré en plus des seuils. Ces seuils sont basés sur les indices IBM (indice biométéorologiques). Cependant, ces seuils sont des seuils fixés pour des températures extérieures, or, on sait que la population passe plus de 80% de son temps dans un espace clos. De plus, les températures intérieures ont de plus faibles amplitudes thermiques que celles extérieures (Pinson L, 2015 AIC). Nous avons donc fait évoluer ces seuils afin qu'ils correspondent à des températures acceptables à l'intérieur. Nous nous sommes donc basés sur des seuils de 24°C de nuit et 28°C de jour.

Cet indice permet de donner les premières pistes de réflexion sur la dangerosité de la canicule. Plusieurs tests ont été réalisés, avec plusieurs seuils et pas de récupération. Les résultats et les illustrations sont présentés dans la partie suivante.

3.2. Analyse des valeurs de l'indice : suivi temporel dans la maille observée

L'indice de dangerosité est appliqué ici aux différents types d'appartements. En fonction du seuil de température choisi et de l'importance que l'on donne à la récupération, les indices évoluent très sensiblement différemment (Figure 4).

Pour les deux premiers graphiques, correspondant aux seuils de 24°C de nuit et de 31°C de jour, les périodes de vagues de chaleur sont moins visibles, mais l'indice met en valeur les particularités des appartements : les appartements 2 et 3 sont bien distincts des autres, surtout pour un pas p_3 de 1. Les appartements 1 et 5 ont des indices similaires sur les deux graphiques. Or, nous avons constaté que les températures de ces appartements sont éloignées surtout dans le temps. D'autres scénarios sont réalisés.

Si on compare le troisième et le quatrième graphiques (mêmes seuils et pas de récupération p_3 différents, une température de 24°C de nuit et de 28°C de jour), on remarque que pour la première période de canicule (de J1 à J5), les courbes vont croître avec une forte ressemblance pour atteindre un indice de 10 la 8^{ème} nuit. Après, en fonction de l'importance de la récupération, les courbes vont croître ou décroître. Dans le cas d'une faible récupération ($p_3 = 0.5$), les courbes augmentent mais avec un faible rythme. Dans l'autre cas, les courbes diminuent. La deuxième période de vague de chaleur (de J15 à J22) est caractérisée par une forte et continue augmentation pour atteindre quasi les indices 16 ou 20, pour les appartements les plus chauds. Dans le cas où la récupération à un pas de 1, on constate que pour les appartements 3 et 5, il suffit de 5 jours pour revenir à un indice de 0. Il faudra attendre encore quelques jours de plus pour les autres appartements (2 et 4) puisque leur inertie thermique est plus importante.

De ces graphiques, on peut observer que l'appartement 1 aura un indice toujours très faible, voir 0. Cela est dû, en grande partie, à la localisation. De plus, pour ces

SAGEO'2016 – Nice, 6-9 décembre 2016

seuils (24°, 28°, p 3 = 0.5 ou 1), on constate que lors des vagues de chaleur (J1-5 et J16-19) les seuils, de jour comme de nuit, sont dépassés puisque les courbes sont toutes ascendantes (à dérivée positive) (à l'exception du type 1).

Ces analyses nous permettent de mettre en valeur l'importance des seuils de température et de récupération :

- Lorsque les seuils 24°C-31°C sont utilisés, les courbes ne sont pas satisfaisantes. Les températures intérieures nocturnes sont généralement supérieures à 21°C (dus à l'inertie thermique des bâtiments) et les températures diurnes inférieures à 31°C. Les seuils de 24°C et de 28°C, eux, permettent de mieux mettre en valeur les vagues de chaleur et le comportement thermique de chaque type d'appartement.
- Les graphiques sont sensibles à la valeur du seuil de récupération qui par ailleurs met bien en évidence la configuration spatiale de chaque appartement et l'inertie. Avec ces graphiques, surtout les deux derniers, on constate que même si une période de « rafraîchissement » est présente cela ne veut pas dire que le risque n'existe plus. Il faut plusieurs jours, en fonction des personnes et des configurations, pour atteindre une situation sans risque.

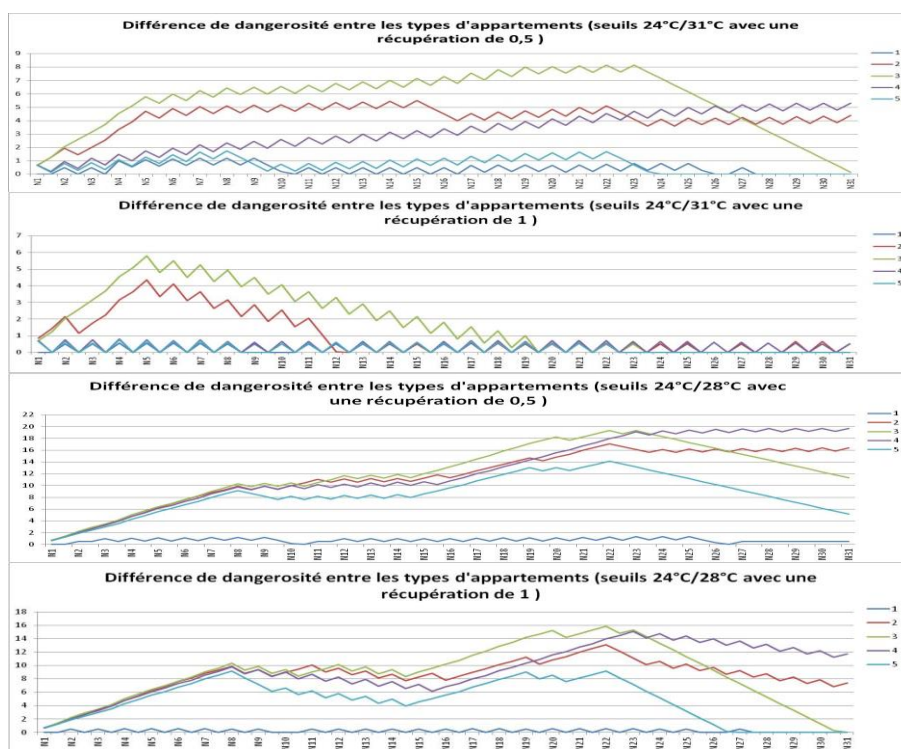


FIGURE 4. Indices de dangerosité sur 30 jours pour les 5 types d'appartements

Caractériser et représenter la dangerosité des canicules

3.3. Analyse de valeurs de l'indice : analyse spatio-temporelle

La méthode de fusion de données permet de calculer les températures intérieures sur tout Paris pour différents types d'appartement. Dans cette étude nous nous focalisons davantage sur les températures intérieures, souvent sous estimées. L'analyse de la durée des phénomènes, pour en déduire des zones de dangerosité «hot spot» à l'échelle du quartier, est possible.

La représentation cartographique permet de détecter visuellement les zones de forte intensité à chaque temporalité, ainsi qu'au cours du temps (Figure 5). Le type 1, comparé aux deux autres, est celui qui concentre les hot-spot de plus faible intensité. Pour ce type, les indices sont faibles mais les seuils de températures sont tout de même dépassés pour la plupart des mailles étudiées. Pour l'ensemble des types, les indices augmentent sur toute la période et commencent à décroître à partir de la 25^{ème} nuit. Ces cartographies nous permettent, par conséquent de comparer avec l'état précédent, afin de désigner si la canicule est en période ascendante (si la température continue à augmenter dans la durée), stagnante ou descendante (si la température régresse pour passer en dessous des seuils fixés de jour et de nuit). Cette représentation illustre bien que les appartements en rez-de-chaussée (type 1) résistent bien mieux à la canicule, les appartements très peu ventilés (type 4) subissent plus longtemps les vagues de chaleurs que les autres. On voit également que les configurations spatiales et les aménagements semblent jouer un rôle non négligeable puisque la température et a fortiori la dangerosité connaissent d'importantes variations spatiales.

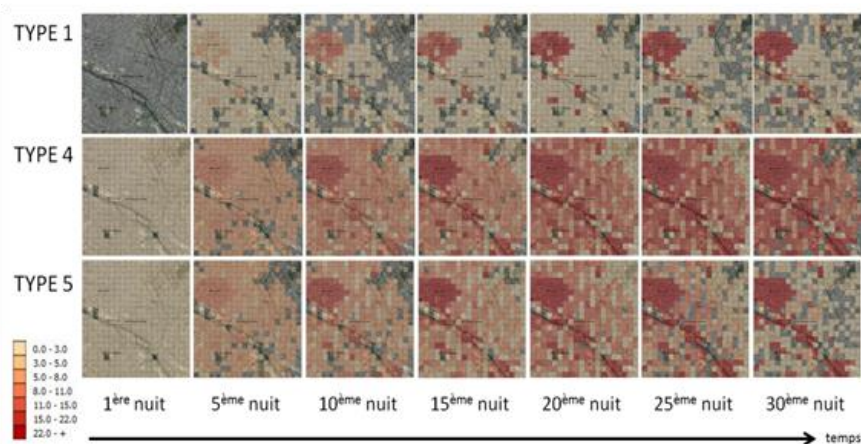


FIGURE 5. Représentations cartographiques de l'indice de dangerosité (avec les seuils 24/28°C et une récupération de 0.5) utilisés pour 3 types d'appartements

Par la suite, des requêtes sont effectuées permettant de repérer les hot-spots via le langage de requête SQL. L'exemple de requête en spatial-SQL ci-dessous sélectionne les mailles où l'indice de dangerosité est supérieure à 15 et où la canicule est ressentie pour la nuit du 10 juillet :

SAGEO'2016 – Nice, 6-9 décembre 2016

```
SELECT canicule.*,indice.*
FROM canicule, indice
WHERE canicule.nuit10=12 AND indice.nuit10>=15
```

Ces requêtes permettent de réaliser une agrégation spatiale. Comme nous souhaitons faire ressortir les lieux où le phénomène dure, l'utilisation de l'indice de dangerosité est parfaite puisqu'il est bâti sur la durée. L'accumulation des zones de fortes chaleurs produit des zones de hot-spot dues à la durée du phénomène. L'analyse de la requête spatiale entre la carte de la dangerosité et celle de la canicule montre les zones les plus vulnérables.

Les analyses nécessaires au suivi des canicules produisent d'importants volumes de données et de couches graphiques à superposer. Afin de faciliter la compréhension des phénomènes, nous développons un site web qui permettrait de mieux informer sur les caractéristiques du phénomène représenté et de mieux identifier les zones et types d'appartement inquiétants. Cela permettrait, par exemple, des actions de communications ciblées sur certains quartiers. Cette conception de nouvelles représentations nécessite un développement de structures de données adhoc (Cheaib et al. 2013).

4. Mise en place du service web

La mise en œuvre du service web d'exploration des canicules repose sur un ensemble de logiciels open source (Figure 6) (Tyler M. 2005) :

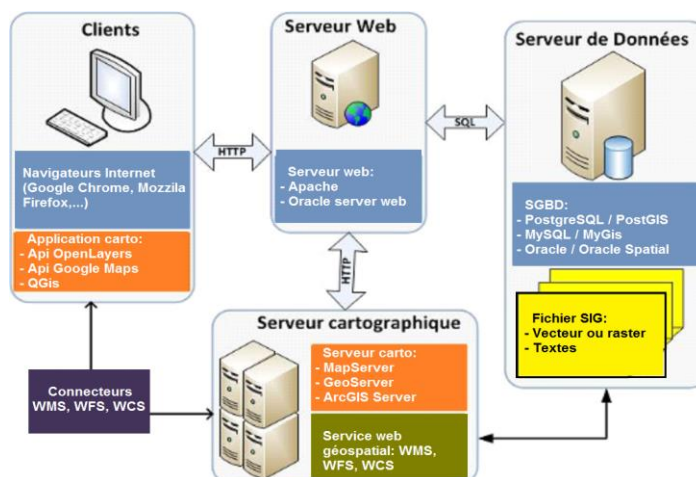


FIGURE 6. Principe général de fonctionnement du Web mapping © (Millogo 2012, issus des travaux de Tyler M. 2005)

² Dans notre table de données 1 correspond à des mailles canicules et 0 à des mailles non canicules

Caractériser et représenter la dangerosité des canicules

- PostgreSQL/PostGIS pour stocker et manipuler les données avec
- QGIS pour visualiser les résultats et créer des fichiers de styles (.sld)
- GeoServer pour publier des couches cartographiques à partir de données stockées sous PostgreSQL/PostGIS
- OpenLayers pour agréger les couches cartographiques à l'aide de la bibliothèque
- HTML, CSS, JavaScript pour créer l'interface web

4.1. Symbolisation des données dans QGIS

Les données sont visualisées dans QGIS sous forme de mailles colorées. La symbolisation QGIS est écrite dans un fichier .sld utilisé à l'étape suivante. Un fichier SLD est un schéma XML indiqué par l'OGC (The Open Geospatial Consortium) qui permet de décrire le style des couches de carte. Il définit des règles de symbolisation pour des ensembles d'entités spécifiées au sein d'une couche WMS (Web Map Service) donnée. Le fichier peut comporter plusieurs règles, et chaque règle peut compter un filtre décrivant quelles entités doivent être affichées, ainsi que plusieurs symboles décrivant comment ces entités doivent être affichées.

4.2. Publication des couches de données avec GeoServer

GeoServer est un serveur cartographique. Il publie des couches cartographiques à partir des données stockées dans PostGIS et utilise les fichiers .sld pour définir la symbolisation de ces couches. Une couche cartographique de GeoServer correspond à une table de données dans PostGIS.

4.3 Mise en place du service web avec OpenLayers

L'interface cartographique est la fenêtre qui va contenir notre carte. Son développement est réalisé grâce à l'utilisation du client cartographique OpenLayers avec JavaScript. Cette interface cartographique contient la carte et les outils de manipulation. La carte est composée de *données topographiques* (données topographiques, orthophotographies) et de *données thématiques* (de la température ou de l'indice de dangerosité)

- L'information topographique (bâtiments, routes, rivières, parcs, etc) sont des données invariables pour notre application. Elles décrivent l'espace support. Elles sont écrites dans une table sur PostGIS et correspondent à une seule couche dans GeoServer. Notre carte est l'agrégation des couches graphiques, une couche de données topographiques est ajoutée dans la carte par la fonction :

map.addLayers([chemin de données/nom de couche])

SAGEO'2016 – Nice, 6-9 décembre 2016

- Les données thématiques sont les données météorologiques composées des températures et de l'indice de la dangerosité. Elles sont dynamiques alors nous les sauvegardons dans plusieurs tableaux sur PostGIS, chaque table correspond à la température ou à l'indice pour une période (nuit ou jour) (un état). Initialement, la couche cartographique « température » ou « indice » est ajoutée dans la carte comme une couche de données statiques, ses paramètres sont ceux du premier état. Lorsqu'on lance la simulation, le programme met à jour le chemin de données de la couche de l'état suivant et l'affiche dans la carte.

*map.layers[indice_de_couche].mergeNewParams({nouveau
chemin de données/nom de couche});*

Le service web en cours de conception permet de choisir de visualiser la température extérieure ou intérieure ou l'indice de dangerosité. Pour la température intérieure et pour l'indice, on peut choisir le type de bâtiment étudié. On peut ensuite soit lancer la séquence des états, soit se déplacer à un état donné. La figure 7 représente la température intérieure moyenne pour la nuit du 4 juillet pour l'appartement du modèle SURFEX.

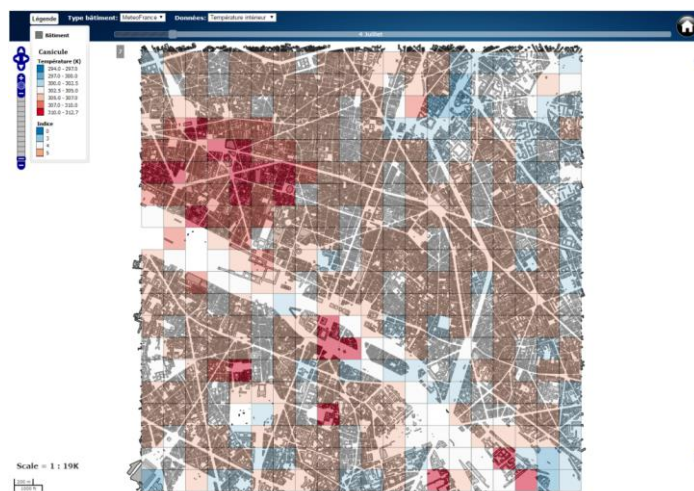


FIGURE 7. Services web de visualisation de la dangerosité d'une canicule (température intérieure de nuit, pour le type 6 SURFEX, le 4 juillet 2015)

Le site web est en cours de finalisation. Il sera disponible sur le site <http://representation-phenomenes.ifsttar.fr>

5. Conclusions et perspectives

Dans un contexte de changement climatique, il est important de mieux identifier la dangerosité du risque caniculaire pour développer les stratégies d'intervention, de

Caractériser et représenter la dangerosité des canicules

communication et d'adaptation les plus appropriées. Dans cette recherche nous avons construit un premier indice à partir d'un modèle et d'observations. Davantage de campagnes d'observations permettraient de consolider l'estimation des températures intérieures mais les résultats sont déjà éclairants.

Les résultats des premiers indices développés sont intéressants : les indices reproduisent bien les observations du comportement des températures intérieures, ainsi que les cycles diurne/nocturne. Les sorties graphiques mais aussi cartographiques montrent bien que chaque bâtiment, appartements, quartier de la vie ne sont pas des blocs homogènes et similaires mais que des différences existent et modifient la dangerosité de la canicule. La encore il devrait être possible d'améliorer l'indice (notamment le choix des seuils) en élargissant la campagne d'observations.

Cette recherche confirme l'intérêt de prendre en considération les températures intérieures et de montrer leur dangerosité en fonction des différentes configurations spatiales, de l'intensité et de la durée du phénomène. Un site web est en cours de conception pour mieux donner à voir ce phénomène.

De plus, une étude sur la perception et le ressenti du risque auprès de la population et des acteurs pourrait apporter des éléments utiles et de nouvelles connaissances.

Bibliographie et références

- Bouttier F. (2007). Arome, avenir de la prévision régionale. *La Météorologie*, 58, p.12–20.
- Cheab N., Ruas A., Gaborit. (2013). From sensor data to the perception of phenomena: Software architecture for online access and offline analysis, *Actes du 26th International Cartographic Conference 2013*, Dresden.
- Masson V. (2000). A physically-based scheme for the urban energy budget in atmospheric models. *Bound.- Layer Meteorol.*, 94, p.357–397.
- Millogo F. (2012). *Mise en place d'une application web mapping de géolocalisation des points d'intérêt de la ville de Ouagadougou*. Mémoire Master2 Informatique Appliquée aux Systèmes d'Information Géographique, Université de Douala.
- Munich Re Group Annual Report 2006* (https://www.munichre.com/publications/302-05252_en.pdf)
- Oke T-R. (1987). *Boundary layer climates*. London et New York : *Methun*, Second edition (first edition 1978), p. 435
- 5^{ème} rapport d'évaluation du GIEC. IPCC Climate Change 2014 : synthesis report. Copenhague 2014 (<http://www.ipcc.ch/report/ar5/index.shtml>)
- Tyler M. (2005). *Web Mapping Illustrated, Using Open Source GIS Toolkits*, *O'Reilly*, ISBN: 0-596-00865-1, p. 367

Analyser des tweets géolocalisés pour explorer les réponses sociales face aux phénomènes météorologiques extrêmes

Réflexions épistémologiques et verrous méthodologiques

Camille Cavalière¹, Paule-Annick Davoine¹, Céline Lutoff², Isabelle Ruin³

1. *Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes
700 Avenue Centrale, F-38401 Saint-Martin-d'Hères, France
{Camille.Cavaliere,Paule-Annick.Davoine}@imag.fr*

2. *Laboratoire PACTE-Territoires, Université Grenoble Alpes
14 bis, Avenue Marie Reynoard, F-38100 Grenoble, France
Celine.Lutoff@univ-grenoble-alpes.fr*

3. *Laboratoire LTHE, Université Grenoble Alpes
CS 40700, F-38058 Grenoble, France
Isabelle.Ruin@univ-grenoble-alpes.fr*

RESUME. En automne, de violentes perturbations météorologiques frappent régulièrement le sud-est de la France ; pendant de tels événements, habitants et autorités locales diffusent de l'information via les médias sociaux. Cette information constitue une source de données gratuite et utile pour l'analyse sociale de l'événement, de la manière dont les personnes l'ont vécu. Cet article repose sur l'étude des tweets géolocalisés émis dans huit départements à l'automne 2014 : nous proposons de traiter l'information selon une posture de recherche originale : le raisonnement abductif.

ABSTRACT. In fall, the south-east of France is regularly affected by heavy weather disturbances. During those natural disasters, citizens and authorities turn to social media platforms to share information. They can therefore provide free and relevant in-field information. This paper aims at exploring georeferenced tweets created within eight departments in the fall 2014: we intend to process the twitter data using an original research approach, called abductive reasoning.

MOTS-CLES : MEDIAS SOCIAUX · TWITTER · PHENOMENES EXTREMES · ABDUCTION

KEYWORDS: SOCIAL MEDIA · TWITTER · EXTREME PHENOMENA · ABDUCTION

1. Introduction

Au cours de la dernière décennie, les géographes ont assisté à la transformation de la nature, de la disponibilité et du volume de l'information géographique. La rencontre entre les systèmes de géolocalisation, dont l'usage s'est largement répandu grâce aux *GPS* et aux *Smartphones*, et le *Web 2.0*, constitue le facteur décisif ayant favorisé l'explosion de la disponibilité de données géographiques publiques (Miller, 2007). L'environnement de travail du chercheur est désormais constitué d'un monde extrêmement riche de données (Kitchin, 2013). Twitter est une plateforme du *Web 2.0* via laquelle les utilisateurs, nommés *tweeters*, créent et échangent quotidiennement de l'information relative à leurs activités ou opinions, tout en ayant la possibilité d'ajouter une information géographique, sous forme de coordonnées ou de *hashtags* spatiaux (villes, quartiers, etc.). Créé en 2006, Twitter a bénéficié d'un succès rapide et, en 2015, la plateforme enregistre 304 millions d'utilisateurs mensuels actifs pour un effectif quotidien de tweets émis estimé à 500 millions¹. Le service fournit ainsi une source intarissable de données numériques centrées sur l'individu, ses opinions et ses interactions avec son environnement.

Ces dernières années, les travaux fondés sur l'étude de tweets géolocalisés ont été entrepris dans des thématiques variées : l'étude de la répartition spatiale des *tweeters* et la définition de profils sociaux (Li *et al.*, 2013), la mobilité internationale (Hawelka *et al.*, 2014), les relations entre individus, temps et espace (Andrienko *et al.*, 2013), ou encore l'étude de la dynamique et du cycle de vie des *hashtags* (Lin *et al.*, 2013). L'objectif de ces études consiste à identifier des structures régulières et répétitives dans le temps et dans l'espace (Erwig, 2004), les changements qui peuvent survenir, ou encore des comportements de *tweeting* qui révèlent les pratiques et habitudes individuelles ou collectives des *tweeters* (Andrienko *et al.*, 2013). D'autres auteurs se sont intéressés à l'exploration des réponses sociales lors de la survenue de phénomènes extrêmes, sociaux ou environnementaux : les réactions du public face à des événements exceptionnels qui affectent des milliers d'individus à l'échelle d'un pays (Lee Hughes et Palen, 2009), les réactions du public face à la criminalité (Kounadi *et al.*, 2015) ou encore les risques naturels (Dashti *et al.*, 2014). Les travaux effectués dans ce domaine ont démontré la pertinence de ces données particulières comme source d'information spatio-temporelle pour la détection et la caractérisation d'événements naturels en temps réel (De Longeville *et al.*, 2009 ; Dittrich et Lucas, 2014).

Pour autant, l'observation des réponses sociales aux phénomènes extrêmes, qu'il s'agisse de catastrophes naturelles ou d'attentats, soulève l'interrogation de nos capacités à appréhender les caractéristiques de la donnée tweet afin d'identifier les obstacles qui contraignent l'analyse et l'interprétation des données. En effet, si les tweets peuvent constituer une source de données utile à une nouvelle et meilleure compréhension des phénomènes sociaux et des comportements des individus en période de crise, leur analyse requiert une réflexion approfondie sur la nature des données, les problèmes méthodologiques et l'approche scientifique à adopter : en

¹ <http://www.blogdumoderateur.com/chiffres-twitter/>

SAGEO'2016 – Nice, 6-9 décembre 2016

effet, un certain nombre de paramètres peuvent biaiser la pertinence de la donnée comme l'état psychologique de la personne qui crée l'information au moment où elle vit l'événement, l'incomplétude ou l'imprécision de l'information produite ou encore la distance entre l'individu qui crée la donnée et l'événement (Hertfort *et al.*, 2014). Dans un second temps, la donnée tweet appartient à ce que l'on désigne aujourd'hui sous le nom de *Big Data*, en référence aux données qui dépassent nos capacités et méthodes d'analyse traditionnelles (Miller et Goodchild, 2015). Il ne s'agit plus de collecter des données pour répondre à une question préalablement identifiée, mais de collecter d'abord pour identifier ensuite les questions auxquelles les données peuvent répondre (Miller et Goodchild, 2015). Cette posture se traduit par la primauté de l'observation, suivie par la formulation d'hypothèses.

Dans cet article, nous nous intéressons aux tweets géolocalisés produits dans le contexte des précipitations extrêmes et plus particulièrement des pluies et inondations cévenoles. Nous proposons ainsi une méthodologie originale d'approche et de traitement des données, destinée à favoriser la découverte de phénomènes par le raisonnement abductif qui s'inscrit dans la posture précédemment évoquée. Nous menons cette recherche dans le but d'estimer la pertinence du tweet comme indicateur de survenue d'une crise et de détecter les problèmes sous-jacents à son utilisation. Notre méthodologie est appliquée à l'événementiel hydro-météorologique - les épisodes cévenols - qui a affecté huit départements du sud-est de la France entre octobre et décembre 2014. Ces événements, nommés crues rapides, se caractérisent par des orages convectifs qui provoquent de courts épisodes de pluies torrentielles entraînant un ruissellement intense au sein de tous les bassins versants et l'élévation subite du niveau des rivières.

L'article est organisé de la manière suivante : la partie 2 présente les challenges associés à l'exploitation de telles données, l'ouverture que propose l'approche abductive et l'application que nous soumettons dans le cadre de cette recherche. La partie 3 présente les étapes successives de l'analyse des données. Finalement, la partie 4 conclut par une évaluation globale de la recherche et propose de prochains travaux envisageables.

2. L'abduction comme méthode pour l'exploration de la donnée tweet

Cette partie introduit les problèmes posés par la donnée tweet dans le cas particulier des phénomènes extrêmes et décrit la posture de recherche adoptée vis-à-vis des données. Elle présente les réflexions préalables à leur traitement, illustrées par une démarche destinée à favoriser la découverte de motifs intéressants.

2.1. Regard sur la donnée tweet dans un contexte événementiel

Les tweets géolocalisés sont considérés comme des capteurs sociaux qui détectent et enregistrent les perturbations survenues (De Longeville *et al.*, 2009). La communauté scientifique a montré un intérêt croissant, notamment au travers de la *Volunteered Geographic Information* (VGI) et du *crowdsourcing*, pour les données

Cpcnř ugt 'rgu'vy ggw'ı 2qręcrkř u

produites par des individus équipés d'appareils géolocalisables, qui communiquent en cas de survenue d'événement exceptionnel (Dittrich et Lucas, 2014). Néanmoins, contrairement à la VGI, les tweets ne sont pas forcément créés dans l'intention de diffuser une information ultérieurement utilisée par un tiers. En d'autres termes et dans la plupart des cas, les tweets ne sont pas créés dans un objectif de satisfaction d'exigences scientifiques. Par conséquent, l'information contenue est potentiellement non structurée, confuse, voire vague par rapport à notre objet d'étude. En outre, ce type de donnée est issu d'une réalité perçue par l'utilisateur : le tweet traduit le regard de l'individu vis-à-vis de l'événement : il révèle ainsi une représentation, qui peut varier sémantiquement (qualité du discours) mais également en fonction de l'expérience vécue de l'individu (l'événement est-il nouveau ou déjà connu ?) et de son degré d'objectivité (exagère-t-il un fait observé ?).

Enfin, l'existence du tweet de crise géolocalisé dépend d'une multitude de facteurs : il est le produit d'un contexte environnemental précis qui interagit avec la sphère sociale, et est donc lié à la réactivité et à l'implication des individus. Ces tweets résultent d'une motivation instantanée induite par la participation du *tweeter* à un événement inhabituel qui rompt avec la routine de sa sphère quotidienne. Cela implique que l'information envoyée concerne l'utilisateur dans sa sphère environnementale et sociale directe. En revanche, nous ne pouvons pas affirmer que la donnée tweet capture l'événement dans sa globalité : le *tweeting* ne concerne que certaines catégories de la population, notamment en fonction de l'âge² ; toute la population qui *tweete* ne dispose pas de la même qualité de réseau et nous supposons que l'implication des *tweeters* dans la création d'information utile est le résultat de facteurs imbriqués comme l'expérience vécue, précédemment évoquée, ou encore la situation et le comportement du *tweeter* à l'instant *t*.

2.2. La place des tweets dans une nouvelle démarche d'analyse abductive

Les travaux précédents (cf. introduction) ont révélé la complexité du contenu des tweets et la nécessité de considérer toutes les facettes de cette information. La donnée tweet présente en effet trois dimensions susceptibles d'introduire une part conséquente d'hétérogénéité : elle est spatiale (De Longeville *et al.*, 2009), mais de précision variable (*hashtag*, coordonnées GPS). Elle est temporelle et peut donc nous permettre de suivre les séquences d'un événement (De Longeville *et al.*, 2009). Enfin, elle est sémantique (Dashti *et al.*, 2014) mais variable dans sa forme (texte, *hashtags*, images, URL) et en fonction des deux premiers paramètres : le contenu des tweets varie donc en fonction du temps et de l'espace.

Ces propriétés ont incité les chercheurs à soulever l'épineuse question des capacités à traiter et à exploiter ces données (Dashti *et al.*, 2014). Ainsi, l'évolution récente des données, qu'elles soient produites par des individus ou par des machines,

² En 2015, le nombre d'utilisateurs mensuels actifs en France est estimé à 2,3 millions. Source : <http://www.blogdumoderateur.com/chiffres-twitter/>

SAGEO'2016 – Nice, 6-9 décembre 2016

a favorisé l'émergence d'un nouveau paradigme, celui des sciences guidées par les données, ou *data-driven sciences* (Miller et Goodchild, 2015). La perspective d'une nouvelle compréhension des phénomènes géographiques se heurte à des masses de données spatio-temporelles et sémantiques hétérogènes, qu'il est difficile d'explorer de manière approfondie. Ainsi, les chercheurs peuvent difficilement entrevoir la richesse que les données peuvent potentiellement receler (Anderson, 2008). Cette richesse, il faut la découvrir : or, Kitchin (2013) note que les méthodes de traitement et d'analyse de données ont peu évolué depuis les années 1990. Ce même auteur affirme même que les *new forms of data science are in their infancy*. C'est pourquoi cette évolution s'est plutôt traduite, dans les sciences humaines, par un questionnement épistémologique sur les possibilités de donner du sens aux données en prenant en compte leurs dimensions géographique et sociale (Kitchin, 2013). Les prémices des *data-driven sciences* ont ainsi préconisé un rejet systématique de l'inférence déductive afin de privilégier l'attention portée à l'émergence et à la détection de motifs inattendus, à la formulation d'hypothèses *a posteriori* (Kell et Oliver, 2003) et à leur test.

Le raisonnement abductif a ainsi été récemment introduit comme inférence clé dans l'exploration des données et la découverte de phénomènes intéressants cachés dans des masses de données (Miller et Goodchild, 2015). Il s'agit d'une démarche explicative fondée sur la perception et la juxtaposition de faits observés pour remonter aux causes qui les produisent (Walters, 2012) et destinée à prévoir le comportement d'un système (Hoffmann, 1999). Elle fait donc référence à la capacité de générer des hypothèses face à un résultat inattendu (Hoffmann, 1999). Cette nouvelle approche a pour but de construire progressivement une connaissance ancrée sur les données (Kitchin, 2013). Nous proposons ici de voir de quelle manière elle nous permet d'étudier les questionnements qui émergent concernant les réponses sociales aux phénomènes naturels extrêmes.

2.3. Description de la démarche de recherche adoptée

La figure 1 illustre la démarche théorique de l'exploration des données que nous souhaitons tester. Nous associons les inférences déductive et abductive. A partir d'un constat établi dans l'état de l'art, nous formulons l'hypothèse suivante (Phase 1) : s'il existe une étroite correspondance entre densités de populations et densités de tweets (Lin *et al.*, 2013), nous supposons que les tweets liés à un phénomène extrême échappent à cette logique et peuvent ainsi constituer le marqueur social de la localisation, voire de l'intensité de ce phénomène, quelle que soit sa nature. Nous focalisons donc notre attention sur les anomalies de *tweeting*, c'est-à-dire les espaces où l'activité de *tweeting* est plus importante que la population, et pensons qu'ils sont davantage présents dans les aires urbaines. Nous faisons néanmoins le pari que les résultats permettront de détecter des motifs non conformes à ces attentes (B).

La phase 2 correspond à l'inférence abductive : l'observation intéressante détectée à la fin de l'étape précédente fait l'objet d'une nouvelle hypothèse (A) fondée sur l'information apportée par la donnée tweet. Cette hypothèse est à son tour soumise à l'expérience. Enfin, si l'existence d'un lien entre (A) et (B) semble se

Cpcnř ugt 'rgu'vy ggw'i 2 qmęcku? u

confirmer, il convient, en phase 3, d'explorer la validité de ce lien pour confirmer ou nuancer l'exactitude de l'hypothèse (A).

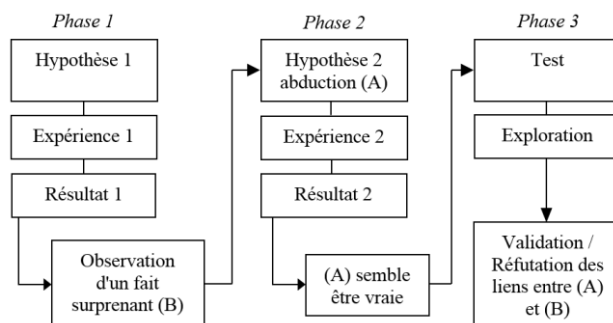


Figure 1. Démarche théorique de construction de la connaissance

3. Exploration et analyse des jeux de données

Nous avons travaillé sur un jeu de données brut, constitué de 597 741 tweets géolocalisés émis entre le 8 octobre et le 12 décembre 2014 dans les départements suivants, qui sont fréquemment affectés par les épisodes cévenols : Drôme, Ardèche, Gard, Lozère, Hérault, Bouches-du-Rhône, Var et Vaucluse. Ils ont été collectés par un serveur connecté à l'API Streaming de Twitter. Ce jeu constitue notre base pour l'identification de l'information de crise et l'analyse de données.

3.1. Construction d'un jeu de tweets de crise

Les tweets de crise, qui mentionnent une perturbation cévenole, peuvent être sélectionnés en fonction de leur contenu (Hertfort *et al.*, 2014). La sélection de l'information de crise est alors fondée sur la recherche de mots-clés à travers l'élaboration d'un glossaire. Nous ne nous focalisons pas en particulier sur les tweets qui contiennent des *hashtags* (comme par exemple *#inondation* ou *#intempéries*) : nous craignons en effet qu'il s'agisse d'un critère de sélection trop rigoureux qui n'extraierait qu'une infime partie du jeu de tweets de crise potentiel.

3.1.1. Construction du glossaire d'analyse

Le jeu de tweets de crise est construit en respectant deux contraintes : il doit être le plus complet possible en termes d'effectif et le plus précis possible sur le plan de la qualité de l'information, et de la minimisation du bruit résiduel.

Dans un premier temps, la sélection de mots-clés requiert une étape de *brainstorming*, pendant laquelle nous essayons d'inventorier un maximum de vocabulaire (ou associations lexicales) diversifié sur l'événement considéré. Dans le cas étudié, ces mots-clés ont été recherchés de la manière suivante :

SAGEO'2016 – Nice, 6-9 décembre 2016

- le vocabulaire événementiel contextuel : *orage, éclair, tonnerre, foudre, pluie, inondation* ;

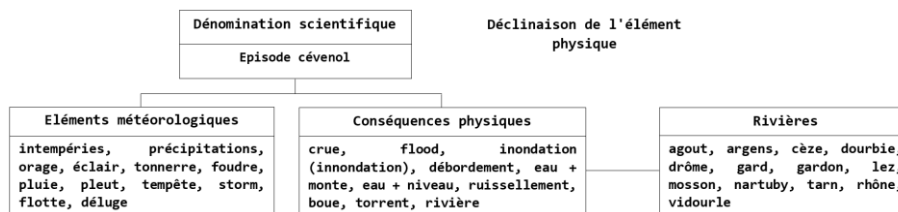
- le vocabulaire officiel ou scientifique qui décrit les phénomènes : *épisode cévenol, intempéries, précipitations*, ou encore le vocabulaire officiel de prévention des risques : *alerte, vigilance*. On peut facilement le relever en consultant la presse en ligne ;

- les noms de communes fréquemment affectées ainsi que les noms des rivières soumises aux crues rapides ;

- le vocabulaire informel et plus ou moins familier que les *tweeters* sont susceptibles d'employer pour évoquer l'événement : *flotte* à la place de *pluie*, *déborder* pour signaler un cours d'eau en crue.

La difficulté consiste ensuite à identifier le vocabulaire spécifique aux effets de ces événements, c'est-à-dire d'inventorier les éléments susceptibles d'être affectés : *l'électricité coupée, la route fermée, la maison isolée, l'intervention des pompiers, l'arbre déraciné*, etc. L'objectif est de mettre en évidence les tweets qui ne mentionnent pas explicitement des éléments physiques (comme l'orage ou l'inondation) mais qui seraient focalisés sur les effets de l'événement. En procédant de cette manière, nous nous attendons à ce que des tweets comme "*j'ai plus d'électricité chez moi*" et "*j'ai plus d'électricité chez moi #orage #intempéries*" soient conservés dans le jeu de données de crise final. Cette première recherche a permis de collecter un ensemble de 80 mots-clés et associations lexicales.

Dans un second temps, nous avons procédé à un examen interne des données, afin de mettre en évidence du vocabulaire employé par les *tweeters* et de compléter la liste précédemment dressée. Nous avons eu recours à des outils de *text mining* (KH Coder) afin de clustériser les mots utilisés dans les tweets et de repérer : du vocabulaire identifié comme directement lié à l'événementiel hydro-météorologique auquel nous n'avions pas songé, comme *grêle* ou *apocalypse* ; des mots que nous avons inscrits dans notre liste de recherche, mais dont l'orthographe est erronée dans les tweets (*tonnaire*) ; des mots ou associations lexicales potentiellement liés aux événements, et dont la validité doit être testée : les mots *murs* et *tremblent* apparaissent dans un cluster contenant le terme *orage*. En revanche, l'association *journée perturbée* n'a identifié que quelques tweets sans lien avec notre problématique. A l'issue de ce double travail, dix catégories de mots-clés sont proposées. Elles sont présentées dans la figure 2 et distinguent le vocabulaire des phénomènes physiques d'une part, ses répercussions sur l'individu et la sphère sociale d'autre part, puis les mots supplémentaires mis en évidence par *text mining*.



Cpcnř ugt'rgu'vy ggw'i 2 qmęcku? u

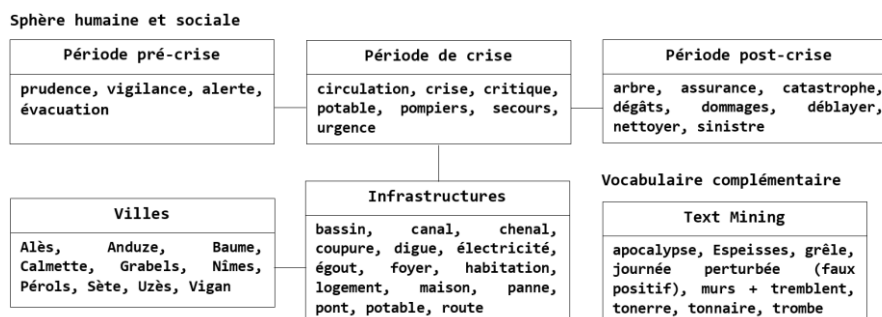


Figure 2. Liste des mots-clés inventoriés par catégories

3.1.2. Elimination du bruit résiduel et constitution du corpus final

Un certain nombre de mots-clés pré-définis ou découverts dans le corpus de tweets peuvent se retrouver dans des contextes variés. C’est le cas par exemple des termes *pompiers*, *secours*, *alerte*, *inondation*. Nous avons donc jugé essentiel de vérifier la correspondance entre les tweets que nous avons extraits et le contexte événementiel considéré. Cette étape de relecture et de « nettoyage » des données a été effectuée manuellement. Étant donné que de nombreux tweets peuvent être sujets à l’interprétation, nous avons défini un ensemble de règles de validation des tweets considérés comme pertinents (tableau 1). Pour être sélectionné, un tweet doit être centré sur l’événement (ses manifestations physiques et/ou ses conséquences) ou sur les réactions de l’individu.

Tableau 1. Liste des conditions et exemples de tweets sélectionnés

Thème du tweet	Exemple de tweet
bulletin ou alerte météorologique	<i>#Nîmes : arrêt des précipitations depuis 40 min. L'épisode devrait décroître d'ici la fin de la matinée</i>
événement en cours ou survenu	<i>Demain j'ai pas cours, mon lycée a été inondé !!</i>
conséquences (individus, biens, infrastructures)	<i>Le pont qui mène en ville est plein d'eau, j'suis coupé du monde ! ??</i>
inquiétude / peur	<i>Il pleut de dingue chez moi omg j'ai peur</i>
comportement	<i>Il fait nuit, il pleut, c'est l'alerte orange mais je fais les magasins</i>
opinion (gestion de la crise, secours)	<i>@meteofrance merci pour cette #alerterouge qui n'a servi à rien. Pas une goutte de pluie ici.</i>

SAGEO'2016 – Nice, 6-9 décembre 2016

Les tweets dont le contenu s'est avéré inintelligible, ambigu, ou exprimait un souhait (un élève suppliant l'annonce d'une alerte afin que ses cours soient annulés) n'ont pas été retenus. Finalement, sur les 10 750 tweets présents dans le jeu intermédiaire après exécution des requêtes, nous n'en avons conservé que 2 789, soit 26%. Au final, le jeu de tweets de crise ne représente que 0,47% du jeu de tweets brut initial pour l'ensemble des événements de la période observée.

3.2. Analyse spatio-temporelle : application de la démarche de recherche

Cette partie teste la démarche présentée au paragraphe 2.3. Nous étudions la distribution spatio-temporelle des tweets dans l'objectif de mettre en évidence des structures spatiales répétitives à travers le temps (échelle mensuelle).

3.2.1. Détection des hotspots d'activité (Phase 1)

La distribution spatio-temporelle des tweets a été analysée par l'étude de densités normalisées : l'objectif consiste à mettre en évidence les foyers (ou *hotspots*) potentiels d'activité en gommant les effets des densités de population, à partir de l'étude du jeu de tweets brut. Nous utilisons la densité de Kernel pour générer des cartes mensuelles des densités de tweets à partir du jeu de données brut. Une carte des densités de population est créée à partir des données carroyées de l'INSEE³. Les cartes finales sont générées en calculant le ratio entre les densités de tweets mensuelles et les densités de population. La figure 3 représente la distribution spatiale des ratios calculés pour chacun des mois collectés, et met en évidence les espaces présentant des anomalies.

Les zones où la valeur des pixels est comprise entre 0 et 1 correspondent à des espaces où l'activité peut être qualifiée de standard, c'est-à-dire que les densités de tweets émis sont inférieures ou égales à la densité de population. Les zones où ce ratio est supérieur à 1 forment nos zones d'intérêt et désignent les espaces qui concentrent plus de tweets que d'habitants.

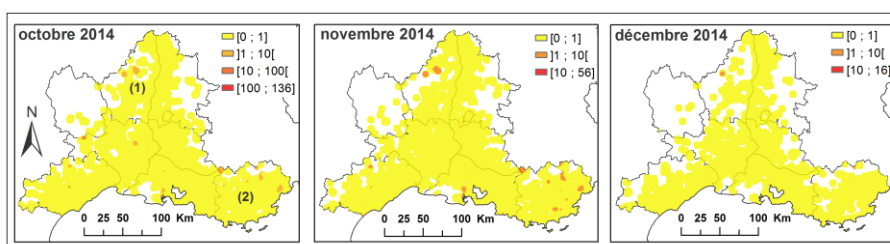


Figure 3. Détection de hotspots mensuels d'activité

³ http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees

Les *hotspots* ainsi détectés sont principalement localisés dans les départements de l'Ardèche (1) et du Var (2), mais absents des agglomérations principales. Nous observons donc un phénomène résultant (B), les anomalies, et recherchons une hypothèse (A) qui pourrait l'expliquer.

3.2.2. Densités mensuelles de tweets de crise (Phase 2)

Nous supposons que l'observation (B) résulte de la survenue d'un événement exceptionnel. Néanmoins, comme (B) apparaît dans des espaces péri-urbains ou ruraux, nous pensons que la perturbation à l'origine de (B) n'est pas liée à un événement d'origine sociale qui aurait la capacité de drainer un flux conséquent de tweets, comme un match de football ou toute autre manifestation. Nous supposons donc que ces anomalies proviennent de perturbations qui peuvent être d'origine hydro-météorologique.

Nous explorons alors les variations spatio-temporelles des densités mensuelles de tweets de crise (figure 4), afin de détecter l'existence éventuelle d'une relation entre les *hotspots* précédemment mis en évidence et les foyers d'émission de tweets de crise.

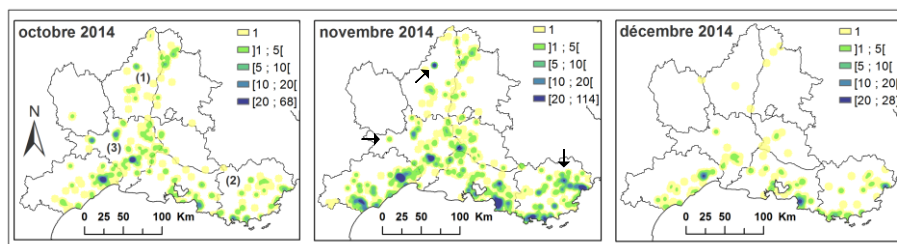


Figure 4. Densités mensuelles de tweets de crise

Sans surprise, les plus fortes densités de tweets sont rassemblées dans les agglomérations régionales et le long du littoral méditerranéen qui concentrent les plus fortes densités de population. Certains foyers d'émission de tweets de crise, qui constituent des motifs répétitifs dans le temps, en particulier dans le nord de l'Ardèche (1), l'ouest du Gard (3) et dans le Var (2) sont néanmoins identifiés en dehors de ces espaces et indiqués par des flèches sur la figure 4.

En comparant ces foyers à ceux mis en évidence lors de l'étape précédente, nous avons découvert des structures communes et choisi d'explorer plus en détail le *hotspot* localisé dans le nord de l'Ardèche afin de vérifier si l'anomalie détectée est bien liée à la manifestation de perturbations météorologiques. En d'autres termes, nous vérifions l'hypothèse selon laquelle des anomalies de *tweeting* permettent d'identifier des lieux particulièrement affectés par les événements hydro-météorologiques extrêmes, et constituent ainsi des marqueurs pertinents.

SAGEO'2016 – Nice, 6-9 décembre 2016

3.2.3. Exploration du hotspot ardéchois (Phase 3)

Nous souhaitons savoir ce qui est précisément arrivé en ce lieu particulier : en explorant ce motif plus en détail, il nous apparaît focalisé sur une commune rurale, le Cheylard, qui compte environ 3 000 habitants. La période étudiée enregistre 3 194 tweets bruts émis par quatre *tweeters*. Trois d'entre eux sont des utilisateurs occasionnels alors que le dernier peut être qualifié de *super tweeter* : il a en effet créé 3 106 tweets sur la période observée, dont 39 tweets de crise. Le foyer mis en évidence provenait donc de la super activité d'un seul utilisateur. Nous ne le considérons pas pour autant comme un *spammer*⁴ : cet individu est en effet le *tweeter* pour lequel nous avons constaté la participation maximale à la création de tweets de crise (à titre comparatif, 90% des personnes enregistrées dans le jeu de tweets de crise ont créé moins de 5 tweets liés aux événements). Pour explorer les habitudes et le comportement de ce *tweeter*, nous avons sélectionné et analysé l'information contenue dans ses tweets de crise afin de les classer par thèmes : quatre ont été identifiés (tableau 2, les thème des tweets sont représentés en gras) :

Tableau 2. Classement thématique et exemples de tweets

Thème du tweet	Exemple de tweet
météo : observation en temps réel	<i>Pluie et gris ici</i>
inondation : le <i>tweeter</i> a été inondé	<i>Là, ça y est c'est l'inondation</i>
alerte : l'Ardèche est en alerte ou vigilance météorologique	<i>noir ici, vigilance orange</i>
information : le tweet est lié à une crise mais ne concerne pas l'environnement local du <i>tweeter</i>	<i>Nouvel accident dans le Gard : un homme retrouvé noyé http://t.co/oBy8tM2X2C via BFMTV</i>

La figure 5 révèle qu'une grande partie de l'information créée reste homogène sur le plan sémantique, et ne fournit, par conséquent, que peu d'information de terrain (qui concerne directement le *tweeter*) utile et pertinente : en effet, lorsqu'il *tweete* pour décrire un événement local, cet utilisateur crée une information relative à la météo, et plus particulièrement à une averse en cours. En outre, signalons que le thème que nous avons nommé *information* dans le tableau 2, qui représente une part non négligeable de l'effectif total de tweets de crise, se réfère à des événements qui ont affecté d'autres territoires, par exemple d'autres départements ou d'autres villes qu'il mentionne dans le tweet, alors que sa géolocalisation indique qu'il est resté dans son quartier de résidence. Nous constatons ainsi un autre problème désigné comme téléprésence asynchrone (Miller, 2007), qui qualifie l'émancipation de la présence physique de l'individu, dans le temps et dans l'espace, pour avoir connaissance d'un événement et s'en faire le témoin indirect.

⁴ Un individu qui envoie une grande quantité de messages inutiles.

Cpcnř ugt 'rgu'vy ggw'f' 2 qmřecrk' u

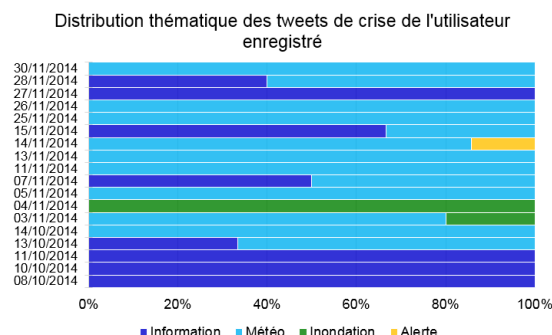


Figure 5. Exploration des tweets de l'utilisateur du Cheylard

Ainsi, la pertinence de la relation que nous cherchions à tester entre la concentration des tweets et la survenue d'un événement hydro-météorologique intense n'est pas complète. L'événement est susceptible d'entraîner une mobilisation locale, signalée par de fortes densités de tweets mais cet effet semble dépendant du comportement des individus et de leur activité sur la plateforme. Nous avons vérifié ce qui se passait en un autre point isolé qui correspond au foyer localisé sur la commune de Valleraugue (ouest du Gard), et qui s'étend jusqu'au Mont Aigoual. Le même phénomène est constaté : un seul *tweeter* crée de l'information de crise et celle-ci reste focalisée sur l'événement local et relative à la sphère sociale (mobilisation des secours, conséquences sur les infrastructures) mais ne fait aucune mention des caractéristiques physiques de l'événement.

L'information de crise d'un utilisateur n'est donc pas forcément exploitable pour qualifier un phénomène physique ou ses réponses sociales observables localement, puisque les tweets peuvent également contenir une information relative à d'autres lieux, parfois très éloignés de celui que l'on observe.

4. Conclusion

Cet article s'est attaché à présenter une démarche de recherche originale que nous avons testée dans l'objectif de problématiser la donnée tweet comme donnée géographique, mais également comme marqueur social d'événements extrêmes. Nous avons ainsi développé une méthodologie réutilisable d'exploitation des tweets, caractérisée par un double objectif : d'une part, en associant déduction et abduction, nous avons testé notre capacité à formuler des hypothèses pour expliquer des liens entre des observations, et d'autre part, nous avons mis en évidence les ambiguïtés du tweet comme marqueur de phénomènes. En vérifiant la correspondance spatiale et temporelle des anomalies de *tweeting* du jeu de tweets bruts et les foyers d'émission de tweets de crise, nous souhaitons savoir si ces anomalies étaient le résultat d'une perturbation d'origine hydro-météorologique et estimer si elles pouvaient constituer un indicateur pertinent. Les résultats ont montré qu'on ne pouvait ni confirmer, ni infirmer cette hypothèse : si une densité anormale de tweets peut être considérée

SAGEO'2016 – Nice, 6-9 décembre 2016

comme un marqueur d'événement, ces tweets ne constituent pas un indicateur fiable pour qualifier la dynamique physique et sociale de l'événement, en raison de trois facteurs : nous avons mis en évidence la forte activité d'individus isolés, et donc une nouvelle problématique comportementale, la téléprésence aux événements ainsi que l'incomplétude de l'information.

Ces résultats nous amènent à considérer les remarques suivantes : il serait préférable d'explorer les *hotspots* d'espaces plus peuplés, comme la vallée de la Nartuby entre Draguignan et Le Muy (Var), qui peuvent offrir un meilleur potentiel d'information diversifiée. Dans un second temps, nous soulevons la question suivante : étant donné le maigre jeu de tweets de crise que nous avons extrait, et dont il resterait à évaluer le degré de complétude, nous pouvons nous demander ce que représente un événement majeur pour un *tweeter* dans le contexte des risques naturels, c'est-à-dire un événement qui est, sur le plan individuel ou collectif, suffisamment important pour devenir un sujet de discussion prépondérant.

Au final, après la mise en œuvre de ces premiers traitements, l'abduction pourrait nous aider à apporter des éléments de réponse à d'autres questions : quels facteurs poussent le *tweeter* à créer une information de crise ? qu'est-ce qui explique la variabilité spatio-temporelle de l'information sémantique et de sa quantité en termes d'effectif ? Néanmoins, à ce stade et avec le peu de données collectées, il est difficile d'affirmer la possibilité de reconstituer l'histoire d'une crise à partir des tweets. Au vu des résultats, nous pensons que le tweet n'a pas la capacité de constituer une donnée autosuffisante et que la construction de connaissances ne doit pas être séparée du contexte de production de l'information, c'est-à-dire qu'elle doit prendre racine sur l'utilisation simultanée des tweets et de données complémentaires, rattachées au contexte physique de l'événement (cumuls pluviométriques, hauteur d'eau des rivières, etc.).

Remerciements.

Le travail présenté dans ce papier s'inscrit dans le cadre du projet ANR MobiClimEx (ANR-12-SENV-0002-01) et du projet MISTRAL-HyMeX.

Les auteurs remercient l'équipe SLIDE du Laboratoire d'Informatique de Grenoble pour la mise à disposition des tweets via leur serveur de collecte.

Bibliographie

- Anderson C. (2008). *The end of theory: The data deluge makes the scientific method obsolete*, <http://www.wired.com/2008/06/pb-theory/>
- Andrienko G., Andrienko N., Bosch H., Ertl T., Fuchs G., Jankowski P., Thom D. (2013). Thematic patterns in georeferenced tweets through space-time visual analytics. *Computing in Science and Engineering* vol. 15, n° 3, p. 72-82.
- Dashti S., Palen L., Heris M. P., Anderson K. M., Anderson T. J., Anderson S. (2014). Supporting disaster reconnaissance with social media data : a design-oriented case study

Српнф угр'ггу'уу ггу'г' 2 qmcecku

- of the 2013 Colorado floods. *Proceedings of the 11th International ISCRAM Conference*, University Park, USA.
- De Longeville B., Smith R. S., Luraschi G. (2009). "OMG, from here, I can see the flames!" : a case use of mining Location Based Social Networks to acquire spatio-temporal data on forest fires. *Proceedings of the 2009 International Workshop on Location Based Social Networks*, Seattle, USA.
- Dittrich A., Lucas C. (2014). Is this Twitter event a disaster ? *Proceedings of the AGILE'2014 International Conference on Geographic Information Science*. Castellón, Spain.
- Erwig M. (2004). Toward Spatio-Temporal Patterns. *Spatio-Temporal Databases*, Berlin Heidelberg, Springer, p. 29-53.
- Hawelka B., Sitko I., Beinat E., Sobolevsky S., Kazakopoulos P., Ratti C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science* vol. 41, n° 3 , p. 260-271.
- Hertfort B., Porto de Albuquerque J., Schelhorn S.-J., Zipf A. (2014). Exploring the geographical relations between social media and flood phenomena to improve situational awareness. *Connecting a digital Europe through Location and Place*, Springer International Publishing, p. 55-71.
- Hoffmann M. (1999). Problems with Peirce's concept of abduction. *Foundations of Science*, vol. 4, n° 3, p. 271-305.
- Kell D. B., Oliver S. G. (2003). Here is the evidence, now what is the hypothesis ? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays*, vol. 26, n° 1, p. 99-105.
- Kitchin R. (2013). Big data and human geography. Opportunities, challenges and risks. *Dialogues in Human Geography*, vol. 3, n° 3, p. 262-267.
- Kounadi O., Lampoltshammer T.J., Groff E., Sitko I., Leitner M. (2015). *Exploring Twitter to analyze the public's reaction patterns to recently reported homicides in London*, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121848>
- Lee Hughes A., Palen L. (2009). Twitter adoption and use in mass convergence and emergency events. *Proceedings of the 6th International ISCRAM Conference*, Gothenburg, Sweden.
- Li L., Goodchild M. F., Xu B. (2013). Spatial, temporal and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, vol. 40, n° 2. p. 61-77
- Lin Y. R., Margolin D., Keegan B., Baronchelli A., Lazer D. (2013). #Bigbirds never die : understanding social dynamics of emergent hashtags. *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, Cambridge, USA.
- Miller H. J. (2007). Placed-Based versus People-Based Geographic Information Science. *Geography Compass*, vol. 10, n° 3, p. 503-535.
- Miller H. J., Goodchild M. F. (2015). Data-driven geography. *GeoJournal*, vol. 80, n° 4, p. 449-461.
- Walters B. B. (2012). An event-based methodology for climate change and human-environment research. *Geografisk Tidsskrift-Danish Journal of Geography*, vol. 112, n° 2, p. 135-143.

Détection automatique de séquences sismiques par analyse spatio-temporelle de catalogues de sismicité

Jenny Trévisan¹, Lionel Matteo¹, Anthony Sladen¹

*I. Univ. Nice Sophia Antipolis – CNRS – IRD – Observatoire de la Côte d’Azur, Géoazur UMR 7329
250 rue Albert Einstein, F-06560 Valbonne, France
trevisan@geoazur.unice.fr*

RESUME. Tous les séismes importants sont généralement accompagnés d’une séquence sismique, c’est à dire d’évènements sismiques voisins dans l’espace et dans le temps. Ces séquences font l’objet de nombreuses études en sismologie, mais leur analyse se fait au cas par cas. Nous proposons ici une méthode de détection automatique de ces séquences à partir de catalogues de sismicité, par analyse spatio-temporelle, en mixant savoir-faire sismologique et savoir-faire géomatique. Cette détection a pour but de quantifier et de caractériser les séquences de manière systématique pour, à terme, améliorer notre compréhension de la physique et des mécanismes de déclenchement de la sismicité.

ABSTRACT. Large earthquakes are usually associated with a seismic sequence, i.e. additional seismic events defined by their proximity in both space and time. Numerous seismological studies are focused on these sequences, but their analysis usually results from case-by-case detections. We propose here to detect seismic sequences automatically from seismicity catalogs, by spatio-temporal analysis, using both seismology and geomatic expertise. This systematic detection aims at quantifying and characterizing the sequences in order to improve our understanding of earthquake physics and triggering mechanisms.

MOTS-CLES : Sismicité, Sismologie, Séisme, Séquence sismique, Analyse spatio-temporelle, Cluster, Triangulation, Delaunay

KEYWORDS: Seismicity, Seismology, Earthquake, Seismic sequence, Space-Time Analysis, Clustering, Delaunay Triangulation

1. Les séquences sismiques : une problématique en Sciences de la Terre

Les techniques d'observation de la Terre ont fait un bond pendant ces dernières décennies. Elles nous permettent de disposer de catalogues de sismicité de référence, sur le monde entier ou sur des zones régionales ou locales, et ce sur des périodes de temps supérieures à 20 ans. Un catalogue de sismicité recense les séismes passés en donnant au minimum, pour chaque séisme, les informations suivantes : la longitude, la latitude, la profondeur, la date et l'heure, et la magnitude qui reflète l'amplitude du séisme.

Ces séismes ne sont pas localisés au hasard. Ils se déclenchent dans les zones où la croûte terrestre accumule trop de contraintes. Quand le seuil de résistance du milieu est atteint, le séisme a lieu. Cette rupture sismique, parce qu'elle génère des ondes qui se propagent dans la croûte terrestre, engendre des déformations élastiques ou inélastiques. D'autres séismes peuvent alors se produire en réponse à ce séisme initial. Ce phénomène engendre la formation d'une séquence sismique, c'est-à-dire d'un ensemble d'événements sismiques liés dans l'espace et dans le temps. Le plus gros événement d'une séquence, en magnitude, est appelé choc principal, les événements antérieurs sont dits précurseurs, tandis que les événements postérieurs sont des répliques.

Ces séquences sismiques sont très étudiées en sismologie, comme le montre Bouchon et al (2013 ; 2016) sur les séquences de précurseurs. Cependant, ces études analysent les séquences de manière ponctuelle, c'est-à-dire manuellement, au cas par cas. Nous cherchons donc ici à détecter les séquences sismiques systématiquement et automatiquement à partir de catalogues de sismicité mondiaux, régionaux, voire locaux.

Nous proposons une méthode de détection qui mixe savoir-faire sismologique et savoir-faire géomatique. Cette méthode, dite par « TimeGeoClusters », amas géographiques et temporels, est une détection spatio-temporelle en 2D des événements sismiques autour d'un choc principal. Elle est basée (1) temporellement, sur une période de temps autour du choc principal, et (2) spatialement, sur le voisinage des séismes environnant ce choc.

2. Les données de référence nécessaires à la détection de séquences sismiques

Nous avons besoin d'identifier deux types d'événements sismiques : (1) les chocs principaux potentiels et (2) les précurseurs ou répliques composant les séquences sismiques. Pour ces deux types d'événements, certaines propriétés sont nécessaires en vue de la détection des séquences.

2.1. Les propriétés des événements sismiques

Voici la liste des propriétés qui servent de point de départ à notre détection de séquences sismiques. Elle aidera au choix du ou des catalogues de sismicité à utiliser.

SAGEO'2016 – Nice, 6-9 décembre 2016

2.1.1. La magnitude

Les chocs principaux potentiels sont des séismes de forte magnitude. Le terme « forte magnitude » est relatif, dans la mesure où il dépend de la zone d'étude envisagée. Dans les zones de subduction (cf. chapitre 4), les fortes magnitudes sont considérées comme supérieures à 7, alors qu'en France, une magnitude de 4.5 est déjà considérée comme forte. La magnitude est donc une propriété nécessaire pour identifier les chocs principaux potentiels, à partir d'une valeur seuil donnée.

Disposer de la magnitude la plus précise possible sur chaque évènement (au moins pour le choc principal) est un plus, car la magnitude sera utilisée par la détection comme point de départ pour évaluer l'emprise spatiale de la sismicité d'une séquence (cf. paragraphe 3.2.2). Nous conseillons d'utiliser, de préférence, la magnitude de type M_w , c'est-à-dire la magnitude de moment, qui est calculée à partir des basses fréquences des ondes émises par le séisme et qui qualifie précisément l'énergie dégagée. Comparativement aux autres types de magnitudes (magnitude locale M_l , magnitude des ondes de surface $M_s...$), qui ont tendance à sous-estimer la magnitude, M_w est en effet moins soumise aux phénomènes de saturation lors de l'enregistrement des ondes. Tous les catalogues de sismicité ne fournissent pas cette magnitude, car elle reste plus difficile à calculer. Notre détection supposera cependant que la magnitude utilisée est une M_w .

2.1.2. Le temps de référence

Outre le soin à apporter à la magnitude, il est également important de disposer du temps de référence de chaque évènement sismique, appelé encore temps origine. Les catalogues de sismicité donnent un temps en date/heure/minutes/secondes. Suivant les stations sismologiques et les traitements utilisés pour calculer ce paramètre, un même évènement dans deux catalogues distincts peut avoir un temps origine différent à, en moyenne, 10 secondes près.

2.1.3. La localisation

Qui dit détection spatiale dit localisation : la position de l'hypocentre ou de l'épicentre de chaque évènement sismique est donc nécessaire. L'hypocentre correspond au foyer de la rupture sismique sur la faille (dimension 3D), alors que l'épicentre est la projection de l'hypocentre à la surface de la terre (dimension 2D). Les évènements sismiques continentaux, c'est-à-dire à terre, sont généralement mieux localisés (incertitude 2D de 1 à 5 km) que ceux en mer (incertitude 2D de 5 à 20 km). En effet, ils sont mieux contraints grâce à la proximité des stations sismologiques permettant de déduire leur position (stations essentiellement installées à terre). Notre détection proposera donc un paramètre d'incertitude sur la localisation des évènements.

2.2. Un appariement éventuel de catalogues

Les propriétés nécessaires à la détection peuvent éventuellement conduire à utiliser non par un, mais deux catalogues de sismicité : un pour les chocs principaux

Détection automatique de séquences sismiques

et un pour les autres évènements. Dans ce cas, il est indispensable de pouvoir naviguer facilement entre ces deux catalogues.

Nous avons donc développé un code d'appariement de catalogues de sismicité, avec une méthode standard spatio-temporelle à conservation de caractéristiques : chaque choc principal du premier catalogue a été apparié, via son épigcentre, avec son plus proche évènement sismique dans le second catalogue, en temps, en localisation et en magnitude. Nous utilisons l'épigcentre (2D) pour l'appariement spatial, car le paramètre de profondeur (Z) sur l'hypocentre des séismes n'est pas suffisamment précis pour envisager un appariement 3D.

Seul un appariement unidirectionnel « Catalogue Chocs Principaux → Catalogue Evènements Sismiques » est calculé ici, car le but final n'est pas d'optimiser l'appariement mais de disposer d'un nombre suffisant de chocs principaux appariés pour détecter les séquences sismiques. Notre méthode d'appariement permet cependant de faire de l'appariement bidirectionnel, « Catalogue Chocs Principaux ↔ Catalogue Evènements Sismiques », et même d'apparier plus de deux catalogues entre eux. Notons que les chocs principaux qui n'ont pas été appariés automatiquement (ou dont l'appariement est ambigu) peuvent être revus manuellement. Si la reprise manuelle ne trouve pas de solution, ou si on décide de ne pas faire de reprise manuelle, alors le choc principal n'est pas pris en compte pour la suite de la détection.

Comme nous le verrons en pratique dans le paragraphe 4.1.4, cet appariement simple au plus proche voisin ayant des caractéristiques similaires permet de traiter la grande majorité des cas.

3. La détection de séquences sismiques par TimeGeoClusters

A partir des données de référence ainsi préparées, nous pouvons envisager la détection effective de séquences sismiques. Pour chaque choc principal, nous procédons à une détection (1) temporelle dans un laps de temps donné, puis (2) spatiale en mixant connaissances sismologiques et connaissances géomatiques. Les amas de séismes ainsi récupérés autour du choc principal forment les TimeGeoClusters.

3.1. La détection temporelle

La détection temporelle récupère les évènements du catalogue d'évènements, à moins d'une durée d_t du temps origine du choc principal. Les évènements ainsi récupérés peuvent être situés avant (précurseurs) et/ou après le choc (répliques). La durée d_t peut être fixe pour tous les chocs principaux considérés : 48 heures, 30 jours... en fonction de ce qu'on souhaite étudier au final. Elle peut également être variable, dépendant alors de chaque choc. Auquel cas, elle doit être stockée comme propriété du choc principal.

SAGEO'2016 – Nice, 6-9 décembre 2016

3.2. Une détection spatiale multi-passes

Il est actuellement difficile de relier les informations sur le milieu de propagation à la répartition de la sismicité elle-même. C'est justement l'un des enjeux de notre détection des séquences sismiques : leur étendue, leur densité sont-elles liées au milieu de propagation ? Il est important de noter que la répartition des précurseurs et des répliques dans une séquence sismique peut être plus ou moins étendue et de densité non homogène, de forme variée, et qu'il peut même y avoir plusieurs clusters de sismicité.

3.2.1. Cluster adapté à l'environnement via Delaunay

Pour la création des clusters spatiaux, il nous est donc impossible de connaître a priori soit un paramètre de distance ou de densité, soit le nombre de clusters à trouver, ou encore si les clusters seront complexes ou non en forme. Ce postulat élimine, pour la détection spatiale des séquences, les algorithmes de clusterings de type « densité » (ex : DBSCAN d'Ester et al., 1996), « partitionnement » (ex : K-means, de MacQueen, 1967) ou encore « hiérarchique » (ex : BIRCH, de Zhang et al., 1996).

Nous avons donc besoin d'une méthode qui s'adapte d'elle-même à la répartition sismique, encore imprévisible, survenant autour du choc principal. De là est venue l'idée de s'appuyer sur la notion de voisinage et nous nous sommes naturellement tournés vers les algorithmes de clustering utilisant la triangulation de Delaunay (Delaunay, 1934), méthode reconnue pour évaluer le voisinage. Ces algorithmes, comme AUTOCLUST d'Estivill-Castro & Lee (2002) ou encore ASCDT de Deng et al. (2011), forment des clusters par suppression de certaines branches de la triangulation de Delaunay du jeu de points donné, suivant des critères statistiques élaborés.

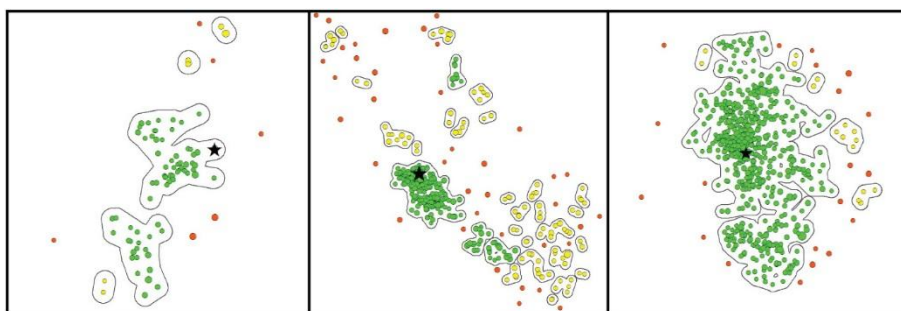


FIGURE 1. Tests de clustering sur des événements sismiques, via Delaunay coupé par la moyenne des branches internes. En rouge : les événements isolés, en jaune : les événements dans des petits clusters, en vert : les événements dans les clusters retenus comme étant significatifs. L'étoile noire représente le choc principal.

Détection automatique de séquences sismiques

Nous avons repris ce principe en choisissant de simplifier le critère de suppression des branches Delaunay : nous supprimons les branches supérieures à la moyenne des longueurs des branches internes de la triangulation. L'utilisation de la moyenne a un sens dans la mesure où le jeu de points considérés correspond aux événements sismiques situés autour du choc principal (voir 3.2.2). Parmi ces événements, certains font partie de la sismicité ambiante, c'est-à-dire la sismicité habituelle à cet endroit, en dehors de la séquence sismique. La sismicité ambiante étant généralement plus diffuse que celle d'une séquence sismique, l'utilisation de ces points va permettre d'augmenter la moyenne des longueurs des branches Delaunay et donc d'identifier les clusters plus resserrés en distance. Des tests empiriques concluants, montrés en Figure 1, nous ont permis de vérifier ce choix.

3.2.2. Recherche des précurseurs/répliques potentiels à partir de la longueur de rupture sismique théorique

Reste à résoudre la question de la récupération préalable des événements autour du choc principal, ceux qui vont servir de base à ce clustering spatial : comment quantifier « autour » ? Il se trouve que nous disposons de la magnitude M_w du choc principal. Or la loi d'échelle la plus utilisée, concernant la longueur de rupture sismique théorique associée à choc principal, est justement fonction de ce paramètre. C'est la loi de Wells et Coppersmith (1994). Elle permet de déduire la longueur de rupture en surface $L_{W\&C}$ (en km) par l'équation (1).

$$L_{W\&C} = 10^{(0.58M_w)-2.42} \quad (1)$$

Nous décidons de prendre cette longueur comme base pour la recherche spatiale. La rupture pouvant se propager d'un côté ou de l'autre de l'épicentre, $L_{W\&C}$ correspond donc à un rayon de recherche autour du choc principal. Pour s'affranchir des incertitudes, de la loi de Wells et Coppersmith (1994) d'une part, et de la localisation des épicentres d'autre part, nous augmentons ce rayon de recherche à la valeur $R_{rech_clusters}$ de l'équation (2), c'est-à-dire à 110% de $L_{W\&C}$ auquel est ajoutée la précision δ_{loc} de la localisation des événements. L'idée est de rechercher les clusters d'événements sismiques, les « TimeGeoClusters », situés à cheval sur ce rayon de recherche $R_{rech_clusters}$.

$$R_{rech_clusters} = \delta_{loc} + \frac{(100+10)L_{W\&C}}{100} \quad (2)$$

Cependant, pour constituer ces TimeGeoClusters, nous devons récupérer les événements potentiels dans une zone de recherche encore plus large que ce dernier rayon pour être sûr de ne pas exclure une séquence qui serait particulièrement étalée dans l'espace, et pour prendre en compte des événements de la sismicité ambiante, en vue du clustering (voir 3.2.1). A la suite de plusieurs tests empiriques, nous avons estimé pertinent de fixer cette zone de recherche des événements à deux fois la zone de recherche des clusters, soit un rayon de recherche R_{rech_events} (équation 3) de plus de 2 fois le rayon théorique de rupture de Wells et Coppersmith (1994).

$$R_{rech_events} = 2R_{rech_clusters} \quad (3)$$

SAGEO'2016 – Nice, 6-9 décembre 2016

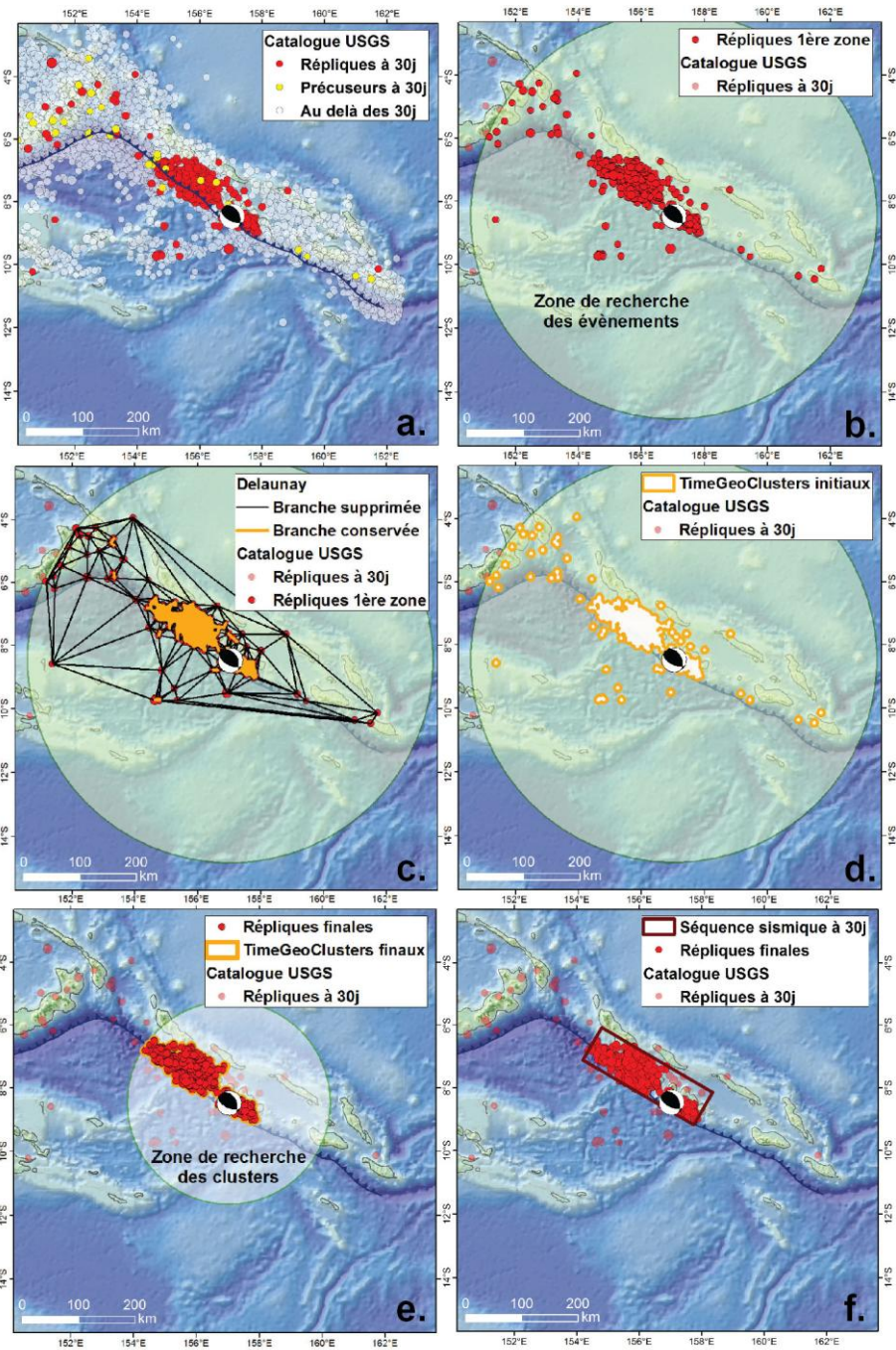


FIGURE 2. Méthode de détection sur le séisme de Solomon, 01/04/2007 (cf. 3.3)

3.3. Récapitulatif de la détection par TimeGeoClusters

En résumé, les grandes étapes de cette détection par TimeGeoClusters sont :

- Fig.2a : récupération des événements à une durée d_t du choc principal ;
- Fig.2b : extraction de ceux à moins de R_{rech_events} du choc principal ;
- Fig.2c : triangulation de Delaunay ;
- calcul de la longueur moyenne des branches Delaunay internes et suppression des branches Delaunay supérieures à cette moyenne : les événements encore reliés par des branches Delaunay forment les TimeGeoClusters initiaux ;
- Fig.2d : création d'une emprise pour chaque TimeGeoCluster initial ;
- Fig.2e : suppression des TimeGeoClusters initiaux dont l'emprise n'intersecte pas la zone de recherche de rayon $R_{rech_clusters}$ centrée sur le choc principal.

Les événements des TimeGeoClusters ainsi retenus sont les composants de la séquence sismique correspondant au choc principal considéré. Ils sont alors filtrés pour ne retenir que le TimeGeoCluster principal, à savoir le plus proche du choc principal, contenant le plus d'événements (Fig.2f) ou, en cas d'équivalence, avec la densité la plus importante (Matteo, 2016).

4. Application sur les séismes de subduction : les séquences de répliques

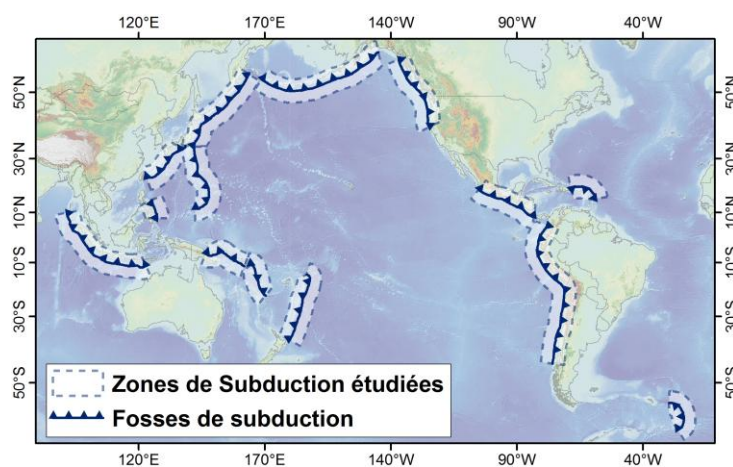


FIGURE 3. les zones de subductions étudiées, représentées en projection Mercator centrée sur le méridien 180.

Les zones de subduction (Fig. 3), là où une plaque océanique rencontre une plaque continentale, sont à l'origine des plus gros séismes mondiaux, encore appelés mega-séismes (comme le séisme de Tohoku-Oki au Japon en 2011, M_w 9.1). Elles sont par conséquent très soumises à l'aléa tsunami. Nous avons focalisé notre

SAGEO'2016 – Nice, 6-9 décembre 2016

première étude applicative sur la détection des séquences de répliques sismiques associées à ces séismes de subduction. Nous montrons ci-après les résultats préliminaires qui ouvrent des perspectives d'approfondissement pour la recherche en sismologie.

4.1. Les catalogues de sismicité et le paramétrage utilisés

Nous avons tout d'abord choisi avec soin les catalogues de sismicité et le paramétrage nécessaires pour la détection.

4.1.1. Le catalogue des chocs principaux potentiels

Le catalogue américain Global Centroid Moment Tensor¹, GCMT, recense les évènements sismiques mondiaux avec une magnitude de complétude à 5, c'est-à-dire la magnitude à partir de laquelle tous les séismes sont systématiquement recensés dans le catalogue. Outre la magnitude de moment M_w , le catalogue GCMT fournit également des informations sur le mécanisme au foyer du séisme (ou CMT). Le CMT explique comment s'est passée la rupture sur la faille : normale, inverse, ou décrochante. Ces informations nous permettent de filtrer le catalogue pour se restreindre aux séismes majeurs des zones de subduction, à savoir les séismes en faille inverse sur la plaque continentale. Nous l'avons donc adopté comme catalogue pour les chocs principaux potentiels.

Au départ de la détection, le nombre de séismes GCMT retenus comme candidats potentiels « choc principal » est de 7407 évènements, après filtrage en magnitude ≥ 5 , pour des mécanismes inverses sur la plaque continentale de subduction, et sur une période de temps choisie entre 1990 et 2014. Ce chiffre nous donne un potentiel d'analyse conséquent.

4.1.2. Le catalogue des évènements sismiques précurseurs ou répliques

Le catalogue d'évènements sismiques doit être plus complet pour détecter les séquences sismiques associées aux chocs principaux. Dans notre cas, il doit donc avoir une magnitude de complétude inférieure à 5, magnitude minimale des chocs GCMT considérés. Deux catalogues de sismicité mondiaux ont retenu notre attention : (a) le catalogue américain U.S. Geological Survey, USGS², dont la magnitude de complétude est autour de 4 et (b) le catalogue britannique International Seismological Centre³, ISC, dont la magnitude de complétude est autour de 3. Nous avons utilisé le catalogue USGS pour valider notre méthode (moins d'évènements donc temps de traitements plus rapides). Le catalogue ISC nous servira ensuite à caractériser plus précisément les séquences déduites. Ces catalogues sont filtrés à magnitude 3, sur une période de temps de 1990-2014 pour l'USGS, et 1990-2012 pour l'ISC (disponibilité du catalogue), soit

¹ 1. <http://www.globalcmt.org/>

² 2. <http://earthquake.usgs.gov/earthquakes/>

³ 3. <http://www.isc.ac.uk/iscbulletin/>

Détection automatique de séquences sismiques

235 000 événements sur les zones de subduction pour l'USGS, contre 609 000 pour l'ISC. Dans la suite, nous donnerons des exemples sur les séquences USGS. Les premiers résultats obtenus avec l'ISC sont similaires à ceux obtenus avec l'USGS.

Nous aurions pu utiliser directement l'USGS comme catalogue de référence pour les chocs principaux potentiels, mais ce catalogue (tout comme l'ISC) ne fournit pas forcément de magnitudes M_w , et ne donne pas d'indications sur les mécanismes au foyer, nécessaires à notre filtrage des séismes de subduction.

4.1.3. Incertitudes de localisations et géo-référencement

Dans le cadre des zones de subduction, nous travaillons sur des données couvrant le monde entier. Il est difficile dans ces conditions d'utiliser un système de coordonnées projeté pour traiter toutes les données simultanément. Nos données sont donc en WGS84 et tous les calculs de distance sont menés en utilisant la distance géodésique, c'est-à-dire la longueur de la ligne la plus courte entre deux points quelconques de la surface de la terre sur l'ellipsoïde.

De plus, les zones de subduction étant à cheval sur la terre et la mer, nous avons pris une incertitude moyenne de localisation des événements de 10 km.

4.1.4. Appariement du GCMT avec l'USGS

Comme nous utilisons deux catalogues différents pour la détection des séquences, nous avons fait tourner le code d'appariement pour trouver une correspondance entre les chocs principaux potentiels GCMT et leurs homologues USGS, avec un écart maximal très tolérant de 30 secondes, 100 km et 1 de magnitude. Sur la période 1990-2014, le nombre de chocs principaux GCMT potentiels appariés avec l'USGS est alors de 7306, soit supérieur à 98% : on reste toujours dans des ordres de grandeurs intéressants pour des analyses statistiques. La différence avec les 7407 GCMT de départ est essentiellement due aux problèmes d'appariement aux bords des zones de subduction, événements qui ne nous intéressent pas car trop éloignés de la fosse de subduction.

4.2. Les séquences sismiques détectées

Une fois les données préparées, nous lançons la détection des séquences. Rappelons que seules les séquences de répliques, de magnitude supérieure ou égale à 3, sont calculées ici. Les précurseurs n'ont pas été pris en compte pour cette première étude sur la subduction.

Voici deux exemples de résultats préliminaires à approfondir.

4.2.1. Comparaison globale des séquences détectées sur des laps de temps différents après le choc principal

Plusieurs laps de temps ont été testés pour détecter les séquences de répliques :

SAGEO'2016 – Nice, 6-9 décembre 2016

– « temps court », sur une durée fixe de 48 heures après le choc principal, dans le but de vérifier par la suite la longueur de la rupture co-sismique en surface (Well et Coppersmith, 1994), qui se mesure sur les 24 à 48h après le choc principal ;

– « temps moyen », sur une durée fixe de 30 jours après le choc principal, pour regarder l'évolution de la propagation des répliques dans le temps en comparaison avec les 48h ;

– « temps variable », sur une durée basée sur la loi d'Omori-Utsu (Omori, 1894 ; Utsu, 1961). Cette loi prédit le temps de décroissance du nombre de répliques à la suite du choc principal, en fonction de la sismicité ambiante et du nombre de répliques dans les 24 heures.

Comme le montre la Figure 4, environ deux-tiers des séismes de subduction de magnitude M_w supérieure ou égale à 5 ne semblent pas avoir de répliques de magnitude supérieure à 3, et ce quelle que soit la durée de répliques considérée.

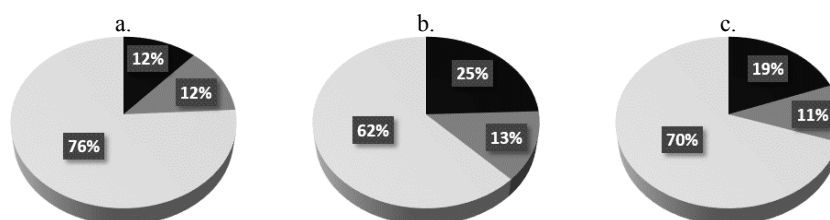


FIGURE 4. Pourcentage du nombre de séquences sismiques des répliques USGS de magnitude supérieure ou égale à 3, détectées à partir des 7306 chocs principaux GCMT potentiels. a. Résultats à 48h. b. Résultats à 30 jours. c. Résultats à durée variable Omori-Utsu. En gris clair, aucune séquence n'est détectée. En gris foncé, les séquences détectées ont moins de 10 répliques. En noir, les séquences détectées ont plus de 10 répliques.

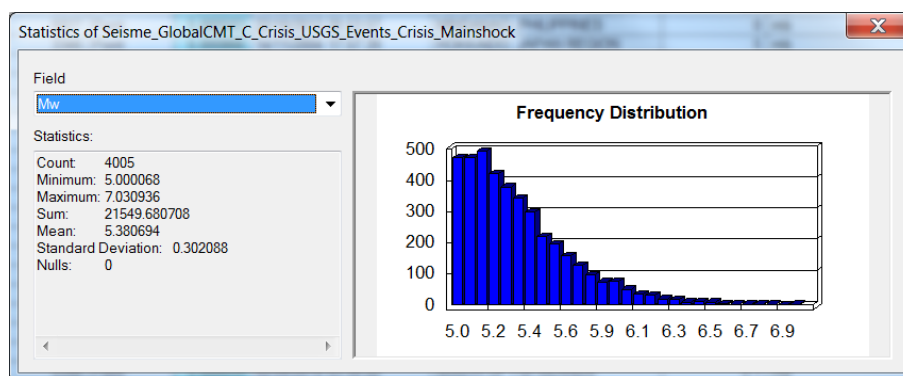


FIGURE 5. Fréquence des magnitudes M_w des chocs principaux n'ayant pas de séquences sismiques USGS associées à 30 jours.

Détection automatique de séquences sismiques

La loi de Bath (Bath M., 1965) explique que, statistiquement, la plus importante réplique a une magnitude inférieure de 1.2 par rapport à celle du choc principal. Comme le catalogue USGS d'événements n'est complet qu'à partir de la magnitude 4, on peut supposer que les séquences de chocs 5 à 5.5 sont effectivement sous-estimées par notre détection. Cependant, comme le montre la Figure 5, un nombre significatif (34%) de chocs supérieurs à 5.5 n'aurait pas de séquence sismique associée à 30 jours, ce qui contredit la loi de Bath. La question du pourquoi est posée...

4.2.2. Zoom sur la séquence sismique de Maule

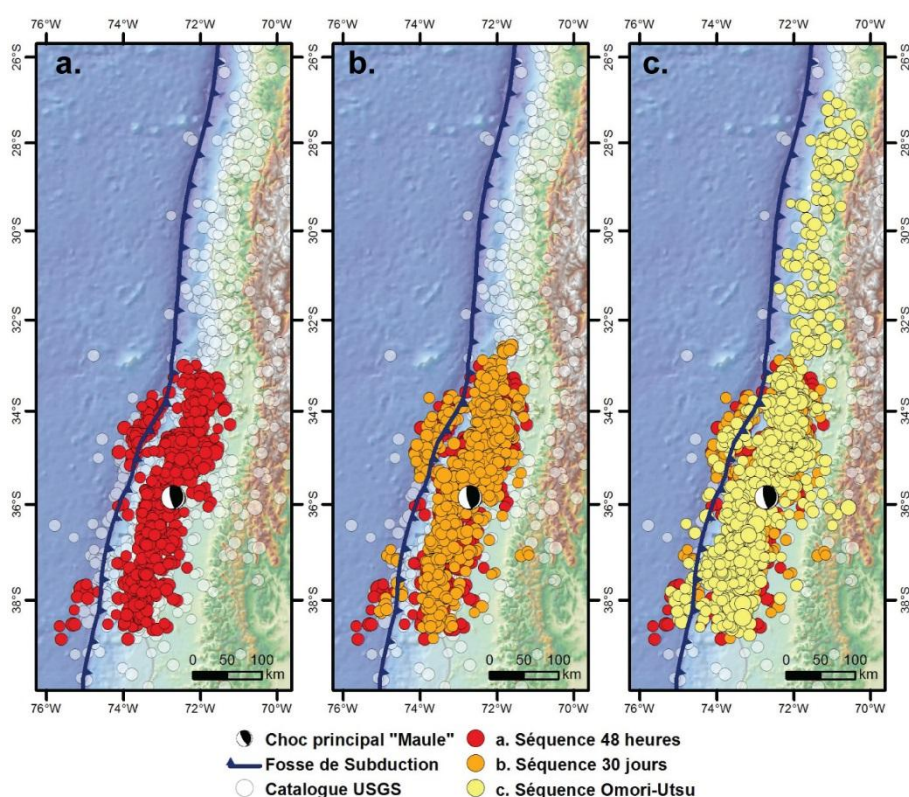


FIGURE 6. Evolution dans le temps de la séquence sismique du séisme de Maule, 27 février 2010. a. séquence à 48 heures, b. séquence à 30 jours, et c. séquence en fonction du délai Omori-Utsu de décroissance des répliques (soit 932 jours ici). Séismes représentés proportionnellement à leur magnitude.

Le 27 février 2010, un séisme de subduction de magnitude M_w 8.8 a secoué la région de Maule au Chili, avec d'importants dégâts matériels et de nombreuses

SAGEO'2016 – Nice, 6-9 décembre 2016

victimes. Nous montrons, dans la Figure 6, une cartographie dans le temps de la séquence sismique de Maule, issue de notre détection automatique.

On constate une légère expansion de la zone de failles réactivée entre 48 heures et 30 jours. Sur le plus long terme, la zone réactivée est uniquement au Nord.

Disposer des séquences sismiques dans le temps nous permet, par exemple, de quantifier l'expansion, et donc d'en déduire des caractéristiques, à la fois propre au choc principal (comme ici pour Maule), mais également systématiques puisqu'on peut travailler sur un nombre conséquent de séquences.

5. Vers de nombreuses perspectives d'études sismologiques

La méthode que nous avons proposée détecte automatiquement, à partir de catalogues de sismicité, les séquences sismiques associées à des chocs principaux, en fonction d'une durée donnée. La prise en compte de la dimension 3D lors de la création des clusters de sismicité est en cours de développement. L'utilisation de la triangulation de Delaunay par l'intermédiaire d'un TIN (Triangulated Integrated Network) permet en effet une adaptation rapide de notre méthode pour la 3D, et nous aidera à détecter des séquences en profondeur. De plus, nous comptons tester la méthode ASCDT complète de Deng et al. (2011) pour détecter les séquences sismiques sans choc principal, les swarms. Parce que cette méthode utilise des critères statistiques permettant de prendre en compte la dimension globale et la dimension locale de la répartition (multi-niveaux), tout en permettant des clusters de densité non homogène, elle nous paraît adaptée pour les swarms.

Le but sismologique final de cette étude est de pouvoir caractériser les séquences sismiques afin de mieux comprendre leur comportement. Leur détection systématique et homogène, sur des catalogues à petite échelle, va nous permettre de disposer d'échantillons conséquents caractérisés avec de nouvelles observables (Fig. 7), comme la directivité, la longueur de rupture, ou encore

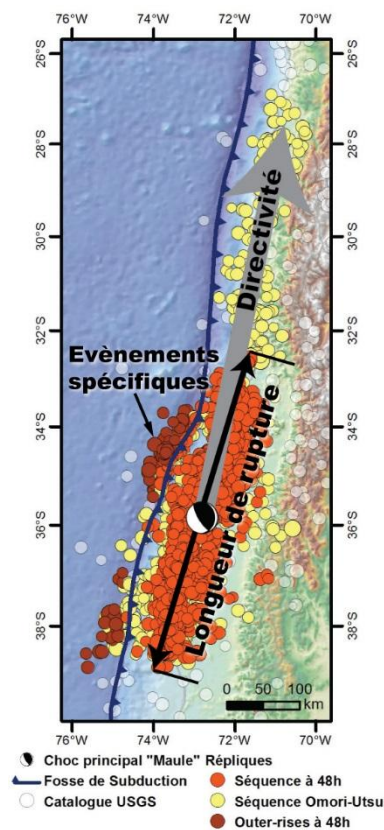


FIGURE 7. Possibilités d'études sismologiques découlant de la détection et de la caractérisation des séquences sismiques.

Détection automatique de séquences sismiques

la présence d'évènements spécifiques comme les précurseurs ou encore les outerrises (Christensen et Ruff, 1988). Nous pourrions ainsi envisager des cycles d'analyses descriptive, exploratoire et si possible prédictive via les méthodes statistiques du Big Data Mining.

Enfin, l'utilisation de catalogues différents sur une même zone nous permettra également de faire des comparatifs pour renforcer les potentielles conclusions à venir. De nombreuses possibilités sont ouvertes.

Remerciements :

Nous remercions Sébastien Mustière (IGN-COGIT) pour ses commentaires avisés

Compléments techniques :

Parce qu'historiquement notre laboratoire de recherche est doté de licences ArcGIS advanced, la méthode décrite ci-dessus a été développée en langage python, avec le package ArcPy, à partir de données stockées en geodatabase ESRI.

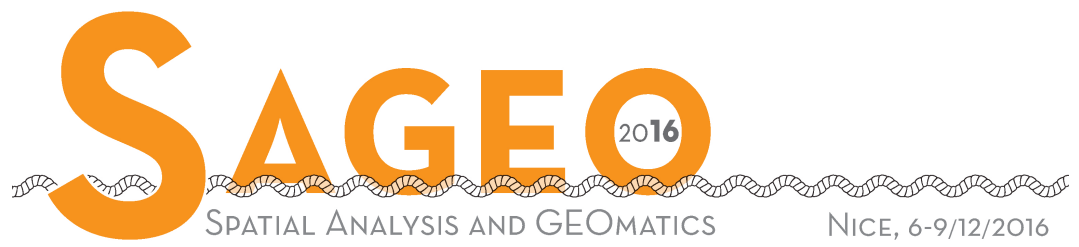
Bibliographie

- Bath M. (1965). Lateral inhomogeneities of the upper mantle. *Tectonophysics*, vol. 2, p. 483-514.
- Bouchon M. et al. (2013). The long precursor phase of most large interplate earthquake. *Nature Geoscience*, vol. 6, p. 299-302.
- Bouchon M. et al. (2016). Potential slab deformation and plunge prior to the Tohoku, Iquique and Maule earthquakes. *Nature Geoscience*, online first.
- Christensen, D. H. et Ruff, L. J. (1988). Seismic coupling and outer rise earthquakes. *J. Geophys. Res. Solid Earth 1978–2012* vol. 93, p. 13421–13444.
- Delaunay B. (1934). Sur la sphère vide. *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk*, vol. 7, p. 793-800.
- Deng M. et al. (2011). An adaptive spatial clustering algorithm based on Delaunay triangulation. *Computers, Environment and Urban Systems*, vol. 35, n° 4, p. 320-332
- Ester, M., Kriegel, H.-P., Sander, J., et Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with Noise. *In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining 1996*, p. 226–231, Portland, OR.
- Estivill-Castro V. et Lee I. (2002), Argument free clustering for large spatial point-data sets, *Computers, Environment and Urban Systems*, vol. 26, p. 315–334.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *In proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability 1967*, University of California Press, p. 281–297, California.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Matteo L. (2016). *Estimation de la loi d'échelle L-Mw par analyse automatique des catalogues mondiaux de sismicité*. Mémoire de Master 2 en Sciences de la Terre, Univ. Nice Sophia Antipolis, UMR Géoazur.
- Ōmori F. (1894). On aftershocks. *Rep. Imp. Earthquake Invest. Comm.*, vol. 2, p. 103 -109.
- Utsu T. (1961). A statistical study on the occurrence of aftershocks. *Geophysical Magazine*, vol. 30, n° 4.
- Wells D.L. et Coppersmith K.J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, vol. 84, n° 4, p. 974-1002.
- Zhang, T., Ramakrishnan, R., et Livny, M. (1996). BIRCH: An efficient data clustering method for very large databases. *In proceedings of the ACM SIGMOD International Conference on Management of Data 1996*, p. 103–114, Canada, Montreal.

[ISBN : 2-910545-11-3]



Session 6

Environnement, traitement d'image, résolution spatiale

Génération d'environnements artificiels pour la simulation spatiale d'arboviroses en milieu urbain : application à la dengue et au virus Zika

Misslin R.¹, Daudé É.¹

1. UMR-6266 IDEES, Université de Rouen
5 rue Thomas Beckett, 76821CEDEX Mont-Saint-Aignan, France
renaud.misslin1@univ-rouen.fr, eric.daude@univ-rouen.fr

RESUME. *Aedes aegypti* est le principal vecteur de la dengue et du virus Zika. Avec la recherche vaccinale, le contrôle vectoriel est un levier important de lutte contre ces maladies. Étudier les relations dynamiques et non-linéaires entre l'environnement et le vecteur permet de guider les stratégies de surveillance et de contrôle. L'objectif de MODE (Model of Dynamical Environment) est d'identifier, par la simulation, les contextes environnementaux qui influencent la distribution spatiotemporelle d'*Aedes aegypti* en milieu urbain. Cet article présente (1) le concept d'un « habitat basé sur les ressources » et son adaptation à *Aedes aegypti*, (2) les méthodes génériques d'estimation des facteurs environnementaux de cet habitat et (3) les dynamiques de ces facteurs. Ces concepts et méthodes sont illustrés en (4) par une application sur Bangkok (Thaïlande). Nous concluons par une discussion sur l'apport d'une modélisation d'environnements dynamiques pour l'étude des arboviroses.

ABSTRACT. *Aedes aegypti* is the main vector of dengue and Zika virus. Along with vaccine research, vector control is a major lever for fighting these diseases. Studying dynamical and non-linear relationships between the environment and vectors provides guidance for monitoring and control strategies. MODE's (Model of Dynamical Environment) objective is to identify environmental dynamics that contribute to the spatiotemporal distribution of *Aedes aegypti* in urban areas. This paper presents (1) the resource-based habitat concept and its application to *Aedes aegypti*, (2) the generic methods for estimating the environmental factor of this habitat and (3) the factors dynamics. These concepts and methods are illustrated in (4) with an application on Bangkok, Thailand. We conclude with a discussion on the contribution of environmental dynamics modelling for the study of arboviruses.

MOTS-CLES : *maladies à transmission vectorielle, resource-based habitat concept, biogéographie, simulation spatiale, SIG, télédétection, simulation à base d'agents*

KEYWORDS: *vector-borne disease, resource-based habitat concept, biogeography, spatial simulation, GIS, remote sensing, agent-based simulation*

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

Aedes aegypti est un moustique présent dans les zones tropicales et intertropicales d'Afrique, d'Amérique et d'Asie. Il est le vecteur principal de la dengue, du virus Zika et, dans une moindre mesure, du Chikungunya et de la fièvre jaune. Sa nature synanthropique (Rodhain et Rosen, 1997) font des villes les principales zones de propagation de ces maladies. Constat accablant de l'OMS (2016), la propagation de ces virus, notamment de Zika, est le résultat de l'abandon depuis les années 1970 des politiques de contrôle des moustiques dans de nombreux pays¹. Si à moyen terme les vaccins contribueront au recul de ces maladies, les mesures anti-vectorielles restent d'ici là le moyen de contrôle principal. Celles-ci consistent à agir sur l'environnement dans le but (a) d'empêcher le développement des moustiques (nettoyage, drainage et destruction des gîtes larvaires potentiels), (b) de réduire les populations (traitements chimiques des gîtes, fumigation, introduction de prédateurs ou de mâles stériles) et (c) de limiter les interactions entre hôtes et vecteurs (utilisation de moustiquaires). L'amélioration des moyens de surveillance des aires de distribution des vecteurs en milieu urbain et péri-urbain doit permettre d'optimiser l'utilisation de ces moyens dans le temps et l'espace (Gubler, 2002).

Les facteurs qui influencent le développement de populations d'*Ae. aegypti* sont inégalement distribués dans les espaces urbains et leur présence évolue dans le temps. La variété de combinaisons de ces facteurs rend alors difficile la mise en évidence de causalités simples qui permettraient d'expliquer et d'estimer les densités vectorielles. Les précipitations, par exemple, sont relativement homogènes en milieu urbain. De ce fait, elles ne permettent pas d'expliquer la présence de *hotspots* de moustiques, marqueurs d'une distribution spatiale hétérogène.

Devant la difficulté de réaliser sur le terrain des mesures de cette diversité et l'incapacité des expériences de laboratoire à la reproduire, la simulation informatique, plus particulièrement la simulation à base d'agents, apparaît comme une alternative intéressante à ces obstacles. Notre objectif est de modéliser le moustique *Aedes aegypti* dans un environnement réaliste pour étudier, par le biais de la simulation, divers scénarios telles que les conditions de croissance d'une population de moustiques ou les effets de diverses techniques de lutte anti-vectorielle. Plusieurs modèles à base d'agents ont été développés pour explorer le comportement du moustique *Aedes* (Almeida *et al.*, 2010 ; Cummins *et al.*, 2012). Dans ces modèles les relations vecteur-environnement sont explorées dans un contexte micro-local qui ne permet pas la généralisation directe des connaissances à l'échelle urbaine. Or, ce sont ces dernières qui permettent de guider les campagnes de lutte anti-vectorielle dans la ville. D'autres auteurs proposent une approche à base d'agents où les moustiques, les hôtes et l'environnement sont modélisés à résolution fine et simulés à l'échelle urbaine (Daudé *et al.*, 2015 ; Karl *et al.*, 2014). Le modèle MOMA (Maneerat et Daudé, 2016a), développé en parallèle du modèle présenté

¹ <http://who.int/dg/speeches/2016/wha-69/en/>

SAGEO'2016 – Nice, 6-9 décembre 2016

dans le cadre de cet article, intègre par exemple un environnement hétérogène et dynamique, construit à partir de données collectées sur le terrain dans un quartier de Delhi (Inde) (Maneerat et Daudé, 2016b). La collecte de données est cependant une contrainte si l'on veut couvrir une vaste zone ou appliquer le modèle sur un autre territoire. C'est là la limite que MODE (*Model of Dynamical Environment*) propose de pallier : son objectif est de proposer un cadre de modélisation pour générer des environnements dynamiques, spatialement représentatifs des zones d'études.

MODE est développé dans le cadre du projet DENFREE (FP7). Sa conception implique des épidémiologistes et des entomologistes de l'Institut Pasteur intégrés au projet. Il s'agit d'un modèle générique, transposable d'une zone urbaine à une autre, tout en étant spécialisé pour la modélisation d'*Ae. aegypti*. Ce modèle a été conçu comme le support environnemental dynamique avec lequel interagissent des moustiques et des hôtes. Il peut également être utilisé pour produire des cartes détaillées des niches écologiques potentielles d'*Ae. aegypti* à différentes périodes de l'année, et ainsi orienter les opérations de lutte anti-vectorielle (ex : suivi plus régulier des populations de moustiques ou mise en place de systèmes de lutte anti-vectorielle participative dans les espaces les plus attractifs pour les moustiques).

Nous présentons dans un premier temps le concept de l'habitat basé sur les ressources (*ressource-based habitat*) qui permet de définir l'ensemble des actions et interactions possibles entre *Ae. aegypti* et l'environnement. Dans un second temps, nous présentons les 5 étapes de construction du modèle MODE permettant de générer différents contextes environnementaux et leurs évolutions. Une troisième section est consacrée à une application de MODE à Bangkok (Thaïlande).

2. Concept de l'habitat-basé sur les ressources appliqué à *Aedes aegypti*

Le *Resource-Based Habitat Concept* (RBHC) (Dennis *et al.*, 2003), utilisé en biologie de la conservation, consiste à se baser sur les besoins primaires d'un organisme pour en déduire son habitat fonctionnel ou sa niche écologique potentielle. Elle a été portée dans le champ de la prédiction du risque d'occurrence des maladies à vecteur par (Hartemink *et al.*, 2015). Selon ces auteurs, l'approche RBHC permet de dépasser les limites des modèles inductifs basés sur les relations statistiques mesurées entre des variables environnementales observées et la présence d'un vecteur ou d'un agent pathogène, ainsi que les limites des modèles déductifs constitués d'équations ne tenant pas compte du caractère spatial du phénomène modélisé. Dans le champ des maladies à transmission vectorielles, le RBHC permettrait de reconstruire les espaces potentiels de transmission du virus en rassemblant les espaces fonctionnels des vecteurs et des hôtes. Avec MODE, nous nous concentrons sur la construction d'un modèle dynamique de l'environnement pour le moustique vecteur de la dengue. Il s'agit donc de mettre en relation les caractères perceptifs de ce moustique et les ressources ou conditions qui influent son développement biologique et ses comportements. Sur la base de la classification proposée par Hartemink *et al.* (2015), les éléments de l'environnement perceptibles par *Ae. aegypti* peuvent en effet être séparés en deux catégories. D'une part les ressources qui représentent l'ensemble des éléments environnementaux qui sont

Génération d'environnements artificiels

consommables par le moustique (ex : nectar, sang). D'autre part les conditions qui représentent l'ensemble des états de l'environnement qui influencent le moustique dans son comportement (présence de gîtes de ponte) ou dans son fonctionnement biologique (température).

Ces ressources et conditions varient au cours de la vie d'*Ae. aegypti*. Son cycle de vie est divisé en 4 stades, de la phase aquatique (œuf, larve et nymphe) à la phase aérienne (adulte). Au cours de sa phase aquatique, les conditions de l'eau (qualité chimique et température) ont un impact important sur la survie et la durée de développement du moustique. Une fois adulte, celui-ci est capable de se déplacer, de se reproduire, de se nourrir de nectar et, pour les femelles, de prendre des repas sanguins. Les protéines contenues dans le sang permettent à cette dernière de porter ses œufs à maturité. Au cours de son cycle gonotrophique (cycle repas sanguin – maturation des œufs – ponte), elle dédie une grande partie de son temps au repos dans des lieux frais et ombragés (Vezzani *et al.*, 2005). La ponte se fait dans un élément de l'environnement contenant (ou susceptible de contenir) de l'eau. Il peut tout aussi bien s'agir d'un objet créé par l'homme (un pot de fleur) que d'un objet dont l'existence est liée à la végétation (creux dans un arbre). Par la suite nous désignerons par « gîte larvaire potentiel » (GLP) ces types d'éléments environnementaux. En milieu urbain, les GLP liés aux activités humaines peuvent être très nombreux. Ils sont un facteur primordial de la prolifération des populations d'*Ae. aegypti* dans les villes (Arunachalam *et al.*, 2010).

Si l'importance des différentes ressources et conditions pour *Aedes* varie au cours de sa vie, celle-ci est en partie déterminée par ce qu'il est capable de percevoir à chaque stade de développement. La perception d'*Ae. aegypti* adulte est limitée à une trentaine de mètres environ (Gillies et Wilkies, 1972). Il dispose de capacités olfactives, visuelles, tactiles, chimio-taxiques, thermo-sensorielles et photo-sensorielles. Celles-ci lui permettent d'identifier ou de réagir selon son état à la présence de certains éléments environnementaux : l'humain, l'eau, le nectar des plantes, la lumière, l'humidité et la température. Les relations entre les caractères perceptifs du moustique à l'état pré-adulte et adulte et les objets (conditions et ressources) sur lesquels ils portent permettent ainsi de définir les types de données qui devront être mobilisées pour décrire ces relations. Par exemple, au stade adulte, le moustique peut percevoir des gîtes en eau ou susceptibles de contenir de l'eau lorsqu'il a besoin de pondre. Cinq facteurs environnementaux locaux ont ainsi été définis comme faisant partie du RHBC d'*Aedes* : l'eau (liée aux précipitations ou aux activités humaines), les températures, la végétation, la présence humaine et les gîtes de ponte. Nous décrivons dans la section 3 les méthodes d'évaluation de ces facteurs et les modèles de leurs dynamiques.

3. MODE : Model Of Dynamical Environment

MODE est un modèle à base d'agents où chaque agent représente une unité de base de l'environnement. Celle-ci est décrite par des attributs qui représentent les ressources et conditions du RBHC d'*Ae. aegypti*. Elle dispose également de

SAGEO'2016 – Nice, 6-9 décembre 2016

méthodes qui décrivent les dynamiques de mise en eau des gîtes et la variation des températures (Figure 1).

La résolution spatiale d'une cellule est de 30 m sur 30 m, elle correspond à celle des images Landsat 8 (OLI). Cette résolution est conforme avec la superficie moyenne de « l'espace de vie » du moustique observée par des études de type *mark-release-recapture* (marquage-libération-recapture) (Getis *et al.*, 2003). Elle correspond également à la distance de dispersion moyenne maximale de cohortes de moustiques simulée dans un contexte urbain (Manerat et Daudé, 2016a). Cette résolution est utilisée dans d'autres modèles de simulation (Karl *et al.*, 2014).

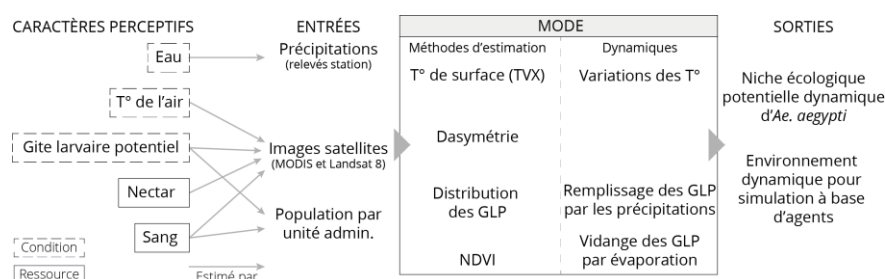


FIGURE 1. Fonctionnement général de MODE : des caractères perceptifs à la niche écologique potentielle dynamique d'*Ae. aegypti*

3.1. Méthode d'estimation des ressources et des conditions

De façon générale, les méthodes d'estimation des caractéristiques environnementales intégrées à MODE sont développées dans le but de pouvoir être appliquées à différentes zones urbaines. Les images satellites, entrées majeures du modèle, sont issues de satellites couvrant la quasi-totalité de l'espace terrestre (Terra et Aqua (MODIS) et Landsat 8) dont les données sont mises à disposition librement. Bien que l'utilisation d'images haute résolution ne soit pas nécessaire au bon fonctionnement de MODE, celles-ci constituent une perspective intéressante (ces données restant toutefois très coûteuses). Par ailleurs, les données sur le recensement, les précipitations et les températures peuvent être acquises librement ou auprès d'autorités locales.

3.1.1. L'eau : un élément essentiel du cycle de vie d'*Ae. aegypti*

La présence d'eau est une condition *sine qua non* de la présence de moustiques. Dans les zones colonisées par *Ae. aegypti* soumises à un climat de mousson, la saison des pluies correspond à la période de prolifération maximale des moustiques. Les données de précipitation utilisées dans MODE sont issues de relevés par station météorologique en accès libre dans la plupart des cas. Ces données correspondent à la quantité (mm) de précipitations enregistrées par station et par jour. Les systèmes de stockage d'eau constituent une seconde opportunité pour le développement du

moustique. Ces systèmes peuvent être liés à l'absence de réseau d'adduction d'eau ou aux pratiques culturelles telles que l'arrosage des fleurs.

3.1.2. La température de l'air : les zones de confort d'*Aedes aegypti*

La température de l'air influence les dynamiques biologiques du moustique et du virus (Lambrechts *et al.*, 2011), mais aussi le comportement des hôtes (utilisation de climatiseurs et vie à l'intérieur ou l'extérieur). De nombreuses études soulignent les relations qu'elles entretiennent avec les épidémies de dengue (*e.g.* Descloux *et al.*, 2012 ; Hopp *et al.*, 2003). Les températures locales évoluent durant la journée selon de nombreux facteurs dont le principal est le mode d'occupation du sol (Misslin *et al.*, 2016). Dans le cadre de MODE, les données de température de l'air (T_a) sont estimées à partir d'images MODIS de température de surface (LST). Cette dernière exprime la chaleur émise par la surface terrestre et diffère de quelques degrés par rapport à la T_a . Cependant, grâce à une méthode appelée *temperature-vegetation index* (TVX), les fortes corrélations entre la LST et la T_a et entre la couverture végétale et la T_a permettent d'estimer cette dernière lorsque des relevés météorologiques issus de stations sont disponibles. La méthode TVX est utilisée et décrite par plusieurs auteurs (*e.g.* Vancustem *et al.*, 2010 ; Zhu *et al.*, 2013). Les données de température résultant de l'estimation de la T_a sont affectées aux cellules environnement. Cette température évolue en fonction des relevés de LST MODIS (MOD11A1 ou MYD11A1). Lorsque ceux-ci sont disponibles, la température de la cellule environnement prend la valeur qui lui correspond. Lorsque les conditions nuageuses ne permettent pas l'acquisition d'images de LST, la T_a de la station météorologique la plus proche est affectée à la cellule environnement en question.

3.1.3. La végétation : les zones refuges productrices de nectar

Aedes aegypti peut être attiré par les zones végétalisées pour se reposer, pour pondre ou lorsqu'il recherche du nectar pour se nourrir. Les espaces végétalisés proches des espaces bâtis sont des espaces très producteurs de moustiques (Vezzani *et al.*, 2005). Pour évaluer la présence de ces espaces à une résolution fine, nous utilisons le NDVI (Normalized Difference Vegetation Index) (Tucker, 1979). Cet indice de végétation est couramment utilisé en télédétection. Il est calculé à partir des bandes rouge et infrarouge d'une image satellite (Landsat 8). Les résultats sont compris entre -1 et 1. Les valeurs négatives indiquent généralement une absence de végétation. Une valeur de 1 indique une forte présence de végétation.

3.3.4. Les densités humaines : les possibilités de repas sanguin

Les densités humaines représentent des stocks de sang pour le moustique. Il s'agit d'un élément essentiel pour son cycle gonotrophique. MODE estime ces densités au lieu de résidence, le rôle de la mobilité des hôtes dans la propagation des épidémies n'est donc pas abordé ici. Le niveau d'agrégation des données de recensement des populations pose une première difficulté : dans MODE les cellules-environnements sont définies à une résolution de 30 m sur 30 m alors que les données de recensement sont souvent agrégées sur des unités administratives de

SAGEO'2016 – Nice, 6-9 décembre 2016

plusieurs km². La cartographie dasymétrique permet de résoudre ce premier problème (Yuan, 1997). Elle est principalement utilisée pour affiner des données démographiques en les redistribuant d'une échelle large à une échelle plus fine, cette dernière étant définie par rapport à des données satellites (images brutes, occupation du sol, température de surface, indices spectraux ...). La méthode de cartographie dasymétrique est décrite dans plusieurs publications (Li et Weng, 2010). Schématiquement la méthode revient à ventiler les données de populations (nombre d'habitants et nombre de ménages dans MODE) au lieu de résidence fourni à un échelon administratif dans des zones bâti qui occupent cet échelon administratif. Dans le cadre de MODE, ces données satellites sont extraites d'une image satellite Landsat 8. La classification d'occupation du sol est obtenue en appliquant une classification non-supervisée ISODATA.

3.3.5. Les gîtes larvaires potentiels : les lieux de ponte

La présence de GLP est une donnée difficile à estimer : il est pratiquement impossible de la mesurer de façon directe à partir d'images satellite (Moloney *et al.*, 1998) et il est coûteux de collecter cette donnée sur le terrain. En effet, de nombreuses études ont montré qu'*Ae. aegypti* pond de manière préférentielle dans des petits gîtes domiciliaires ou péri-domiciliaires contenant de l'eau fraîche et stagnante. La répartition de ces gîtes dans l'espace urbain est très largement liée à la présence des hommes et à leurs activités (Ooi *et al.*, 2006). Karl *et al.* (2014) proposent d'utiliser le nombre de ménages et la végétation pour estimer la présence de gîtes larvaires dans l'environnement en construisant un *breeding site abundance index* (indice d'abondance de GLP). Notre démarche s'en inspire, le nombre de gîtes potentiels dans une cellule environnement est une valeur de stock estimée en partant de l'hypothèse que les ménages en sont les principaux producteurs (les cellules environnement associées à de l'eau ne produisent pas de GLP) :

$$Gpotint_i = D_i \cdot \alpha_i \cdot \Delta \quad (1)$$

$$Gpotext_i = \mu_i + (D_i \cdot \alpha_i)(1 - \Delta) + D_{j(i)} \cdot \beta_{j(i)} \quad (2)$$

$$Gpottot_i = Gpotint_i + Gpotext_i \quad (3)$$

Dans l'ensemble d'équations précédant, i désigne une cellule environnement, $j(i)$ désigne les cellules voisines d'ordre 1 de la cellule i , i_{occsol} correspond à l'occupation du sol de i (classification non-supervisée réalisée sur une image Landsat 8), $Gpotext_i$ correspond au nombre de GLP situés à l'extérieur d'un bâtiment dans i , $Gpotint_i$ est le nombre de GLP situés à l'intérieur d'un bâtiment dans i , $Gpottot_i$ désigne le nombre de GLP total dans i et D_i correspond au nombre de ménages de i . Quatre paramètres sont utilisés dans ces équations : μ_i correspond au nombre de GLP de i produits quel que soit le nombre ménages habitants i (entier aléatoire), il permet de générer des gîtes dans des cellules de type végétalisé par exemple ; α_i représente le nombre de GLP créés dans i par un ménage habitant i , Δ_i représente la part des gîtes intérieurs parmi les GLP créés par les ménages habitant i ($0 < \Delta < 1$) et $\beta_{j(i)}$ correspond au nombre de GLP créés dans i par les ménages habitant j .

Génération d'environnements artificiels

Les paramètres μ , α et β sont des entiers dont la valeur est issue d'une loi de Poisson avec une espérance mathématique λ fixée par l'utilisateur. Cette valeur dépend du contexte local. Ces paramètres indiquent en moyenne et en écart-type le nombre de GLP que l'on trouve par unité de surface dans des espaces non-bâties (μ), par unité de ménage dans les espaces bâtis (α) et par densité de voisinage (β). La production de GLP par ménage est ici homogène. Un critère de discrimination, selon le statut économique des ménages par exemple, pourrait être introduit.

MODE simule les dynamiques des deux stocks GLP intérieurs et extérieurs de chaque cellule à partir de deux processus : le remplissage naturel (pluie) ou artificiel (arrosage/climatiseur/collecte d'eau) et la vidange (évaporation ou manuelle).

Le remplissage s'effectue différemment selon la situation du gîte (intérieur d'un bâtiment / extérieur). Les gîtes intérieurs sont remplis par les hommes. Ils ont une contenance maximale et une contenance minimale (en mm) fixées par l'utilisateur. La hauteur d'eau initiale du stock de gîtes intérieurs est donnée par tirage aléatoire dans une loi uniforme entre 0 et 1. Lorsque le niveau d'eau atteint la contenance minimale, celui-ci est rempli jusqu'à la limite de sa contenance maximale. Le stock de gîtes extérieurs est rempli par les précipitations. Si un stock de gîtes situés à l'extérieur atteint sa contenance maximale et reçoit encore de l'eau via les précipitations, celui-ci déborde.

La vidange des gîtes s'effectue par évaporation pour les gîtes intérieurs et les gîtes extérieurs. Celle-ci est calculée à l'échelle de la ville grâce à l'équation de Linacre (1977) qui prend notamment en entrée les horaires de lever et de coucher du soleil, la température de l'air et les coordonnées géographiques. Les gîtes intérieurs peuvent également être vidés par consommation d'eau. Le rythme de la consommation peut-être définie par le modélisateur selon différentes stratégies : par exemple, chaque jour une certaine proportion des gîtes est vidée. Seuls les GLP intérieurs sont à nouveau remplis lorsqu'ils atteignent leur contenance minimale.

Ces dynamiques de remplissage et de vidange des gîtes larvaires potentiels, à l'intérieur comme à l'extérieur, permettent de prendre en compte les processus menant à l'accumulation de l'eau durant plusieurs jours consécutifs, ce qui favorise potentiellement la mise en eau des œufs et le développement du moustique. Elles permettent également de prendre en compte le débordement des gîtes situés à l'extérieur, ce qui entraîne potentiellement la perte d'une partie de la population de larves des gîtes actifs. Enfin, elles permettent de prendre en compte la présence de gîtes en dehors de la saison des pluies du fait des usages humains.

4. Application de MODE à Bangkok (Thaïlande)

La zone d'expérimentation choisie est la ville de Bangkok (Thaïlande) située dans la zone intertropicale (13°45'22"N, 100°30'06"E). Son climat est tropical avec une saison sèche (5 à 44 mm de précipitations) de novembre à mars et une saison humide (160 à 342 mm de précipitations) d'avril à octobre. Les températures sont

SAGEO'2016 – Nice, 6-9 décembre 2016

comprises entre 26 et 30°C (moyenne de 28°C). La ville de Bangkok compte 8,2 millions individus (recensement NSO 2010²) pour un territoire administratif de 1568 km². Celui-ci présente une grande diversité en termes de densité de population et d'occupation du sol. *Aedes aegypti* est détecté tout au long de l'année avec un minimum durant la saison sèche et un maximum au cours de la mousson.

Le logiciel R (3.1.3) a été utilisé pour constituer les jeux de données. MODE est développé sous GAMA (1.7). Les images de télédétection utilisées sont issues du satellite Landsat 8 OLI (datant du 17/01/2014) et des satellites MODIS Aqua et Terra (MYD11A1 et MOD11A1, deux images par jour entre 2009-2011). Ces dernières sont téléchargées et prétraitées par le biais du package R MRTtools). Les données démographiques proviennent du recensement de 2010 (BMA 2010 Census). Enfin, les données météorologiques sont accessibles librement sur le site web du Thai Meteorological Department³. Elles ont été acquises pour la période 2009-2012.

4.1 Les précipitations : données météorologiques

Les données de précipitations sont issues d'une station météorologique située au centre de Bangkok. Elles sont collectées pour la période 2009-2012. Ces données sont à la base des dynamiques de remplissage des gîtes larvaires potentiels situés à l'extérieur des bâtiments (Figure 3).

4.2. Température de l'air : estimation par la méthode TVX

Le satellite AQUA survole Bangkok aux alentours de 2h00 et 14h00 (heure locale). Le satellite TERRA passe aux alentours de 10h30 et 22h30. Les images issues du premier satellite sont donc plus à même d'estimer les températures minimales et maximales journalières à Bangkok. Celles-ci sont estimées pour la période 2009-2011 et pour une zone plus large que la BMA afin d'augmenter le nombre de stations de référence dont les données permettent d'appliquer la méthode TVX. Sur cette zone élargie, 19,8% des images MYD11A1 nocturnes et 17,6% des images MYDD11A1 diurnes contiennent au moins la moitié de pixels valides. Ces images sont converties en température de surface puis la température de l'air est estimée. Au total, la température de l'air est estimée pour 217 images thermales nocturnes et 193 images thermales diurnes. La Figure 2a présente un exemple de distribution spatiale de la température de l'air (températures minimales nocturnes au 20/01/2009).

Les données de température de l'air étant fournies à une résolution plus grande que la taille des cellules environnement, ces dernières prennent la valeur de Ta du pixel de 1 x 1 km sur lequel elles se trouvent. Les données de Ta ne sont pas disponibles lorsque la couverture nuageuse est trop importante, durant la mousson

² <http://popcensus.nso.go.th/upload/census-report-6-4-54-en.pdf>

³ <http://www.tmd.go.th/en/climate.php>

notamment. Si la Ta d'un pixel n'est pas renseignée, MODE attribue la Ta mesurée par la station la plus proche géographiquement pour chaque cellule environnement.

4.3. Végétation : estimation par le NDVI

Le NDVI est calculé à partir des bandes rouge et infrarouge de l'image Landsat 8. Les valeurs du NDVI (Figure 2b) varient de -0,32 à 0,82. Les valeurs inférieures à 0 ont été rehaussées à 0. Elles correspondent à l'absence totale de végétation. Bangkok apparaît globalement comme un territoire présentant une couverture végétale importante, notamment dans la périphérie rurale (champs végétalisés). Au centre de Bangkok, les espaces végétalisés sont plus fragmentés. Ils correspondent essentiellement à des parcs et des jardins. Les axes de transport apparaissent distinctement de par les zones de bâti dense qui les bordent. Les données issues du NDVI seront utilisées comme une ressource pour le moustique *Aedes* et considérées comme statiques, la dynamique végétale n'ayant à priori que peu d'impact, en tout cas non renseigné, sur les dynamiques de population vectorielle.

4.4. Densité de population et de ménages : estimation par la cartographie dasymétrique

Les données d'occupation du sol sont extraites de l'image Landsat 8. Deux classes de bâti principales ont été identifiées : bâti dense (correspondant au centre-ville de Bangkok) et bâti lâche (correspondant à la périphérie de la ville), sans distinction quant à leur usage ou leur fonction. Les populations sont ventilées dans ces classes d'occupation du sol potentiellement habitables à partir de la méthode de cartographie dasymétrique (Figure 2c). Cette méthode permet d'estimer la répartition des populations à une échelle plus fine que l'échelle du *Kwaeng* (unité administrative de référence du recensement du NSO dont la superficie est comprise entre 0,15 et 73,92 km²). Les densités estimées les plus fortes sont localisées principalement dans le centre de la ville, à l'Est du fleuve Chao Phraya qui la traverse. Le tissu urbain est très dense tout en comportant quelques zones inhabitées et végétalisées. Dans les espaces périphériques de l'Est et de l'Ouest, les densités les plus fortes sont réparties le long des axes routiers importants. Une grande part de la périphérie Ouest présente un tissu urbain lâche aux densités de populations faibles.

4.5. Gîtes larvaires potentiels : estimation de leur distribution spatiale

Le nombre de ménages par cellule environnement calculé par la méthode dasymétrique est un paramètre essentiel au calcul du nombre de gîtes larvaires potentiels (GLP) par cellule (équation 1). Les zones en eau (canaux, rizières, étangs) sont les seules dans lesquelles il n'est *a priori* pas possible de trouver de GLP. Les paramètres $\lambda\mu=4$, $\lambda\alpha=6$, $\lambda\beta=3$ utilisés à Bangkok donnent une moyenne de 17,8 gîtes/cellule et un écart-type de 28,5 sur l'ensemble de la ville (min.=0, Q1=4, méd.=8, Q3=17, max.=452). L'analyse de sensibilité a montré que le nombre de GLP moyen global varie très peu d'une simulation à l'autre en fixant les valeurs de

SAGEO'2016 – Nice, 6-9 décembre 2016

μ , α et β . À titre d'exemple, pour 100 simulations réalisées avec les valeurs de paramètres ci-dessus, l'écart-type du nombre de GLP moyen produits au cours des simulations ne dépasse pas 0,01.

Les zones les plus pourvues sont principalement situées au centre-ville et dans les espaces d'habitat pavillonnaire situés aux alentours de celui-ci. Les valeurs minimales, moins de 10 gites, se trouvent dans les cellules végétalisées et éloignées des zones habitées, comme dans la périphérie Est et la périphérie Ouest, ou dans des zones urbaines bâties mais peu denses, telle que le Kwaeng de Phra Borom Maha Ratchawang (partie Ouest de l'encart de la Figure 2d) et rassemblant le palais royal, des temples et des ministères.

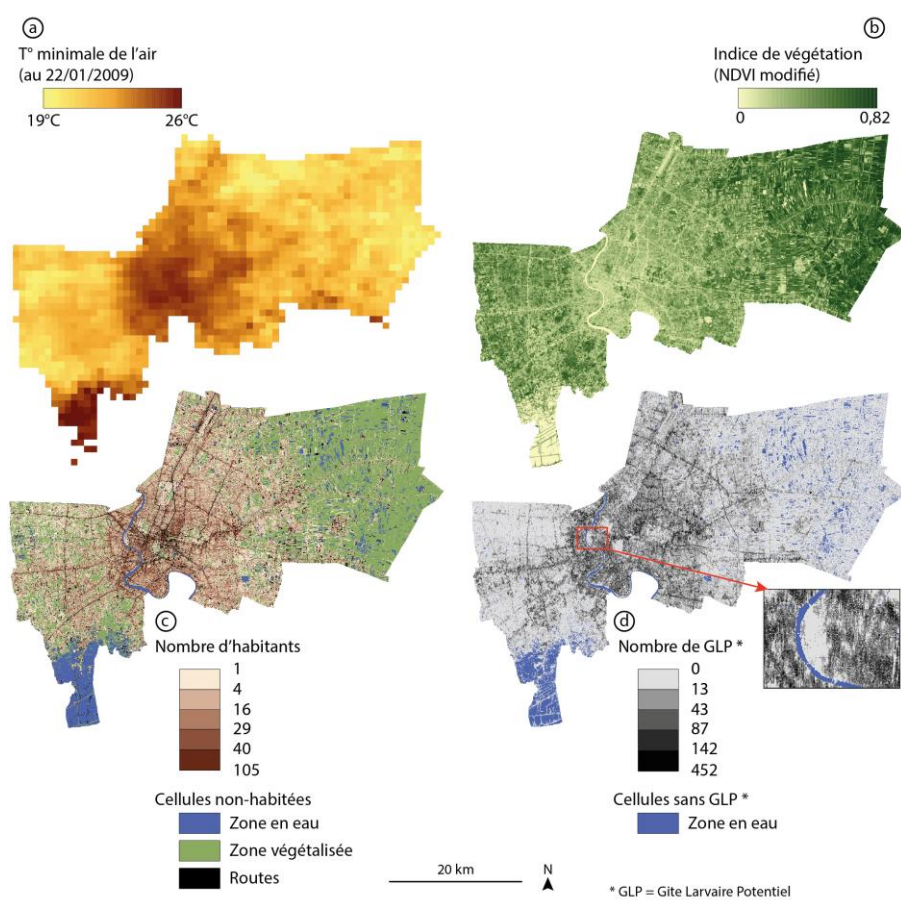


FIGURE 2. Estimation des ressources et des conditions pour *Ae. aegypti* à Bangkok

5. Discussion et conclusion

La méthode définie précédemment permet de quantifier pour chaque cellule environnement 4 attributs environnementaux : le NDVI, le nombre d'habitants, le nombre de gîtes (intérieurs et extérieurs) et la température de l'air. Un cinquième attribut, les précipitations, est fourni globalement en entrée du modèle de remplissage des gîtes potentiels. Parmi ces attributs, deux sont dynamiques : les températures et les précipitations. Ceux-ci agissent directement sur le taux de remplissage des gîtes et la température de l'eau. La Figure 3a représente par exemple la dynamique en eau d'un stock de gîtes situés à l'extérieur. Le phénomène d'accumulation d'eau peut être comparé aux précipitations ponctuelles enregistrées. La possibilité de simuler l'accumulation et l'évaporation est fondamentale pour l'interaction avec le modèle moustique (MOMA). En l'absence de données entomologiques, cette dynamique de disponibilité de gîtes en eau peut être comparée à la distribution temporelle de l'incidence hebdomadaire de la dengue à Bangkok en 2012 (Figure 3b). La forme des deux distributions est similaire. Elles présentent un décalage d'un mois environ qui peut s'expliquer par le temps d'accès des œufs au stade adulte et par le cycle d'infection. Ceci est un résultat encourageant d'un point de vue macroscopique.

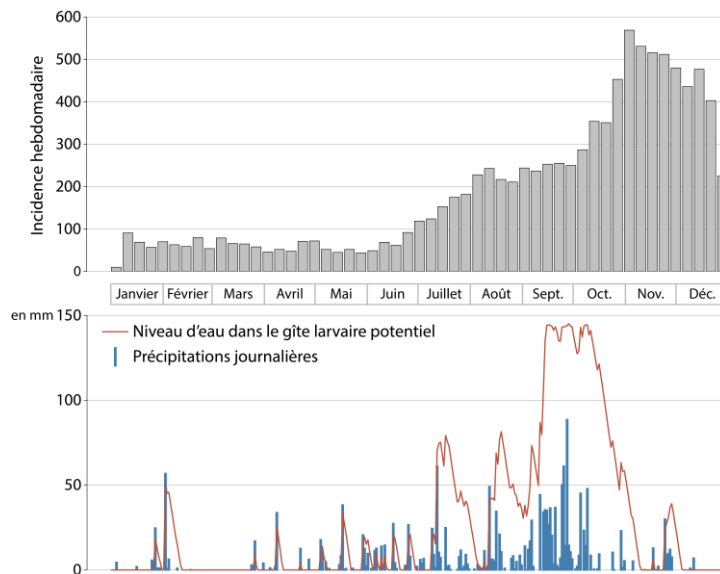


FIGURE 3. Incidence hebdomadaire de dengue, précipitations et évolution simulée du niveau d'eau dans un GLP extérieur de 150 mm de hauteur à Bangkok (2012)

MODE est construit dans l'objectif de pouvoir être transposé d'une ville à une autre. Le fait qu'il soit construit à partir de données libres ou accessibles dans la plupart des villes en font un modèle relativement souple et générique. À court terme,

SAGEO'2016 – Nice, 6-9 décembre 2016

MODE sera couplé au modèle moustique MOMA (Maneerat et Daudé, 2016a) afin de simuler les dynamiques de stocks de moustiques à l'échelle d'une ville, données entomologiques difficiles à obtenir sur le terrain.

Remerciements :

Cet article a été écrit dans le cadre du projet DENFREE (Dengue Research Framework for Resisting Epidemics in Europe, grant agreement : 282 378), septième programme de la Commission européenne (FP7). Les auteurs remercient leurs partenaires Alexandre Cebeillac, Bertrand Lefebvre, Somsakun Maneerat, Patrick Taillandier, Yvette Vaguet, Alain Vaguet (Université de Rouen), Olivier Telle et Richard Paul (Institut Pasteur). Les auteurs remercient également les relecteurs pour l'ensemble de leurs suggestions.

Bibliographie

- Almeida S. (de), Ferreira R., Eirasc Á., Obermayr R., Geier M. (2010). Multi-agent modeling and simulation of an *Aedes aegypti* mosquito population. *Environmental Modelling & Software*, vol. 25, n° 12, p.1490-1507.
- Arunachalam N., Tana S., Espino F., Kittayapong P., Abeyewickreme W., Wai K., Tyagi B., Kroeger A., Sommerfeld J., Petzold M. (2010). Eco-bio-social determinants of dengue vector breeding: a multicountry study in urban and periurban Asia. *Bulletin of the World Health Organization*, vol. 88, n° 3, p. 173-184.
- Bhatti S., Tripathi N. (2014). Built-up area extraction using Landsat 8 OLI imagery. *GIScience & Remote Sensing*, vol. 51, n° 4, p. 445-467.
- Cummins B., Cortez R., Foppa I., Walbeck J., Hyman J. (2012). A spatial model of mosquito host-seeking behavior. *PLoS Computational Biology*, vol. 8, n° 5, e1002500.
- Daudé É., Vaguet A., Paul R. (2015). La dengue, maladie complexe. *Natures Sciences Sociétés*, vol. 23, n° 4, p. 331-342.
- Dennis R. (2003). Towards a functional resource-based concept for habitat: a butterfly biology viewpoint. *Forum*, vol. 102, n° 2, p. 417-426.
- Descloux É., Mangeas M., Menkes C., Lengaigne M., Leroy A., Tehei T., Guillaumot L., Teurlai M., Gourinat A., Benzler J., Pfannstiel A., Grangeon J., Degallier N., Lamballerie X. (de) (2012). Climate-based models for understanding and forecasting dengue epidemics. *PLoS neglected tropical diseases*, vol. 6, n° 2, e1470.
- Getis A., Morrison A., Gray K., Scott T. (2003). Characteristics of the spatial pattern of the dengue vector, *Aedes aegypti*, in Iquitos, Peru. *The American journal of tropical medicine and hygiene*, vol. 69, n° 5, p. 494-505.
- Gillies M.T., Wilkes T.J. (1972). The range of attraction of animal baits and carbon dioxide for mosquitoes. Studies in a freshwater area of West Africa. *Bull Entomol Res*, vol. 61, p. 389-404.
- Hartemink N., Vanwambeke S., Purse B., Gilbert M., van Dyck H. (2015). Towards a Resource-based habitat approach for spatial modelling of vector-borne disease risks. *Biological Reviews*, vol. 90, n° 4, p. 1151-1562.

Génération d'environnements artificiels

- Hopp M., Foley J (2003). Worldwide fluctuations in dengue fever cases related to climate variability. *Climate Research*, vol. 25, p. 85-94.
- Karl S., Halder N., Kelso J., Ritchie S., Milne G. (2014). A spatial simulation model for dengue virus infection in urban areas. *BMC Infectious Diseases*, vol. 14, n° 1, p. 447-464.
- Lambrechts L., Paaijmans K., Fansiri T., Carrington L., Kramer L., Thomas M., Thomas, Scott T. (2011). Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, n° 18, p. 7460-7465.
- Li G., Weng Q. (2010). Fine-scale population estimation: how Landsat ETM+ imagery can improve population distribution mapping. *Canadian Journal of Remote Sensing*, vol. 36, n° 3, p. 155-165.
- Linacre, E. (1977). A simple formula for estimating evaporation rates in various climates, using temperature data alone. *Agricultural Meteorology*, vol. 18, p. 409-424.
- Maneerat S., Daudé É. (2016a). A spatial agent-based simulation model of the dengue vector *Aedes aegypti* to explore its population dynamics in urban areas. *Ecological Modelling*, vol. 333, p. 66-78.
- Maneerat S., Daudé É. (2016b). Étude par simulation à base d'agents des effets des discontinuités intra-urbaines à Delhi sur la dispersion des moustiques *Aedes aegypti*, vecteurs de la dengue, de la fièvre jaune, du chikungunya et du virus Zika. *Cybergéo : Européen Journal of Geographie*. [en cours d'évaluation]
- Misslin R., Telle O., Daudé É, Vaguet A., Paul R. (2016). Urban Climate versus global climate change – What makes the difference for dengue?. *Annals of the New York Academy of Science*. [à paraître].
- Moloney J., Skelly C., Weinstein P., Maguire M., Ritchie S. (1998). Domestic *aedes aegypti* breeding site surveillance: limitations of remote sensing as a predictive surveillance tool. *The American journal of tropical medicine and hygiene*, vol. 59, n° 2, p. 261-264.
- Ooi E., Goh K., Gubler D (2006). Dengue prevention and 35 years of vector control in Singapore. *Emerging infectious diseases*, vol. 12, n° 6, p. 887-893.
- Rodhain, F., Rosen, L. (1997). Mosquito vectors and dengue virus–vector relationships. *Dengue and Dengue Hemorrhagic Fever*. New York, CAB International, p. 45-60.
- Tucker, C. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, vol. 8, p. 127-150.
- Vancutsem C., Ceccato P., Dinku T., Connor S. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, vol. 114, n° 2, p. 449-465.
- Vezzani D., Rubio A., Velázquez S., Schweigmann N., Wiegand T. (2005). Detailed assessment of microhabitat suitability for *Aedes aegypti* (Diptera: Culicidae) in Buenos Aires, Argentina. *Acta Tropica*, vol. 95, n° 2, p. 123-131.
- Zhu W., Lu A., Jia S. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, vol. 130, p. 62-73.

CONTRIBUTION DE LA SEGMENTATION D'IMAGES EN GEOMORPHOLOGIE POUR SPATIALISER LES PROCESSUS HYDRO-PEDOLOGIQUES

Julien Andrieu¹,

*1. UMR ESPACE CNRS, Université Nice Sophia Antipolis
98, boulevard Edouard Herriot, 06204 Nice Cedex 02, France
julien.andrieu@unice.fr*

RESUME. Cet article présente à la fois une méthode originale issue d'une chaîne de traitements d'images géographiques permettant l'établissement d'une segmentation en unités homogènes et continues et une typologie des unités. Il en présente un exemple d'application en géomorphologie sur deux Bassins versants des Cévennes où une typologie de versants a été réalisée.

ABSTRACT. This paper presents on the one hand an original method made of a chain of geographical image processing technics which enable a segmentation in homogeneous and continuous units and a typology of these units. On the other hand it presents an example of application in geomorphology on two water basins in Cevennes (France) where a typology of hillslope units has been realized

MOTS-CLES : Traitement d'images, Segmentation, Ligne de partage des eaux, Télédétection typologie, classification, géomorphologie, hydrologie Cevennes.

KEYWORDS: Image processing, Segmentation, Watershed, Remote sensing, Typology, Classification, Geomorphology, Hydrology, Cevennes.

SEGMENTATION D'IMAGES EN GEOMORPHOLOGIE

1. Introduction

Parmi les nombreux objectifs de la géomatique et de l'analyse spatiale, se trouve la définition et délimitation de régions entendues ici comme simples entités spatiales caractérisées par une continuité et une homogénéité (Pumain et Saint-Julien 2010). Or, cet exercice, pour autant qu'il fût très fréquent dans la cartographie pré-géomatique, (délimitations des régions en géographie rurale ou régionale, zonages manuels de la photo-interprétation, etc.) est, à l'ère de la géomatique, assez rarement pratiqué. Dans le cadre du projet ANR Floodscale et plus précisément de la question des changements d'échelles posée par le programme (Braud et al., 2014), s'est posée la question de la généralisation d'observations ponctuelles (transects de quelques dizaines de mètres) au niveau de petites entités géomorphologiques constituant les versants, puis des versants eux-mêmes jusqu'aux bassins versants emboîtés de taille croissante. Pour les deux premières étapes de cet emboîtement d'échelles, a été explorée (et retenue) l'idée de la réalisation d'une méthodologie originale et géomatique en vue d'une typologie de versants (Braud et al., 2014). En effet, en hydrologie, notamment dans les approches de modélisations hydrologiques spatialement explicites, le bassin versant est très fréquemment subdivisé en entités homogènes et différentes entre elles par des caractéristiques qui conditionnent l'hydrologie (géologie, pédologie, relief, couverture végétale). La mise en place d'une modélisation hydrologique distribuée nécessite en effet la spécification d'un maillage. Deux approches sont dès lors connues et appliquées des hydrologues le maillage régulier et la division en sous-bassins.

De nombreuses approches reposent sur un maillage en grilles régulières définies par le modèle numérique de terrain. Une classification par pixel du bassin versant permet en effet créer des classes de pixels homogènes dans les descripteurs morphologiques. Cependant, ces classes sont, dès lors, sans cohérence spatiale.

Certaines équipes impliquées dans FloodScale travaillent depuis de nombreuses années sur la définition de maillages en unités de réponses hydrologiques (HRUs, Flügel, 1995) ou pédo-paysages (Dehotin et Braud, 2008) sur lesquelles les processus hydrologiques peuvent être considérés comme homogènes. Ceci conduit à des modélisations sur maillages irréguliers. Dans ces applications, c'est une subdivision en sous-bassins versants qui est effectuée. Les mailles présentent alors une grande cohérence spatiale mais elles peuvent être fortement hétérogènes pour les descripteurs (versants asymétriques, géologie transversale au découpage en sous bassin versants, etc). Dans le cadre du projet ANR Floodscale et plus précisément de la question des changements d'échelle posée par le programme (Braud et al., 2014), s'est posée la question de la généralisation d'observations ponctuelles (transects de quelques dizaines de mètres) au niveau de petites entités géomorphologiques constituant les versants, puis des versants eux-mêmes jusqu'au bassin versant. Cette méthode de segmentation spatiale géomorphologique a permis de donner un cadre spatial du Bassin versant qui opère le changement d'échelle entre la facette locale et l'ensemble du bassin versant.

SAGEO'2016 – Nice, 6-9 décembre 2016

La méthode de segmentation spatiale géomorphologique originale et reproductible est ici présentée comme :

- Outil de délimitation de petites entités géomorphologiques homogènes d'échelle locale
- Classification de ces entités en un nombre de types de versants

Cette méthode originale permet donc bien d'assurer à la fois un découpage en entités continues de petite taille (versants) et de définir un petit nombre de types de versants pour paramétrer d'éventuels modèles hydrologiques. L'échelle du versant, intermédiaire entre l'échelle d'observation (locale) et l'échelle d'intérêt (bassin versant de plusieurs centaines de km²) est pertinente pour cet objectif parce qu'elle permet une observation au plus près des processus mis en jeu. Des méthodes assez proches ont été employées pour la détection des glissements de terrain sur des indicateurs uniquement morphologiques et sans l'étape de typologie (Seijmonsbergen et al., 2012)

La méthode ici présentée est totalement généralisable à toute autre application nécessitant la segmentation en entités homogènes et continues avec ou sans la réalisation d'une typologie à partir d'un nombre de variables quantitatives transcrites en imagerie géographique. Elle a déjà été testée pour la distinction d'entités paysagères et serait très probablement parfaitement adaptée pour une segmentation urbaine en quartiers, une segmentation écologique en écosystèmes, etc. En ce qui concerne son adaptation à l'écologie du paysage une modélisation de la chaîne de traitement sera envisagée.

2. Zone d'étude, données

2.1. Zone d'étude

TABLE 2. *Caractéristiques des sites d'études*

<i>Bassins</i>	<i>Valescure (30)</i>	<i>Tourgueille (30)</i>
<i>Superficie</i>	3,83 km ²	
<i>Altitudes</i>	244 – 815 M	
<i>Géologie</i>	Granite	Schiste
<i>Occupation du sol</i>	Forêt	Forêt
<i>Processus hydrologiques de transfert supposé dominant</i>	Flux de sub-surface et ruissellement sur surfaces saturées	Flux de sub-surface et ruissellement sur surfaces saturées

Deux petits bassins versants ont ici permis de tester la méthode : le bassin versant de Tourgueille et celui de Valescure dans les Cévennes (France). Tous deux sont de très petite taille, tous deux présentent plusieurs centaines de mètres de

SEGMENTATION D'IMAGES EN GEOMORPHOLOGIE

dénivelé entre le point culminant et l'exutoire. Tous deux sont, en majorité, couverts de forêts (châtaigneraies, chênaies). Les pentes y sont très fortes dans les deux cas avec localement d'anciennes terrasses en majorité déstabilisées et remobilisées. Le bassin versant de Valescure est orienté Nord-Sud et majoritairement sur roche granitique alors que Tourgueille est orienté est-ouest et majoritairement sur schistes.

2.2. Données

Deux jeux de données issues de la télédétection ici mobilisés. D'une part, un Modèle Numérique de Terrain (MNT) réalisé par LIDAR aéroporté à 1 mètre de résolution. D'autre part quatre images satellitales à Très Haute Résolution PLEIADES mises à disposition du programme par le CNES (février et juillet 2013, pour les deux Bassins versants).

3. Production d'un ensemble de descripteurs géomorphologiques par traitement d'image

La segmentation d'images (figure 1) a été appliquée sur les 7 descripteurs suivants :

1. Altitudes, pour différencier les parties hautes des versants des parties basses
2. Pente, pour distinguer les facteurs favorables au ruissellement (pente forte) ou à l'infiltration (pente faible)
3. Orientation (Nord-Sud), pour décrire les topoclimats
4. Dimension fractale (Eastman, 1985), pour distinguer les facteurs favorables au ruissellement (faible dimension fractale) ou à l'infiltration (forte dimension fractale)
5. Fréquence locale de sol nu ou de roches affleurantes pour distinguer les facteurs favorables au ruissellement (grande fréquence de roches) ou à l'infiltration (petite fréquence de roches)
6. NDVI de février (feuillage important en Février)
7. NDVI de juin (feuillage important en Juin) pour distinguer les facteurs favorables au ruissellement (faible NDVI) ou à l'infiltration (fort NDVI).

SAGEO'2016 – Nice, 6-9 décembre 2016

L'ensemble de ces descripteurs a été « validée à dire d'expert » lors des ateliers de l'ANR Floodscale où la méthode a été présentée aux hydrologues et hydrauliciens qui emploient classiquement ces mêmes indicateurs mais, soit en maille régulière (pixels) soit en sous bassins versants.

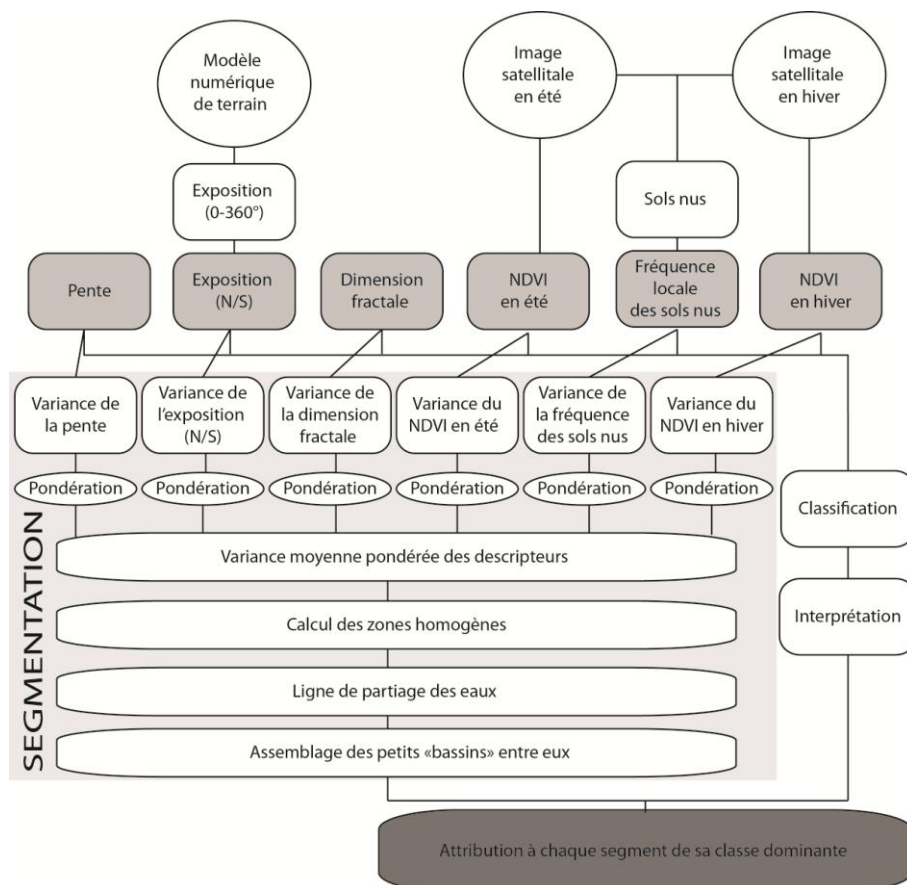


FIGURE 1. Organigramme de la chaîne de traitement

3.1. Traitement du MNT et description morphologique

A partir des images satellites Pleiades, est calculé le NDVI qui est un indice de végétation basé sur la différence normalisée entre la luminance de deux bandes spectrales d'une image produite par radiomètres (ici image du satellite Pleiades). Les deux bandes spectrales sont le rouge (fin du domaine du visible) absorbées par la chlorophylle active et le Proche Infrarouge (début du domaine de l'Infrarouge) réfléchi par les végétaux en phase d'activité photosynthétique. Leur différence

SEGMENTATION D'IMAGES EN GEOMORPHOLOGIE

normalisée permet donc de transcrire en un indice simple l'importance du volume foliaire et de son activité photosynthétique à la date de prise de vue selon une échelle -1 à 1 où les valeurs négatives correspondent aux états de surface non végétalisés, où la végétation très ouverte ou peu active apparaît dans les faibles valeurs positives et où les valeurs proches de 1 correspondent à une végétation très verte et très dense. Dans un tel milieu aux pentes très fortes et aux sols sableux, l'eau ruisselle ou s'infiltre en profondeur. Il n'y a pas de zone humide à détecter. Ainsi, la saisonnalité du NDVI offre une distinction facile des types de couvertures végétales qui reflètent eux même des types de comportements de l'eau dans le sol.

3. Production d'un ensemble de descripteurs géomorphologiques par traitement d'image

3.1. Traitement du MNT et description morphologique

Premièrement, à partir du MNT sont calculées une carte des pentes et une carte de l'orientation des versants. Cependant, la carte d'orientation codée « en degrés de boussole » n'est pas une variable quantitative puisque 0 y est égal à 360. Ainsi, une conversion est faite pour ne transcrire que l'opposition Nord-Sud. Ainsi les degrés de boussole de 0 à 180 sont intouchés et les degrés de 181 à 360 sont inversés et recodés de 179 à 0 de sorte à ce que l'est et l'ouest correspondent à une même valeur de 90 et que la variable devienne quantitative entre 0 le Nord et 180 le Sud.

Deuxièmement, la dimension fractale est calculée selon la méthode d'Eastman (1985) dont la formule est rappelée ci-dessous.

$$\text{Fractional_Dimension} = \frac{\log(2)}{\log(2) + \log\left(\sin\left(\frac{180 - \text{slope}}{2}\right)\right)}$$

Enfin, ces trois descripteurs géomorphologiques et la variable altitude (MNT non traité) sont réétalonnées linéairement sur un octet par pixel soit des valeurs comprises entre 0 pour la valeur minimale 255 pour la valeur maximale.

3.2. Traitement des images satellitaires et description de la couverture du sol

A partir des images satellites Pleiades, est calculé le NDVI qui est un indice de végétation basé sur la différence normalisée entre la luminance de deux bandes spectrales d'une image produite par radiomètres (ici image du satellite Pleiades). Les

SAGEO'2016 – Nice, 6-9 décembre 2016

deux bandes spectrales sont le rouge (fin du domaine du visible) absorbées par la chlorophylle active et le Proche Infrarouge (début du domaine de l'Infrarouge) réfléchi par les végétaux en phase d'activité photosynthétique. Leur différence normalisée permet donc de transcrire en un indice simple l'importance du volume foliaire et de son activité photosynthétique à la date de prise de vue selon une échelle -1 à 1 où les valeurs négatives correspondent aux états de surface non végétalisés, où la végétation très ouverte ou peu active apparaît dans les faibles valeurs positives et où les valeurs proches de 1 correspondent à une végétation très verte et très dense. Dans un tel milieu aux pentes très fortes et aux sols sableux, l'eau ruisselle ou s'infiltrer en profondeur. Il n'y a pas de zone humide à détecter. Ainsi, la saisonnalité du NDVI offre une distinction facile des types de couvertures végétales qui reflètent eux même des types de comportements de l'eau dans le sol.

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$

Le NDVI est calculé à deux dates car les chênaies à chênes verts sont des forêts sempervirentes méditerranéennes avec un feuillage présent toute l'année mais une activité chlorophyllienne assez réduite en été donc un NDVI assez fort en hiver et assez faible en été alors que les châtaigneraies sont des forêts caducifoliées avec un feuillage présent du printemps à l'automne donc un NDVI nul en hiver et très fort en été.

La classe « roches affleurantes et sol nu » a été extraite par seuillage du NDVI à 0 aux deux dates et multiplication des deux images binaires pour ne garder que les pixels couverts de végétation active, ni en été, ni en hiver. Un calcul de fréquence locale sur un voisinage de 5 pixels sur 5 a ensuite permis de produire une carte distinguant les secteurs où les roches affleurent fréquemment et celles où la végétation et les sols forestiers dominent.

4. Méthode de segmentation d'un ensemble hétérogène d'images

4.2. Principe général

La segmentation d'image par l'algorithme issu de la morphologie mathématique dit « de la ligne de partage des eaux » (Beucher et Lantuéjoul, 1979, Beucher et Meyer, 1990) est amplement utilisée en traitement d'images. Elle sert à segmenter une image en un petit nombre de régions homogènes. Il s'agit d'une application générale (sur des images non topographiques) de la traditionnelle recherche de lignes de partage des eaux en hydrologie sur une image topographique (modèle numérique de terrain). En hydrologie, on part des points bas et l'on simule le remplissage progressif de bassins versants. Les points de rencontre des deux masses d'eaux voisines de deux bassins versants voisins sont alors enregistrés et considérés comme des lignes de partage des eaux. La simulation du remplissage se poursuit et

SEGMENTATION D'IMAGES EN GEOMORPHOLOGIE

les grands bassins versants (résultant de la fusion des petits) se remplissent eux aussi jusqu'à une autre ligne de partage des eaux d'un niveau hiérarchique supérieur. Il est également possible de le calculer par la simulation de la trajectoire naturelle de l'eau dans le sens de la plus grande pente depuis chaque pixel de l'image et où chaque pixel de l'image est ensuite attribué au talweg vers lequel l'eau simulée a coulé mais c'est la méthode par remplissage qui est ici utilisée (Beucher et Lantuéjoul, 1979).

En segmentation d'image générale pour la recherche des régions homogènes, on calcule d'abord la variance locale de l'image. Ensuite, c'est sur l'image de la variance qu'est appliqué l'algorithme de la ligne de partage des eaux par remplissage en la considérant comme une surface topographique avec ses points bas et ses points hauts c'est-à-dire ses « talwegs » et ses « lignes de partages des eaux ». Les points bas du « relief » sont donc les zones homogènes de l'image (faible variance locale), les lignes de partage des eaux sont les points de plus forte variance entre deux zones homogènes. Il est ainsi possible de tracer automatiquement la ligne de délimitation entre les zones homogènes voisines, donc de segmenter toute une image puisque les limites nettes, et les légères solutions de continuités sont repérées.

Il est possible de calculer la variance de plusieurs images dans une approche multifactorielle. Dès lors, c'est la moyenne pondérée des différentes variances des images choisies (et pondérées) par l'utilisateur qui est utilisée comme relief. Il est aussi possible d'insérer dans les descripteurs des analyses texturales qui permettent d'intégrer les formes de micro-hétérogénéité dans le calcul et de les considérer à l'échelle des segments comme des descripteurs présentant à l'échelle plus petite une certaine forme d'homogénéité.

4.2. Application de la segmentation

La variance locale sur une fenêtre coulissante de 9 pixels (3x3) a été calculée sur les 7 descripteurs. Une moyenne des 7 images de variances a été calculée en doublant la pondération de la variance de la pente considérée comme le facteur principal de l'hydrologie de versant dans ce site d'étude et à cette échelle. Sur la base de cette image de la moyenne pondérée des 7 descripteurs, la segmentation par la ligne de partage des eaux est calculée pour découper les bassins versants en plusieurs milliers de petites unités homogènes de quelques centaines de mètres carrés.

SAGEO'2016 – Nice, 6-9 décembre 2016

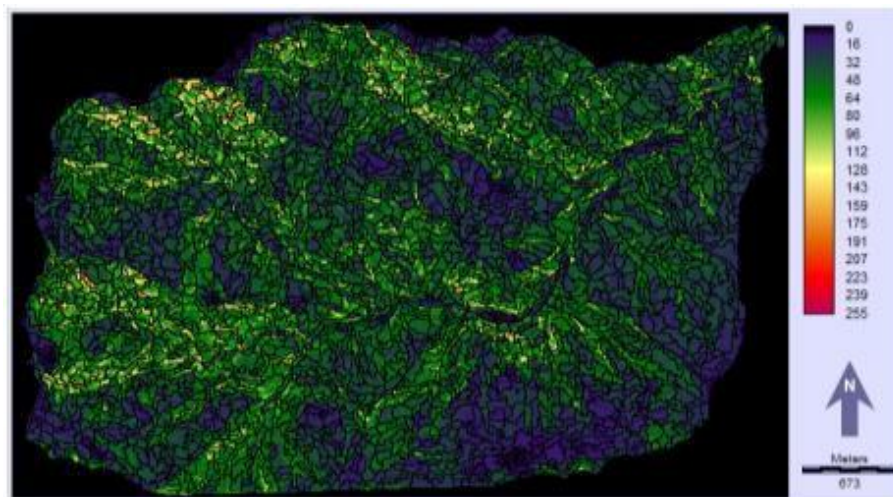


FIGURE 2. Segmentation du Bassin versant représenté sur fond de carte des pentes

L'intérêt, en géomorphologie, hydrologie ou pédologie réside dans une possibilité de réduire un ensemble d'images à échelle fine et à maille régulière (MNT, carte des pentes, carte de l'orientation, etc.) en une image distinguant des segments homogènes dans les valeurs de leurs descripteurs, ce, bien évidemment, en maintenant le géo-référencement. Il est important de dire qu'une certaine liberté dans le paramétrage permet d'une part de pondérer le niveau d'exigence de l'homogénéité des segments. Le calcul débute par une sur-segmentation très importante et nécessite le regroupement des segments. Ce regroupement progressif des bassins versants élémentaires en grands ensembles de niveaux hiérarchisés permet, D'autre part, à l'utilisateur de paramétrer l'échelle de segmentation voulue depuis les tout petits segments très homogènes jusqu'à de grands ensembles plus clairement composites bien que toujours délimités automatiquement en fonction des lignes de forte hétérogénéité entre deux zones plus homogènes. C'est par le facteur de ces regroupements que l'utilisateur choisit l'échelle du résultat obtenu.

Ces segments peuvent correspondre à des réalités géomorphologiques telles que des versants, ou autres unités élémentaires du relief.

SEGMENTATION D'IMAGES EN GEOMORPHOLOGIE

4.3. Réalisation de la typologie

L'algorithme des Nuées Dynamiques (Diday, 1973) permet de classer une population (ici les pixels des 7 images décrivant le bassin versant pour les 7 descripteurs) en un nombre de classes défini par l'utilisateur. Cela permet de créer des profils géomorphologiques homogènes (ensembles cohérents caractérisés par des valeurs proches pour les différents descripteurs). La Figure 2 présente un exemple de classification des 7 descripteurs en 6 classes sur le bassin de Tourgueille.

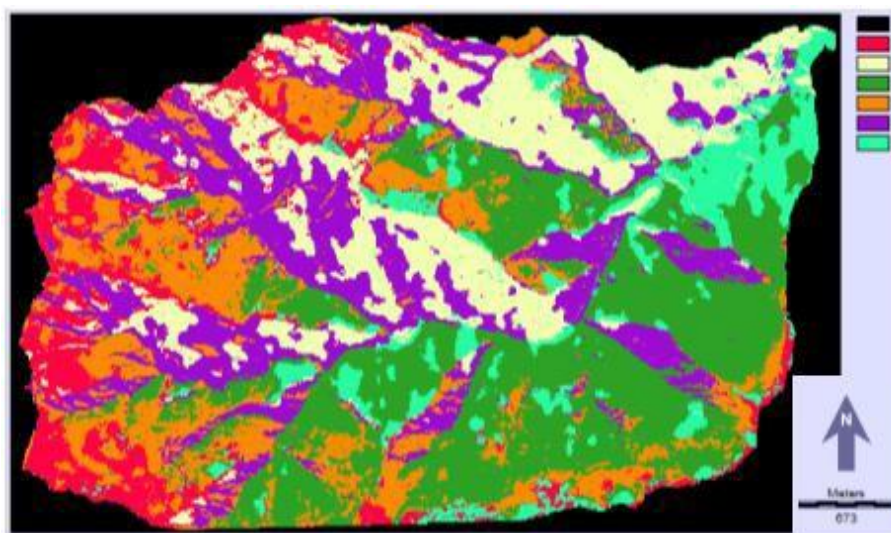


Figure 3 : Classification des pixels en 6 classes en fonction des 7 descripteurs utilisés dans la segmentation.

Cette étape intermédiaire n'est pas à une échelle adaptée pour l'étude des versants, même si dans ce cas d'étude et à cette résolution d'illustration cela apparaît peu, une multitude de pixels appartenant à une certaine classe se retrouvent dans un environnement homogène d'une autre classe et d'autres secteurs sont totalement hétérogènes.

Pour passer de la classification des pixels à la typologie de versants, l'étape de classification par pixel est suivie de l'étape de segmentation. Pour cela, la classe majoritaire de chaque versant (segmenté dans l'étape 1) est attribuée à l'ensemble du segment/versant (Figure 3)

SAGEO'2016 – Nice, 6-9 décembre 2016

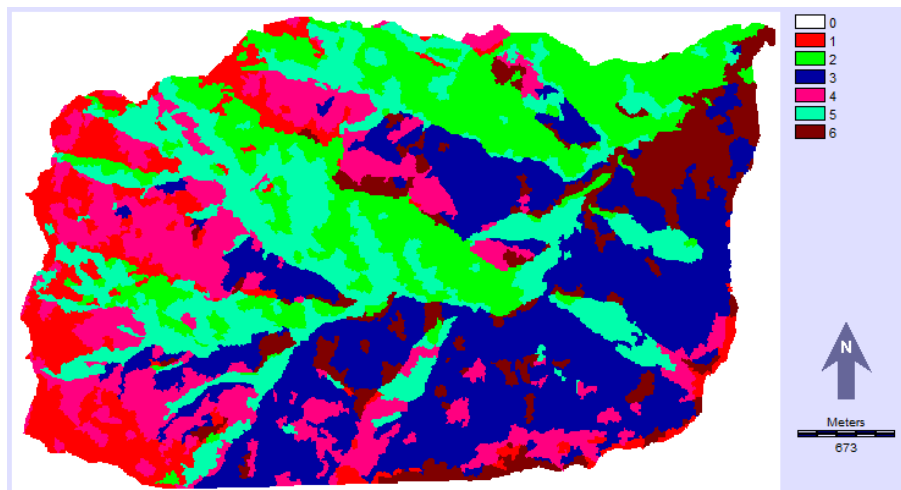


Figure 4 : Typologie des versants de Tourgueille en 6 types.

5. Résultats

5.1. Description des 6 types de versants du bassin versant de Tourgueille

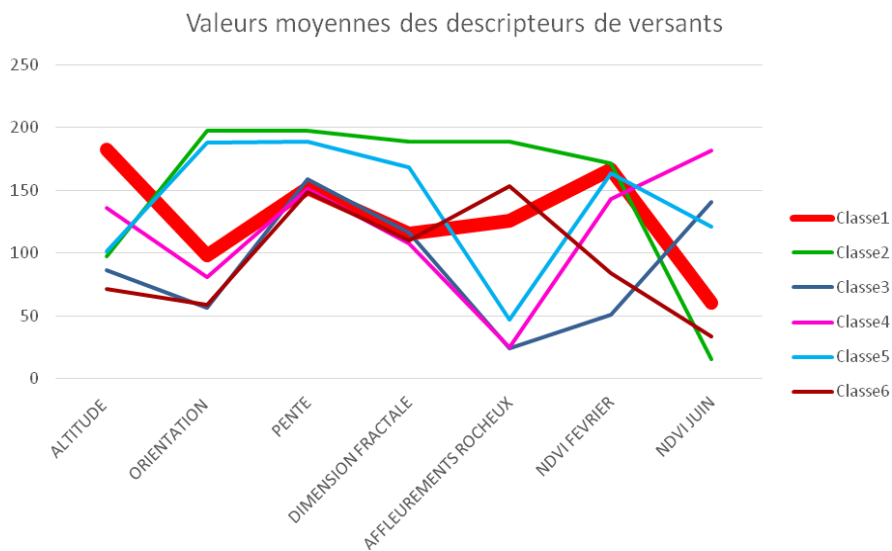


Figure 5 : Valeurs moyennes des descripteurs pour les 6 classes issues de la segmentation du bassin de Tourgueille

SEGMENTATION D'IMAGES EN GEOMORPHOLOGIE

La classe 1 est caractérisée par des altitudes élevées des orientations est ou ouest des pentes moyennes, une dimension fractale faible, quelques affleurements rocheux un NDVI élevé en février mais faible en juin. Ce profil correspond aux lignes de crête à l'amont.

La classe 2 est caractérisée par des altitudes basses des orientations sud des pentes très fortes, une dimension fractale très élevée, de très importants affleurements rocheux, un NDVI élevé en février mais très faible en juin. Ce profil correspond aux versants rocheux abrupts avec des taches de maquis à bruyère arborescente de l'adret.

TABLE 2. *Descripteur des 6 classes issues de la segmentation sur le bassin de Tourgueille*

TOURGUEILLE		Altitude	Orientation	Pente	Dimension fractale	Fréquence locale roche	NDVI 02	NDVI 07
Classe 1	Min	424	0	0	2	0	-1	-1
	Moyenne	1003	71,6	31,1	2,08	18	0,39	0,75
	Max	1223	176	80,4	2,44	37	0,97	1
Classe 2	Min	409	1	0,1	2	0	-1	-1
	Moyenne	719	137	39,9	2,14	26	0,44	0,03
	Max	1187	180	89,9	2,55	37	0,91	0,93
Classe 3	Min	415	0	0,5	2	0	-1	-1
	Moyenne	684	41	32,6	2,08	4,6	-0,44	-0,46
	Max	944	177	80,2	2,45	29	0,82	0,79
Classe 4	Min	422	0	0,1	2	0	-1	-1
	Moyenne	851	58,4	31	2,08	4,9	0,22	-0,32
	Max	1207	176	79,2	2,38	30	1	1
Classe 5	Min	415	0	0,1	2	0	-1	-1
	Moyenne	735	129	38,8	2,13	7,9	0,39	-0,53
	Max	1202	179	87,4	2,59	33	-1	0,88
Classe 6	Min	409	0	0,1	2	0	-1	-1
	Moyenne	636	43,2	31	2,08	22	-0,19	-0,8
	Max	948	179	77,1	2,36	37	0,8	0,77

SAGEO'2016 – Nice, 6-9 décembre 2016

La classe 3 est caractérisée par des altitudes basses des orientations nord des pentes assez fortes, une dimension fractale faible, aucun affleurement rocheux, un NDVI très faible en février mais élevé en juillet. Ce profil correspond aux forêts de hêtres des versants nord à l'amont du bassin versant.

La classe 4 est caractérisée par des altitudes élevées des orientations est ou ouest des pentes assez fortes, une dimension fractale faible, aucun affleurement rocheux, un NDVI élevé en février et très élevé en juin. Ce profil correspond aux forêts de mixtes à l'amont du bassin versant.

La classe 5 est caractérisée par des altitudes moyennes des orientations au sud des pentes très fortes, une dimension fractale faible, d'importants affleurements rocheux, un NDVI faible en février et très faible en juin. Ce profil correspond aux forêts de chênes verts à l'amont et en adret du bassin versant.

La classe 6 est caractérisée par des altitudes très faibles des orientations au nord des pentes moyennes, une dimension fractale forte, très peu d'affleurement rocheux, un NDVI très élevé en février et assez élevé en juin. Ce profil correspond à l'ubac en aval.

Discussion

L'application à la typologie de versants dans les deux bassins versants des Cévennes permet bien d'identifier une typologie de versants avec des caractéristiques spécifiques mais homogènes au sens des critères retenus. Cet article ne présente pas de calcul d'une marge d'erreur ou de justesse ni d'évaluation quantitative de la précision. En effet, la typologie est, en soi, une construction intellectuelle subjective sans réalité factuelle permettant d'en évaluer la valeur. De plus le choix est justement d'une typologie basée sur une classification non dirigée qui recherche mathématiquement la cohérence des classes en laissant s'exprimer les distances euclidiennes des pixels dans l'espace statistique étudié. Ainsi, ne s'agissant pas de classes ayant une pré-définition issue d'une « réalité-terrain » il n'est pas pertinent d'en développer une confrontation au terrain par classes. Enfin, les hydrologues et ingénieurs d'étude possédant une bonne connaissance du terrain ont pu lors des réunions du programme exprimer une certaine confiance dans les classes ainsi produites. Cependant, une discussion plus approfondie sur la plus value de cette méthode par rapport aux méthodes classiques de divisions en sous bassins versants sera développée.

Néanmoins, les classes obtenues sont spécifiques de chaque sous bassin et il est donc difficile de comparer les résultats obtenus sur les deux bassins. On pourrait aussi discuter le choix du nombre de classes ou des descripteurs utilisés et faire une analyse de sensibilité des résultats à ces choix. Se pose aussi la question de l'utilisation de ces résultats sur d'autres bassins, éventuellement plus grands. La méthodologie est applicable à de plus grands territoires, éventuellement avec des

SEGMENTATION D'IMAGES EN GEOMORPHOLOGIE

données ayant une résolution dégradée. Ceci peut être compatible avec la modélisation hydrologique à différentes échelles où le maillage est en général plus raffiné sur de petits bassins que sur de grands bassins, mais toujours avec une recherche d'homogénéité dans la réponse hydrologique (e.g. Dehotin et Braud, 2008). Dans ce cas, il convient néanmoins de rester attentif à la prise en compte des éléments du paysage de petite dimension mais qui peuvent avoir un impact significatif sur la réponse hydrologique (comme par exemple les ouvrages, les fossés, la voirie).

Conclusion

La typologie de versants a bien été réalisée à l'échelle des processus hydro-pédologiques de versant dans deux bassins versants. De simples descripteurs des versants ont été produits par des traitements d'images assez communs. Ils ont été soumis à une segmentation multi-images pondérée divisant chaque Bassin versant en quelques milliers de segments ou versants. Ils ont été soumis à une classification non dirigée par pixels. La classe dominante de chaque segment a été attribuée à l'ensemble des pixels du segment (ou à l'ensemble du segment en format vectoriel).

Cette chaîne de traitements d'images peut être appliquée à toute autre jeu de données quantitatives (ou qualitatives transposée en termes de fréquence locale ou autre indice de texture) à toute échelle dès lors qu'une segmentation en région homogène est recherchée et que ces segments sont ensuite réduits en types.

Bibliographie non numérotée et références

Ayral, P.-A., Sauvagne-Lesage, S., Bressand, F., Contribution à la spatialisation du modèle opérationnel de prévision des crues éclair ALTHAÏR. *Ét. Géogr. Phys*, n°32. p.75-97.

Beucher, S. et Lantuejoul, C. (1979) Use of watersheds in contour detection. In *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*, Rennes, France., September 1979.

Braud, I., Ayral, P.-A., Bouvier, C., Branger, F., Delrieu, G., Le Coz, J., Nord, G., Vandervaere, J.-P., Anquetin, S., Adamovic, M., Andrieu, J., Batiot, C., Boudevillain, B., Brunet, P., Carreau, J., Confoland, A., Didon Lescot J.-F., Domergue, J.-M., Douvinet, J., Dramais, G., Freydier, R., Gérard, S., Huza, J., Leblois, E., Le Bourgeois, O., Le Boursicaud, R., Marchand, P., Martin, P., Nottale, L., Patris, N., Renard, B., Seidel, J.-L., Taupin, J.D., Vannier, O., Vincendon, B., Wijbrans, A., (2014). Multi-scale hydrometeorological observation and modelling for flash-flood understanding, *Hydrology and Earth System Sciences Discussions*, vol. 11, p.1871-1945.

Eastman, J. R., (1985). Single-Pass Measurement of the Fractal Dimension of Digitized Cartographic Lines: In *Annual Conference of the Canadian Cartographic Association*, 1985.

SAGEO'2016 – Nice, 6-9 décembre 2016

Dehotin, J. Braud. I. (2008). Which spatial discretization for distributed hydrological models? Proposition of a methodology and illustration for medium to large-scale catchments. *Hydrology and Earth System Sciences Discussions*, vol. 12 n°3, p.769-796.

Diday, E., (1973). The dynamic clusters method in nonhierarchical clustering, *International Journal of Computer & Information Sciences* vol. 2, n°1, p. 61–88.

Flügel, W.-A. (1995). Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrol. Process.* Vol. 9 n°3–4, p.423–436.

Meyer, F. Beucher S., (1990). Morphological segmentation *J Vis Commun Image*, Vol. 1, p. 21-46

Massari, C. Brocca, L. Moramarco, T. Tramblay, Y. Didon Lescot, J.-F. (2014) Potential of soil moisture observations in flood modelling: Estimating initial conditions and correcting rainfall, *Advances in Water Resources*, Vol. 74, p. 44-53

Pumain, D., Saint-Julien, T. (2010). *Analyse spatiale: Les localisations Armand Colin* 192 p.

Seijmonsbergen A.C., Andersa N.S., Boutena W. (2012) Geomorphological change detection using object-based feature extraction from multi-temporal LIDAR data. *Proceedings of the 4th GEOBIA*, May 7-9, 2012 - Rio de Janeiro - Brazil. p.484

Classification Rules for forest fires loss estimation through data mining techniques. Case study of the Kroumiria region, (Northwestern Tunisia).

Ahmed Toujani^{1,2}, Hammadi Achour^{1,3}, and Sami Faiz²

1. LRSP Laboratory, Silvo-Pastoral Institute of Tabarka, University of Jendouba, Jendouba, Tunisia.

2. LTSIRS Laboratory, National Engineering School, University of Tunis El Manar Tunis, Tunisia.

3. Unity of Geomatics and Geosystems, University of Mannouba, Tunis, Tunisia.

ABSTRACT. In the present study, we present a method integrating Geographic Information System (GIS), Neural Network (NN) and Decision Tree (DT) techniques for forest fires loss estimation in the Kroumiria Mountains (Northwestern Tunisia). Over the last three decade, Kroumiria has experienced major fires, making them the main loss reason of Tunisian forested areas. Forest fire data are retrieved from three different sources: (i) European Forest Fire Information System (EFFIS), (ii) Terra and Aqua Moderate Resolution Imaging Spectrometer (MODIS) active fire products and (iii) national forest inventories. The database includes 1246 fire spots recorded during 30 years (1985–2015). Spatial factors encompassing topo-morphological variables, vegetation combustibility and anthropogenic parameters are determined and associated for each fire occurrence. Two types of NN are chosen and applied. A Kohonen map, common method for data dimensionality reduction, is employed to cluster spatial variables into 12 groups. For each group, a multilayer perceptron (MLP) is used to classify the burned area sizes into four categories notably: limited, considerable, large and massive burned area. Spatial rules are extracted using C5.0 decision tree (DT) from each trained MLP. Finally, a set of pruned rules are converted into fire hazard map using GIS overlay tool. Results show that the classification accuracy is significantly enhanced using Kohonen pre-processing. Spatial rules derived from the trained MLP using C5.0 DT present a mean accuracy of 83%. While the MLP, on its own, records a mean accuracy of 86%. Results from this study are quite crucial in depicting potential hotspots of fire hazard, where prevention practices can be taken in advance.

KEYWORDS: Forest fires, GIS, Rules extraction, Neural Network, Decision Tree

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

Forest covers in Tunisia are relatively limited accounting for 6.7 % of national territory. According to the Global Forest Resources Assessment report (FAO, 2013), the forest area is estimated at 1.041 million ha and its stock volume is 9 million metric tons of carbon in living forest biomass, ranking third at the scale of North Africa after Algeria and Morocco (48 and 241 million metric tons of carbon, respectively). Tunisia belongs to the Mediterranean climate zone where the forest fire season is relatively long starting in May and ending in September. According to the statistics published in European Forest Fire Information System reports (EFFIS, 2014), the countries of the southern arc of the Mediterranean Sea for the year 2009, 2135 forest fires occurred; as a result 29749 hectares of forest were burned. On July 31, 2015, an intensive forest fire occurred and devastated more than 100 hectares of maritime pine (*Pinus pinaster*) forests in Tabarka, northwest Tunisia. The proposed research has been motivated by several earlier works in the literature related to forest fires detection using spatial data, Geographic Information System (GIS) and data mining techniques. A brief description of some of the recent studies is delineated below.

Data mining techniques are widely used for forest fire prevention (e.g. Laneve and Cadau, 2007; Angayarkkani and Radhakrishnan, 2010; Yu et al., 2011). Cortez and Morais (2007) used multiple data mining techniques including multivariate regression, Decision Trees (DT), Random Forest (RF), Neural Networks (NN), and Support Vector Machine (SVM) to predict burned area for forest fires in Portuguese Natural Park. They found that the SVM model with four inputs (temperature, rain, relative humidity and wind speed) is able to predict small fires. Woolford et al. (2009) used generalized additive model (GAM) to develop a spatio-temporal prototype for forest fires caused by people in a portion of boreal forest in northeastern Ontario, Canada. Stojanova et al. (2011) proposed an approach to predict forest fires in Slovenia using different data mining techniques, weather modeling information in conjunction with vegetation modeling and GIS data. They concluded that bagged and boosted DT offer the best results in terms of accuracy. In another study by Sakr et al. (2011), forest fire prediction was implemented with NN and SVM using only meteorological data. Similarly, Ozbayoglu and Bozer (2011) have had recourse to several data mining methods including SVM, Multilayer Perceptron (MLP), and Radial Basis Function Networks (RBFN) to predict burned areas in Turkish forests. They suggested that the MLP model is the most efficient approach, providing accuracy better than a 65% success rate. Arpaci et al. (2014) compared two machine learning algorithms including maximum entropy (Maxent; Phillips et al., 2006) and Random Forest to understand the driving force factors and the actual distribution patterns of forest fires in the province of Tyrol in Australia. They concluded that both algorithms deliver similar results; climate (number of days above a certain fire weather index) and anthropogenic parameters (distance to buildings, population density) are the most important parameters explaining the current fire regime. A recent study by Pourtaghi et al. (2016) examined general indicators influencing forest fires and compared forest fire susceptibility maps based

Classification Rules for forest fires loss estimation through data mining techniques

on the boosted regression tree (BRT), GAM, and Random Forest mining models in the Golestan Province, Iran. Authors found that the annual rainfall, distance to roads, and land use factors are the main drivers of forest fire occurrence. Despite recognizing merits of DM, GIS-based multi-criteria decision analysis (MCDA) for forest fire susceptibility mapping is still the most widely used due to the increasing availability of GIS-ready data (Jaiswal et al., 2002; Dong et al., 2005; Faleh et al., 2012). However some difficulties are faced when too many fire factors are taken as they lead to a higher number of criteria requiring more complex MCDA approach. Moreover, while fire data records are invaluable without data mining, GIS systems cannot tap into their full utility like an effective predictive analysis.

In contrast with these works, we present a DM framework for fire loss estimation with emphasis on three aims: i) using non-costly and available GIS data for forest fire analyzing, (ii) predicting the burned area size by using classification DM techniques, (iii) and transforming classification results into comprehensive rules assisting in decision making. The experiment was carried out in Kroumiria region (Northwestern Tunisia), which extends for 1200 km² and is affected by intense forest fires.

2. Material and Methods

The chief purpose of this study is to integrate data mining techniques with GIS to spatially analyse the fire hazard level. This was done with two objectives: (a) to monitor the fire danger by estimating the burned area sizes and (b) to convert these classification results into explanation rules. These have been achieved through the following methodology:

1. Data preparation involves: defining fire factor parameters, extracting fire data by utilizing GIS, cleaning up the database, transforming the nominal features into numeric parameters, and normalizing the data to eliminate irregularities.
2. Partitioning the data for training and testing (with various configurations to find the best model).
3. Projecting clustering before classification stage to help in eliminating noise points and getting important data points from the large fire data set.
4. Performing classification using an appropriate ANN model for each cluster.
5. After successful training and testing, presenting each ANN classifier (for each cluster) with another fire dataset containing only the input samples and making the generalisation to get a data set with input-output mappings from each trained ANN.
6. Applying the DT algorithm to the input-output mappings to extract rules in the form of “if-then”.
7. Using DT for classification on the original dataset to compare the classification performance of ANN with DT.
9. Analysing the results and selecting the most accurate rules.
10. Converting rules into fire hazard map by using overlay technique in GIS.

SAGEO'2016 – Nice, 6-9 décembre 2016

It must be noticed that practical knowledge in forestry should be applied to indicate the significant spatial rules that could be useful for estimating the burned area sizes and their corresponding spatial patterns.

2.1. Data set

The expert model proposed by (Dagorne, 1994) has been mainly applied to Mediterranean forests. This model is based on the formula of $R=5C+2H+M$ where R iniquities the Wildfire Risk Index, C represents the Combustibility Index (factor related to fuel), H signifies the Human occupation Index (factor related to human activity) and M measures the Topographic Index. In this study, Dagorne's model will serve only to define the data features. The data set concerns relatively various spatial factors including the anthropogenic impact, the vegetation combustibility and the topomorphological conditions. The combustibility factor implies the fuel properties of surrounding vegetation. Whereas, the topomorphological factor contains three topographic parameters: aspect, slope and altitude. The anthropogenic factor comprises two distance variables: the nearest settlement and the closest road. The database is also complemented by two other parameters: the nearby lookout post and the adjacent fire-wall. These added fire factors are strongly correlated with fire incidence and spread.

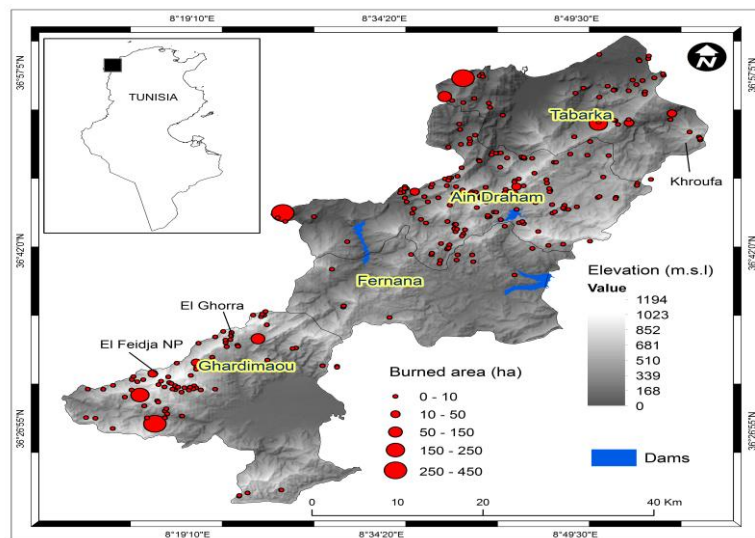


FIGURE 1. Burned areas (ha) in Kroumiria region compiled from the Tunisian General Directorate of Forestry (DGF) between 1985–2010.

The digital elevation models (DEM) and the forest inventories and fire reports (1985-2010) from the Tunisian General Directorate of Forestry (DGF) have been warehoused for this study (Figure 1). Accordingly, experts in the Kroumiria Mountains also helped to determine more precisely the coordinates of fire spots.

Classification Rules for forest fires loss estimation through data mining techniques

Despite this support, a considerable part of fire records is presented with incomplete and noisy data. Forest fire data from NASA FIRMS (1985-2014) (Fire Information for Resource Management System) and European Forest Fire Information System (2010-2015) (EFFIS) have been equally incorporated into the database in the form of vector data. The fire records span from 1985 to 2015 and constitute a complete 30-year fire record. For each fire spot, parameters of fire factors are then extracted from GIS layers (Table 1).

TABLE 1. *The pre-processed dataset attributes.*

Factors	Parameters	Description
Combustibility	Vegetation	Type of tree cover for the area where the fire is located
Topo-Morphology	Aspect	Solar exposure for fire location
	Slope	Degree of slope for fire location
	Elevation	Altitude of fire location
Anthropogenic	Road	Distance from fire location to the nearest road
	Settlement	Distance from fire location to the nearest settlement
Other variables	Look-post	Distance from fire location to the nearby lookout post
	Fire-wall	Distance from fire location to the nearby fire-wall

Lastly, the resulting datasets have been reorganized into tabular database to be more compatible with the proposed data mining algorithms. This database includes 1246 fire spots during the period from 1985 to 2015 and concerning the Kroumiria Mountains. The significance of these data is that they permit the burned area forecasting using historical records. All numeric attributes have been converted into numbers divisible by 10 (for example the elevation will be in the form of 100, 150, 200, 250, 300 meters etc.) for more visibility and efficient processing of data. The datasets are randomly arranged into two categories exploited as learning samples and validating samples. The learning samples are 80 %, and the rest is then tested and evaluated during the mining process.

2.2. Data clustering phase

Classification algorithms perform well on homogeneous data set, so clustering before classification can provide improved results. The Self-Organizing Map (SOM) clustering method has been proposed as a precursor to classification as discussed above. SOM (or Kohonen Map; Kohonen, 1997) is a popular NN and is habitually operated as a clustering tool. However it is used equally as a classification technique (Vesanto and Alhoniemi, 2000). SOM, as a rule, is trained applying unsupervised learning before generating low-dimensional discretized classes, known as a Kohonen map. The objective of SOM is to recognize correlations in their input, and similarly to identify clusters of input vectors. In the course of SOM training step, distances are computed between training samples and neurons. There are various

SAGEO'2016 – Nice, 6-9 décembre 2016

distance functions, such as Manhattan distance or Euclidean distance. The neuron with the smallest distance is the winning neuron or BMU (Best Matching Unit). The BMU updates its associated weight to be nearby to the training sample. As an NN, SOM is capable of learning from complex and large data sets without specification of the output, as desired in our case. Equally, the Kohonen algorithm is uncomplicated enough to make a learning process simple to use and to adjust. A more detailed description of SOM algorithm is given by (Giraudel and Lek, 2001). So in this phase, samples have been clustered by feature similarity into dissimilar sorts without considering any target information. Data normalization was carried out before this process using the standard deviation method, because of the different measurement units of the variables. It follows that the normalized data vectors are entered in the SOM learning process by using Tanagra software (Rakotomalala, 2005) and BMU distance. Tanagra is a free suite of machine learning software for research and academic purposes developed by (Rakotomalala, 2005).

Note that the burned area size is not included in the SOM learning process, since it is the classification results. The SOM outputs are divided into a number k of clusters determined by considering two criteria. The first one is minimizing the number of clusters to have a sufficient amount of data for accurate classification (in each SOM cluster). The second one is to use an appropriate method for optimizing number of clusters (e.g. K-selection, cross-validation or penalized likelihood estimation). Here, we prefer to decide in practice by trying k from k_{min} to k_{max} and choosing the optimal k . As a result, all SOM outputs can be then used to feed the classification module.

2.3. Classification and rules extraction process

As one common approach in data mining, artificial neural networks (NN) have attained until now the highest accuracy for various classification problems even in the situations where model is presented with incomplete and noisy data. As mentioned earlier, the imprecision in the coordinates of fire spots and missing variables in this database make the NN suitable for our purpose. However its major drawback is that NN are black box models, which are not easily interpretable (Krishna and Aramudhan, 2014). This is because the knowledge in the NN is stored as real valued parameters (weights and biases) of the network components (Andrews et al., 1995). Various researchers tried to overcome this problem by extracting rules from trained NN. There are two main approaches for knowledge extraction from trained NN: (a) extracting “if-then rules” by clustering the activation values of hidden state neurons (Fu, 1994) and (b) applying machine learning methods such as DT to the input-output vectors of the NN after its training (Boz, 1995).

DT can be straightforwardly represented in the form of “if-then rules” and hence extracting DT is one of the best methods of interpreting an NN (Fuller et al., 2011). A DT is a set of rules handled to classify data into categories. It looks at the variables in a data set, determines which are most important, and then comes up with a, well, tree of decisions which best partitions the data. The tree is formed by splitting data up by variables and then counting to see how many are in each bucket

Classification Rules for forest fires loss estimation through data mining techniques

after each split. DT is so not a black box and its results are easily interpretable (Wu, 2007). However, it is difficult for DT to handle effectively large data sets with missing values and continuous variables (Wu, 2007). So, with the goal of generating more accurate and comprehensible rules, extracting information from trained NN in symbolic form (if-then rules) could be an appropriate solution using DT. This method offers a way to overcome both the limitation of decision tree and the weakness of neural networks being black boxes. Thus, the target of this process is to extract knowledge from the weights of NN in “if-then” form. After completing the NN training step, DT model is employed for extracting rules from trained NN. For each SOM cluster, the extracted rules will explain the classification of burned area sizes according to their spatial factors. In order to ensure the efficiency of this process, the classification performance of the extracted rules is compared with that of DT alone. Lastly, accurate and generic rules are selected for the process of fire susceptibility mapping.

As an NN algorithm, a Multilayer Perceptron (MLP), for each cluster, is trained and used to forecast the burned area sizes. An MLP contains typically 3 layers including an input layer, a hidden layer and an output layer. Number of input neurons in an input layer is equal to the number of input features in the input vectors. Therefore, the employed MLP model has 8 inputs equal to the number of predictor variables (8 parameters for fire factors). The hidden layer is the group of neurons which has activation function applied on it as well as provide a transitional layer between the input layer and the output layer. There is no exact rule to specify the number of neurons in the hidden layer. Determining the number of neurons in the hidden layer is done by trial and error. Output layer has the output neurons where the number of neurons is same as the number of desired output variables. Here there is one output neuron to compute the burned area size. The back-propagation algorithm is often used to train an MLP network (Haykin, 1994). During the learning, the weights are adjusted based on a set of desired outputs. Each back-propagation iteration contains two sweeps: forward activation to compute a solution and backward propagation of the computed error to adjust the weights. These forward and backward sweeps continue until the MLP solution agrees with the desired value within a pre-specified tolerance. For this classification step, Matlab R2014a® is employed to train the MLP. The output form is an unthresholded linear unit since the burned area is in continuous value, whereas hidden units are sigmoid. Each MLP is learned with the training data of the corresponding SOM cluster. In total, this approach employs k MLP models corresponding to the number of SOM clusters. In order to use an efficient symbolic representation of rules, the burned area has been converted from continuous to categorical form, namely; limited, considerable, large, and massive. For this, z-score transformation with standard deviation normalization method has been applied. The empirical rule asserts that 2/3 (68%) of data lies within one standard deviation unit of the mean, 95% lies within two standard deviation units of the mean, and 99.7% lies within three standard deviation units of the mean. Using standard deviation and zero mean, the transformation is described in the following rules: “If $|\text{normalized burned area}| < 1$, then it is limited; else if $1 \leq |\text{normalized burned area}| < 2$, then it is considerable; else if $2 \leq |\text{normalized burned area}| < 3$, then it is large; else it is massive”.

SAGEO'2016 – Nice, 6-9 décembre 2016

After training, the testing samples are clustered by Kohonen map and fed to the relevant MLP for predicting the burned area size. At this classification step, inputs are associated with outputs. These input-output mappings provide the datasets that will be operated by C5.0 to extract rules in the form of DT. In this step, C5.0 DT is used to create an “if-then model” based on the ID3 (Interactive Dichotomize) (Quinlan, 1993). This method is based on information theory. The ID3 algorithm employs entropy as a benchmark to select the fitting values for branching at the node. In comparison to its predecessors DT methods (ID3 and C4.5), C5.0 creates smaller trees and is therefore more efficient in memory-usage and execution-time (Wu, 2007). In addition, it tends to be easier to use and to understand, since its derived rules have straightforward interpretation. Similarly, it offers the boosting method to improve the model accuracy (Zheng, 1996). A final decision step aims to evaluate the different sets of rules and to find out the most accurate possible. This will be done by conducting C5.0 DT algorithm on data generated by NN. Classification performances are then compared between the two models and for each cluster separately. Accurate forecasting rules are, so, selected and pruned before employing them for the phase of forest fire hazard evaluation.

2.4. Forest fire hazard mapping procedure

After selecting the most accurate rules from NN and (or) DT for each cluster, these spatial patterns have been pruned by removing redundancy and intersecting rules that describe the same class of burned area size. The spatial rules can be summarized into fire hazard map characterizing the fire loss estimation. Based on this map, fire factors and their immediate impacts on the burned surface after fire occurrence are thus determined. For example, an “if-then rule” that looks like this: “if fires occur in areas with (300 m < elevation) and (15 < slope < 20) and (distance to settlement > 2000m), then the burned area will be massive”, can be translated into map layers displaying areas that cover these 3 factors as serious hazard.

This is done with the use of GIS, fire factors are firstly displayed in vector layers. All factor layers are then intersected using buffering and overlay techniques of GIS. The overlay procedure permits to compute the geometric intersection of vector layer inputs. Portions of features that overlap in all input layers are then retained in the output map. In summary, the GIS has been used in the final phase for: (a) developing the thematic maps of fire factors and (b) intersecting these maps by using the overlap technique to highlight fire vulnerabilities.

3. Results and discussion

The results from SOM pre-processing phase are summarised in table 2. Cluster 6 comprises the largest number of training sets (103 samples) and has a ratio of 15.53% of testing sets over training sets. The maximum value of this ratio is obtained in cluster 4 with 47.61%. Then, in accordance with the number of SOM clusters, 12 back-propagation MLPs (BP-MLP) have been employed. Each input neuron corresponds to one parameter of the spatial factors. The output neuron

Classification Rules for forest fires loss estimation through data mining techniques

represents the burned area class. In clusters that include a large number of samples such as cluster 5 and 6, trial experiments have been running with a fixed learning rate of 0.5 and different combinations of learning and testing datasets. The BP-MLP is composed of three layers; an input layer with 8 neurons equal to features (spatial fire factors), a hidden layer with 4 neurons, and the output layer is composed with one neuron for estimating one category of burned area size. The Number of samples was reduced via clustering to enable the net to train on data with missing entries (features). Therefore, it is necessary to reduce the MLP structure which makes one hidden layer sufficient for learning in clusters with a small number of observations. The hidden neurons are then 4 equal to $n/2$ where n equals to the fire factors (8 inputs). The classification result is computed and evaluated in comparison with direct application of MLP (without SOM clustering). The neural network toolbox of Matlab R2014a® software is used to implement this process.

TABLE2. Resulting clusters after the SOM learning. LS is the Learning Sets, and TS is the Testing Sets.

SOM outputs	LS	TS	TS/LS (%)	SOM outputs	LS	TS	TS/LS (%)
Cluster 1	67	13	19,40	Cluster 7	97	27	27,83
Cluster 2	93	34	36,55	Cluster 8	92	23	25
Cluster 3	87	12	13,79	Cluster 9	85	13	15,29
Cluster 4	42	20	47,61	Cluster 10	91	24	26,37
Cluster 5	101	24	23,76	Cluster 11	66	25	37,87
Cluster 6	103	16	15,53	Cluster 12	73	18	24,65
				Total	997	249	20

As shown in table 3, in terms of accuracy, results prove that MLP on its own is the best classifier. Then, the DT model has been exploited for extracting rules from each trained BP-MLP using the C5.0 algorithm from the package C50 in the R Environment (R Development Core Team, 2011). During this step, all training and testing data have been employed with only considering the classification label generated from trained BP-MLP. Here, cluster 2 is a typical example to show the accuracy of rules extracted from it. It contains the largest data samples (93 samples for training and 34 samples for testing) and has an accuracy level of 89% exceeding that of total mean accuracy (86%) and also the accuracy of direct forecasting (68%). From these results it is evident that applying clustering in the preprocessing phase has the potential to increase the classification accuracy, especially if missing data or irrelevant attributes are present.

DT have been similarly applied on the original data set without previously having passed through trained NN. This second classification allows evaluating the performance of our approach and selecting more accurate rules. The C5.0 DT algorithm uses 80% of data as training and testing on 20%. Despite the cases where DT are more precise particularly in three clusters (clusters 1,3 and 3 as in Table 3); the NN provides significantly better performance.

SAGEO'2016 – Nice, 6-9 décembre 2016

TABLE 3. Summary of results in percentage. MLP-C5.0 Accuracy is the forecast accuracy of extracted rules from trained MLP using C5.0 DT. C5.0 Accuracy is the forecast accuracy of extracted rules from original dataset using C5.0 DT. Direct forecasting measures the forecast accuracy without SOM clustering. Accuracy mean indicates the arithmetic mean of forecast accuracy of all clusters.

Clusters	MLP Accuracy	MLP-C5.0 Accuracy	C5.0 Accuracy	Clusters	MLP Accuracy	MLP-C5.0 Accuracy	C5.0 Accuracy
Cluster 1	90	83	91	Cluster 8	87	89	76
Cluster 2	91	89	85	Cluster 9	86	73	81
Cluster 3	82	78	91	Cluster 10	89	85	73
Cluster 4	88	94	84	Cluster 11	85	87	79
Cluster 5	95	100	82	Cluster 12	93	93	86
Cluster 6	86	78	86	Accuracy mean	88	86	83
Cluster 7	95	93	83	Direct forecasting	79	68	77

TABLE 4. Pruned and selected spatial rules from trained MLP or DT C5.0. (--) means that no condition has been detected with the corresponding spatial variable for the designated rule.

Clusters	Rules	Slope (%)	Elevation (meters)	Aspect	Nearby road (meters)	Closest settlement (meters)	Forest vegetation	Adjacent fire-wall (meters)	Nearby look-out post (meters)	Burned are size
7,9,12	Rule 1	<5	_	_	<100	<500	_	<200	<1000	Limited
5,7	Rule 2	>14	_	_	_	<500	_	>1000	_	Large
2,1,3,9,10	Rule 3	_	>650	_	>1000	_	_	>1000	>2000	Large
2,6,8,9	Rule 4	>8	>450	_	<2500	<500	_	>1000	_	Considerable
4,8,12	Rule 5	_	>650	_	>1000	<5000	_	>2000	>2000	Large
10,3	Rule 6	>14	>850	--	>5000	<2000	_	>4000	_	Massive
11,5	Rule 7	--	>350	--	<1000	<500	--	--	--	Limited
4,6,12	Rule 8	--	>350	--	<500	<5000	--	--	--	Considerable

Accurate rules are then used for the phase of forest fire hazard mapping. Table 4 summarizes selected set of rules that maximize the model accuracy. After that these information patterns are translated into visual patterns in fire hazard maps for assessing the hazard according to the estimation of burned area size. As an illustrative example, the following rule (Rule 6) derived, from cluster 3 and cluster 10, is defined: "If fires occur in areas with (distance to roads > 5000 m) and (slope >= 15%) and (elevation > 850 m) and (distance to settlements < 2000 m) and (distance to fire-wall > 4000 m) then the burned area size is massive". This spatial pattern implies a high risk level, caused by 5 conditions. This fire hazard pattern is therefore mapped in two stages. Item conditions are firstly characterized into 5

Classification Rules for forest fires loss estimation through data mining techniques

thematic layers (vector data). Then, these maps are intersected by using the overlap tool of ARCGIS 9.0®. The resulting map indicated zones where there is a high probability that burned area could be massive. The Fire hazard map in Figure 2 characterizes the class of large burned area. This map provides a pertinent decision support to identify the implementation priorities for hazardous zones where burned areas could be large when fires occur into it.

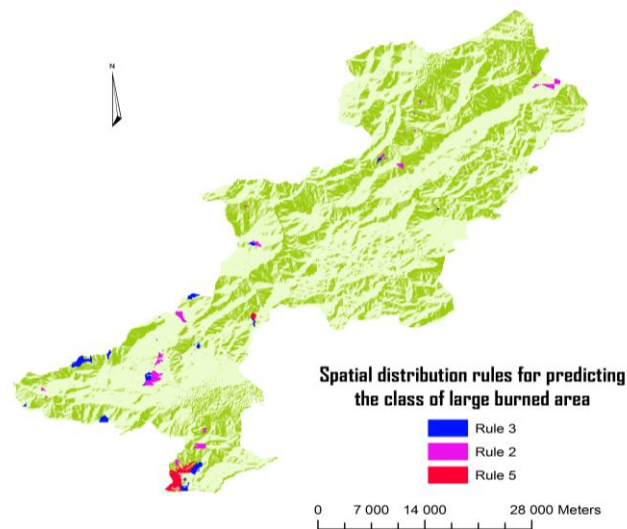


FIGURE 2. Fire hazard map of the class of large burned areas.

From the generated rules (table 4), it is noticed that the combustibility factor has not been patterned. This is, in fact, justified by the intensification of the flammability degree for the Mediterranean trees in summer (Velez, 1990). The difference is accordingly not significant between the combustibility characteristics of highly flammable species during the year (e.g. *Pinus halepensis* or *Quercus canariensis*) and highly flammable species only in summer (e.g. *Pinus pinaster* or *Quercus suber*). This is due to the climatic conditions that are stable during summer (dry season, no precipitation, infrequent variations in the daily minimum and maximum temperature, an almost constant level of humidity, etc.) when the majority of fires occur and the different species have convergent high levels of combustibility and flammability. Likewise the aspect factor has not been detected by the mining process. Correspondingly, it signifies that other spatial parameters have more influence than aspect. This supposition has been attested in GIS studies for forest fire hazard modelling by assigning less weights and ratings to the aspect factor (Jaiswal et al. 2002). The forest fire statistics for the Kroumiria region show that the burned surfaces have been reduced to one-third despite the increased frequency of fire occurrences after 1990 (Boussaidi, 2009). Here, this expert conclusion is reconfirmed and proves the strong human impact in fire prevention. In fact, this result can be justified by the increased population density in the Kroumiria Mountains.

SAGEO'2016 – Nice, 6-9 décembre 2016

In comparison to common methods, GIS applications for fire hazard modelling are often based on satellite images that are difficult to be simply implemented in real time, which is achievable by the proposed method that employs only available vector data. Also, the hazardous areas described by these GIS methods are very large, which make the detection of fires a hard task. For example, Faleh et al. (2012) have characterized more than 66 % of the studied forest as an area of considerable hazard. On the contrary, it is doable using the proposed approach. As an example from the case study, the class of large burned area has been mapped in a surface of 2966.0123 ha that occupy only 1.7% of the studied forest, which makes monitoring possible.

4. Conclusion

In this study a spatial data mining framework has been applied for forest fire prevention. The proposed approach works in three phases: Estimating forest fire loss using NN, extracting spatial rules from trained NN and exploiting these fire patterns for the forest fire susceptibility mapping. The BP-MLP has been successful in the classification of burned area sizes according to their spatial fire factors namely: human activity, fuel and anthropogenic conditions. As the prediction process involves the SOM clustering for reducing the data dimensionality, the prediction accuracy is promisingly increased. The combination of these two neural network algorithms has reduced the negative impact of incomplete and noisy data on classification accuracies. The C5.0 DT model has been convenient for rule extraction from trained BP-MLP. The results ascertain that BP-MLP on its own has shown the best performance, and its combination with C5.0 DT can approximately keep its given accuracy level. Thus, we assume that NN and DT models are crucially able to resolve the task of fire loss estimation and they can be useful to prevent other natural hazards where many incomplete and noisy spatial data are present.

To our knowledge, this is the first time the burned area size is predicted in the form of spatial rules using only available GIS data and fire records. In parallel with giving the possibility to transfer this kind of classification rules into hazard maps, some other novel techniques for forest fire susceptibility mapping could be employed to improve fire fighting resource management. In fact, the zonation of an estimated burn area, based on spatial factors, provides an essential cue for minimising fire losses. A short-term management would be better directed such as resource allocation and spatially when simultaneous fires occur. Furthermore, the long-term management would be improved such as creating firewalls and planting resistant species in areas vulnerable to massive fires.

This spatial DM method could be more effective if other fire factors will be taken into account. For example climatological data could be tested with data mining algorithms for temporal reasoning. Also, the method could be extended with integrating more efficient pre-processing techniques like feature selection algorithms, in order to further increase the prediction accuracy.

Classification Rules for forest fires loss estimation through data mining techniques

Acknowledgements:

This work is part of a program of research supported by the laboratory of Silvopastoral resources (Silvopastoral Institute of Tabarka). The authors would like to thank the researcher Abdelaziz Hamdi for helping in the identification of fire locations, and the forest engineers in Tunisian general directorate of forestry for providing fire reports and forest inventories.

References

- Andrews R., Diederich R., Tickle A., (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge Based System*, 8, pp. 373-389.
- Arpaci A., Malowerschnig B., Sass O., Vacik., H 2014. Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. *Applied Geography* 53: 258-270.
- Boussaidi N. (2009). *Impacts de l'action anthropique sur la subéraie tunisienne: essai de projection dans le futur d'un écosystème*. PhD thesis, Agronomic Sciences, National Agronomic Institute of Tunisia, Tunis, Tunisia, pp. 226.
- Boz O. (1995). *Knowledge integration and rule extraction in neural networks*. EECS Department, Lehigh University.
- Cortez P. Morais A. (2007). A data mining approach to predict forest fires using meteorological data. In *Proceedings of the 13th Portuguese Conference on Artificial Intelligence*, 512-523.
- EFFIS, (2014). *Forest Fires in Europe, Middle East and North Africa 2013*, EUR 26791 EN, Publications Office of the European Union, Luxembourg, p. 107.
- Dagorne D. (1994). Protection des forêts contre les incendies et SIGs. *Revue forêt méditerranéenne t. XV, n4*.
- Faleh A., Lakhouaja H., Sadiki A., Chaaouan J., (2012). Sig, teledetection et evaluation de risque d'incendie de forets : exemple du massif forestier khezana (province de Chaouene). *Papeles de geografia, ISSN-e 1989-4627, N° 55-56, 2012 :37-48*.
- FAO, (2013). *State of Mediterranean Forests (SoMF)*. Rome, Italy, 194 p.
- Fu L. (1994). Rule generation from neural networks, *IEEE Transactions on Systems, Man and Cybernetics*, 24(8), pp. 1114-1124.
- Fuller C., Wilson R., (2011). Extracting Knowledge from Neural Networks. *Encyclopaedia of Knowledge Management*: 320-330.
- Giraudel J., Lek S., (2001). A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecological Modelling* 146: 329– 339.
- Haykin S. (1994). *Neural Networks - A Comprehensive Foundation*. New York: Macmillan College Publishing Company.
- Jaiswal R., Mukherjee S., Raju D., (2002). Forest fire risk zone mapping from satellite imagery and GIS. *International Journal of Applied Earth Observation and Geoinformation* 4 pp: 1-10.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Kohonen T. (1997). Self-Organizing Maps. *Springer, Berlin, Heidelberg, 1995. Second Extended Edition.*
- Krishna R., Aramudhan M., (2014). Decision Support System Using Fuzzy Min-Max Neural Network with the Modified Genetic Algorithm," *International Review on Computers and Software (IRECOS)*, vol. 9.
- Laneve G., Cadau E., (2007). Quality assessment of the fire hazard forecast based on a fire potential index for the Mediterranean area by using a MSG/SEVIRI based fire detection system. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 2447-2450.
- NASA FIRMS (2014). *MODIS Active Fire Products*. <http://earthdata.nasa.gov/firms>.
- Phillips S., Anderson R., Schapire R., (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190, 231–259.
- Pourtaghia Z., Pourghasemib H., Aretanoc R., Semeraroca T., (2016). Investigation of general indicators influencing on forest fire and its susceptibility modeling using & different data mining techniques. *Ecological Indicators* 64: 72–84.
- Quinlan J. (1993). C4.5: Programs for Machine Learning. *Morgan Kaufmann 1993*, ISBN 1 55860-238-0.
- R Development Core Team (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. ISBN 3-900051-07-0.
- Sakr G., Elhajj I., Mitri G., (2011). Efficient forest fire occurrence prediction for developing countries using two weather parameters. *Engineering Applications of Artificial Intelligence* 24 (5), 888-894.
- Rakotomalala R. (2005). Tanagra: Un logiciel gratuit pour l'enseignement et la recherche. *In Actes de EGC'2005*, RNTI-E-3.vol-2. Pp.697-702.
- Vesanto J., Alhoniemi E., (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3):586600.
- Velez R. (1990). Les incendies de forêt dans la région méditerranéenne. *Panorama régional.Unasyuva*, 162 (41), 3-9.
- Woolford D., Stanford D., Kulperger R., Boychuk D., Wotton, B., (2009). Erlangian approximations for the transient analysis of a fluid queue model for forest fire perimeter. *INFOR* 47:309-317.
- Wu X., (2007). Top 10 algorithms in data mining. Knowledge information, 22-24. *Springer Verlag London Limited.*
- Yu Y., Omar Y., Harrison R., Sammathuria M., Nik A., (2011). Pattern clustering of forest fires based on meteorological variables and its classification using hybrid data mining methods. *Computational Biology and Bioinformatics Research* Vol. 3(4), pp. 47-52.

Évolution altitudinale à 25 m de résolution spatiale de l'isotherme 0°C des températures moyennes mensuelles hivernales dans les Alpes du Sud françaises durant le XXI^e siècle.

MARTIN Nicolas¹, CARREGA Pierre¹, ADNES Cyriel¹

1. UMR ESPACE 7300, Université de Nice Sophia-Antipolis
98 Bd. Edouard Herriot, 06204 Nice cedex 3, France
nicolas.martin@unice.fr

RÉSUMÉ. Dans les Alpes du Sud françaises les températures vont augmenter durant le XXI^e siècle. Pour savoir dans quelle mesure ce réchauffement se répercutera à fine échelle spatiale localement sur chaque domaine skiable, cet article propose d'étudier l'évolution de l'altitude de l'isotherme 0°C entre 1961-1990, 2021-2050 et 2071-2100 en fonction des scénarios A1B, B1 et A2. À l'aide des sorties climatiques d'ALADIN-Climat à 12 km de résolution une étape de downscaling statistique fondée sur la régression multiple permet de simuler les températures à 25 m de résolution. Les résultats montrent une nette remontée de 250 à 350 m d'altitude de l'isotherme 0°C des températures moyennes mensuelles de décembre à mars pour la moitié du siècle et jusqu'à 600 m pour la fin du siècle.

ABSTRACT. In the French South of the Alps the temperatures are going to increase during XXIth century. To know to what extent this reheating will echo in fine spatial scale locally on every ski slopes, this article suggests studying the evolution of the height of the isotherm 0°C between 1961-1990, 2021-2050 and 2071-2100 according to the scenarios A1B, B1 and A2. By means of the climatic outputs of ALADIN-Climat at 12 km of resolution a step of downscaling statistics based on the multiple regression allows to simulate the temperatures at 25 m of resolution. The results show a clear rise of 250 to 350 m of height of the isotherm 0°C for monthly mean temperatures from December till March for half of the century and until 600 m for the end of the Century.

MOTS-CLES : modélisation climatique, descente d'échelle statistique, températures moyennes mensuelles, stations de ski.

KEYWORDS: climate modeling, statistical downscaling, mean monthly temperatures, ski resorts.

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

Les évolutions climatiques, en cours et attendues pour les décennies à venir, inquiètent les acteurs économiques et les décideurs à l'échelle des territoires. Dans le département des Alpes-Maritimes une grande partie de l'économie repose sur le tourisme, qu'il soit estival ou hivernal. Or la hausse des températures déjà observée dans l'extrême sud-est de la France comme ailleurs dans le monde, aura nécessairement des répercussions sur l'ensemble de ces activités touristiques. L'été par exemple le confort thermique en ville, mesuré notamment par l'humidex qui combine température et humidité, pourrait augmenter ; l'hiver le manteau neigeux recouvrant le Massif du Mercantour, si précieux pour les stations de ski, pourrait malheureusement s'amincir et ne recouvrir que les versants les moins exposés au soleil et les sommets les plus élevés.

Les projections climatiques, établies grâce aux modèles globaux et régionaux, ne proposent pas aujourd'hui des résolutions spatiales suffisamment fines afin de déceler ces évolutions majeures pour le devenir des espaces. Quelles sont les trajectoires à suivre pour l'aménagement du territoire ? Quelles sont les échelles de temps ? Quels sont les investissements à faire ou à ne pas faire ? Ces questions sont actuellement posées par les décideurs et les gestionnaires à l'échelle locale mais elles ne trouvent pas nécessairement de réponses explicites. Ainsi l'enjeu est de proposer à la société des éléments tangibles afin de contribuer à une meilleure anticipation des évolutions climatiques qui permettra de mieux décider aujourd'hui pour demain. Le travail proposé ici est donc de fournir, à une résolution spatiale de 25 m, une cartographie de l'évolution de l'isotherme 0°C des températures moyennes mensuelles de décembre à avril dans la zone montagneuse des Alpes-Maritimes. Cette variable climatique, sans être bien entendu la seule à régir le comportement du manteau neigeux, constitue le facteur le plus influant sur la variabilité spatiale et temporelle de la neige.

2. Contexte et objectifs

Les stations de ski dans les Alpes ont connu plusieurs saisons catastrophiques économiquement au début des années 1990 en raison d'un enneigement insuffisant (Marty, 2013). Toutefois le manteau neigeux est caractérisé par une forte variabilité interannuelle en raison d'une occurrence variable d'une année sur l'autre des types de circulation atmosphérique favorable ou non aux chutes de neige (Scherrer, 2006). Malgré cette variabilité reconnue par tous, de nombreux chercheurs ont mis en évidence un amincissement du manteau neigeux et une réduction de la durée de présence de la neige au sol depuis environ 25 ans dans les Alpes, notamment aux altitudes moyennes en Suisse, en Autriche, en Italie et en France (Marty, 2013).

En Italie une tendance à la baisse des hauteurs de neige et des durées d'enneigement sur les 20 dernières années a été décelée à partir d'un réseau de 30 stations de mesures disposant d'observations entre 1920 et 2009 (Valt et Cianfarra,

Évolution de l'isotherme 0°C dans les Alpes du Sud durant le XXI^e siècle

2010). La hausse de températures ferait diminuer la proportion de chutes de neige dans les précipitations totales et expliquerait cette tendance (Serquet et al., 2011). Ce constat n'est pas seulement enregistré aux basses et moyennes altitudes mais également entre 2000 m et 3000 m dans une petite partie du nord de l'Italie dans 20 postes de mesures (Bocchiola et al., 2009). Dans les Alpes françaises les études montrent une baisse de l'épaisseur du manteau neigeux aux basses et moyennes altitudes, et ce signal est plus faible dans le sud du massif que dans le nord (Durand et al., 2009). Actuellement les stations de ski françaises sont économiquement viables au-dessus de 1200 m au nord et de 2000 m au sud.

Si le raccourcissement de la saison des sports d'hiver et la baisse de l'épaisseur de la couverture de neige sont observés depuis plus de deux décennies dans les Alpes, quelle sera la suite ? Les modèles globaux ou régionaux qui simulent l'évolution climatique ne proposent pas de résolutions spatiales suffisamment fines pour véritablement modéliser la variabilité spatiale de la neige. Pour surmonter cet obstacle, deux solutions existent :

- Un premier type d'approche estime les conditions d'enneigement futures en s'appuyant sur une analyse de sensibilité de la variabilité actuelle de la neige (Marty, 2013). Par exemple Hantel et Hirtl-Wiekle (2007) ont mis en valeur une baisse de 33% du nombre de jours de neige (jours avec plus de 5 cm de neige au sol) par degré de hausse de la température dans les Alpes, estimation établie sur 40 ans dans 268 postes de mesures. Une autre étude réalisée sur 666 stations de ski dans les Alpes (Abegg et al., 2007) calcule la baisse du nombre de zones skiabiles en s'appuyant sur l'hypothèse d'une remontée de l'altitude de la limite de la neige naturelle de 150 m par degré d'élévation de la température. Les simulations ont été faites pour 1°C, 2°C et 4°C de hausse. Le nombre de zones skiabiles baisserait respectivement de 25%, 40% et 70%. En fonction de l'intensité de l'élévation de la température de très nombreuses stations seraient contraintes de cesser leurs activités reposant sur l'or blanc. Selon les conclusions de cette étude, et pour le cas le plus pessimiste, seule la station d'Isola 2000 pourrait voir son activité perdurer dans les Alpes-Maritimes. Enfin une autre étude démontre à partir d'un modèle liant les précipitations aux températures dans les Alpes autrichiennes que seulement très peu de stations de ski, les plus hautes, surmonteraient l'impact d'un réchauffement de 2°C (Breiling et Charamza, 1999).

- Une seconde manière de se projeter est d'utiliser des modèles physiques à échelle spatiale très fine alimentés par des conditions météorologiques générées et potentiellement probables dans le futur pour simuler les épaisseurs de neige et la durée de l'enneigement continu au sol. Par exemple les Alpes suisses Beniston et al., (2003) prévoient, en se servant d'un modèle de couverture neigeuse, une baisse de 50% du volume de neige et une fin de saison 50 jours plus tôt à 2000 m d'altitude en supposant une hausse de la température de 4°C. Plus récemment, pour la Suisse et pour la même élévation du thermomètre, le recours à un modèle climatique régional fournissant les données d'entrée nécessaires à un modèle de simulation de la couverture neigeuse prédit une neige éparse au pied des pistes et une baisse drastique dans plus de la moitié des stations du nombre de jours avec plus de 30 cm de neige, seuil déterminant pour la pratique du ski (Uhlmann et al., 2008). Pour les

SAGEO'2016 – Nice, 6-9 décembre 2016

Alpes françaises la réduction du manteau neigeux serait de 50% en dessous de 1500 m d'altitude et la superficie couverte par la neige diminuerait de 30% au cœur de l'hiver avec seulement 1,8°C de hausse de température et à précipitation constante (Martin et Etchevers, 2005).

L'objectif de notre travail est d'intégrer des facteurs topographiques à fine résolution qui régissent très fortement la variabilité spatiale des températures en montagne (altitude, orientations, pentes et encaissement notamment). Comme cela a été exposé dans les paragraphes ci-dessus, les températures ne sont pas le seul facteur explicatif de la présence ou de l'absence de neige, les précipitations mais également leur fréquence et le calendrier des chutes de neige jouent un rôle central, entre autres, dans l'épaisseur du manteau neigeux naturel. Cependant les températures sont modélisées avec bien moins d'incertitude que les précipitations par les modèles climatiques globaux et régionaux ; le choix de ne retenir que cette variable climatique ne nous permet donc pas de simuler véritablement le manteau neigeux mais, par contre, de calculer précisément l'altitude de l'isotherme 0°C pour les températures moyennes mensuelles (établies par la moyenne des températures minimales et maximales mensuelles), qui par extrapolation peut servir de point de repère dans l'espace afin d'apprécier globalement la limite basse potentielle du manteau neigeux.

Pour se projeter au cours du XXI^e siècle et tenter de percevoir l'évolution future des températures dans les Alpes du sud françaises, notre démarche est alimentée par un modèle climatique régional du CNRM (Centre National de Recherches Météorologiques) appelé ALADIN-Climat (Aire Limitée Adaptation dynamique Développement InterNational) et dont les simulations ont été spécialement établies pour les zones montagneuses en France à 12 km de résolution à l'occasion de l'ANR SCAMPEI (Déqué, 2012). Trois scénarios d'évolution socio-économiques et démographiques ont été retenus (A1B, B1 et A2) ainsi que deux échéances temporelles, 2021-2050 et 2071-2100. Une étape appelée downscaling (descente d'échelle) statistique a été introduite pour que chacun des pixels de 12 km de côté (144 km²) soit subdivisé en 230400 nouveaux pixels de 25 m de côté (625 m²). Cette amélioration de la finesse spatiale des températures est essentielle pour que les cartes finales puissent servir d'outil d'aide à la décision aux gestionnaires et décideurs locaux.

3. Méthode, données et mise en œuvre de la descente d'échelle statistique

Les méthodes de descente d'échelle statistique sous-tendent que le climat régional ou local dépend de deux facteurs : le climat à échelle continentale et les caractéristiques locales ou régionales comme la topographie, la répartition terre/mer ou l'occupation du sol (Von Storch, 1999). Cependant dans la quasi-totalité des travaux réalisés à l'aide d'une méthode de downscaling statistique, les auteurs considèrent que les caractéristiques locales ou régionales se limitent uniquement aux variables météorologiques et ils ne prennent pas en compte d'autres variables environnementales telles que celles citées ci-dessus (Mullan, 2012 ; Khalili, 2013). Le point commun à la grande majorité de ces études réside dans le fait qu'en fin de

Évolution de l'isotherme 0°C dans les Alpes du Sud durant le XXI^e siècle

traitement les prévisions climatiques ne sont établies que pour des points précis de l'espace. En règle générale ces points sont les stations météorologiques qui ont fourni les variables locales pour établir les relations statistiques avec les sorties de modèles climatiques.

Dans la bibliographie un grand nombre d'articles traite de cette question de descente d'échelle statistique. Il s'agit fréquemment d'améliorer la précision spatiale des projections climatiques concernant les précipitations journalières (Liu, 2013 ; Lutz, 2012 ; Wilby, 2000). En effet cette variable alimente de nombreux modèles environnementaux comme des modèles hydrologiques par exemple (Boé, 2007 ; Sachindra, 2013). Concernant les températures les travaux sont également abondants : régionalisation des températures dans l'état du Rajasthan en Inde (Goyal, 2011), dans le sud-ouest du Québec et le sud-est de l'Ontario (Khalili, 2013), dans le nord-est des Etats-Unis (Ahmed, 2013) et dans le nord du Chili (Souvignet, 2010) par exemple.

Pour reconstituer un champ thermique selon une démarche statistique à partir soit d'un réseau de stations météorologiques, soit d'un ensemble de pixels de modèles climatiques (et donc pour affiner la résolution de ces pixels), il est illusoire de s'appuyer sur des outils fondés sur l'autocorrélation spatiale. Celle-ci suppose en effet que la distance à un point dont la valeur est connue régit intégralement le calcul de la température. Or divers travaux ont montré que cette approche trouve ses limites en relief contrasté (Laborde, 1984 ; Lhotellier, 2005). Ainsi la méthode choisie s'appuie sur la régression multiple que l'on peut qualifier de régression « environnementale » puisque les variables explicatives retenues font référence à l'environnement propre des individus (stations ou pixels) constituant l'échantillon de départ du modèle statistique. Bien sélectionnées, les variables explicatives intègrent implicitement dans l'équation de régression les mécanismes physiques qui influencent les températures. En relief contrasté la prise en compte de l'orientation et de la pente des versants permet par exemple d'intégrer l'énergie solaire captée par le sol, l'altitude quant à elle incorpore l'intensité de l'absorption d'énergie solaire par la masse d'air qui voit sa pression baisser avec l'éloignement par rapport au niveau de la mer, etc. De nombreux chercheurs ont eu recours à la régression multiple pour spatialiser les températures minimales (Tn) ou maximales (Tx) avec de très bons résultats, notamment sur des normales climatiques trentenaires (Carrega, 1982, 1998, 1995 ; Madelin, 2004 ; Lhotellier, 2005 ; Joly, 1987 ; Joly et al., 2011 ; Hengl, 2009).

Les données alimentant les modèles de régression font parties de deux grandes catégories, des sorties de modélisation climatique et des données d'observation topographique ou climatique :

- D'une part les calculs sont alimentés par les sorties de modèles d'ALADIN-Climat concernant les Tn et les Tx des mois de décembre à mars et pour les périodes 1961-1990 (référence), 2021-2050 et 2071-2100 selon trois les scénarios les plus utilisés, le A1B, le B1 et le A2. La version exploitée a été celle développée dans le cadre du projet ANR-SCAMPEI (Déqué, 2012). Les sorties de ce modèle ont été corrigées selon la technique quantile-quantile (Déqué, 2007). Cette procédure

SAGEO'2016 – Nice, 6-9 décembre 2016

permet d'ajuster au mieux les valeurs des quantiles calculés par le modèle ALADIN-Climat à ceux calculés à partir des observations de Météo-France. A chacun des scénarios sont associées des émissions de gaz à effet de serre (GES) d'intensité différente. Le scénario A2 envisage une multiplication par 4 des émissions d'ici à 2100, le scénario A1B prévoit un doublement de celles-ci pour 2050 puis une stabilisation, et le scénario B1 suppose des émissions quasi constantes durant tout le XXI^e siècle (IPCC, 2007). La période dite « de référence » est la « normale » 1961-1990 qui a permis de valider la capacité d'ALADIN-Climat à reproduire le climat considéré comme actuel (ou très légèrement passé étant donné la vitesse de l'évolution), ainsi que l'aptitude du modèle de régression environnementale à affiner le champ thermique. Cette validation se double d'une confrontation avec les données observées de 13 postes météorologiques situés dans le département des Alpes-Maritimes.

- D'autre part les modèles de régression sont nourris par des prédictors liés à la topographie et issues du M.N.T. (modèle numérique de terrain BD Alti) à 25 m de l'I.G.N. (Institut National de l'Information Géographique et Forestière) ; outre la longitude, la latitude et l'altitude ce sont près d'une cinquantaine de variables qui ont été testées pour être incorporées ou non dans les modèles de régression : exposition (découpée en deux catégories : nord/sud et est/ouest variant à chaque fois de -1 à 1), pente, distance à la mer (y compris son logarithme), une trentaine de formulations de l'encaissement (profondeur d'un talweg sur divers rayons) et une douzaine de types de rugosités (écart-type de l'altitude sur divers rayons). Enfin l'étape de validation des résultats pour la période de référence 1961-1990 a eu recours aux températures minimales et maximales mensuelles de décembre à mars observées dans le réseau de stations de Météo-France présent dans la partie montagneuse des Alpes-Maritimes (13 postes).

Pour établir nos modèles de régression, un modèle par mois (décembre à avril) et par variable (Tn et Tx), soient 10 modèles pour la période de référence 1961-1990, les Tn et Tx de chaque pixel d'ALADIN-Climat sont associées aux valeurs de chaque descripteurs topographiques exposés précédemment. L'équation de régression ainsi calculée à 12 km est ensuite appliquée à 25 m pour effectuer la descente d'échelle et donc pour affiner l'information spatiale relative aux températures. Ce « saut spatial » entre l'établissement de l'équation et l'utilisation de cette même équation pour produire la carte pose inévitablement un problème de principe : les gradients (coefficients de régression) ne sont pas nécessairement les mêmes selon l'échelle. Pour minimiser le plus possible ce problème, les descripteurs des pixels ALADIN-Climat n'ont pas été calculés à l'échelle 12 km, mais pour chacun de ces pixels ils sont la moyenne, la valeur minimale et la valeur maximale de tous les pixels de 25 m qui le composent et pour lesquels chaque prédictor a été évalué ; cela signifie qu'un prédictor donné se décline en 3 versions (moy., min. et max.). Ainsi, la valeur moyenne de l'encaissement maximum à 5 km d'un pixel ALADIN est la moyenne des encaissements maximums à 5 km de chacun des 230400 pixels compris dans les 144 km² ; de même la valeur minimale de l'encaissement maximum à 5 km d'un pixel ALADIN est le minimum de ce prédictor rencontré pour les 230400 pixels de 25 m inclus dans le pixel ALADIN et

Évolution de l'isotherme 0°C dans les Alpes du Sud durant le XXI^e siècle

ainsi de suite pour tous les descripteurs. Cette démarche permet de fonder le downscaling sur une relation forte entre les deux échelles, puisque l'une alimente l'autre avant même l'établissement de l'équation de régression.

Une fois la carte des températures à 25 m de résolution produite, la validation de cette modélisation est établie à l'aide des observations climatiques réalisées par Météo-France au cours de la période 1961-1990. C'est alors que l'écart moyen entre les normales climatiques observées dans chaque station, et les pixels de 25 m de côté correspondant à ces postes de mesures, permet de juger de la pertinence et de la robustesse de l'ensemble de la démarche de downscaling statistique. La validation de l'enchaînement logique des traitements a déjà été très largement exposée dans une précédente publication (Martin et al., 2013) et n'est volontairement pas aussi détaillée ici. Cependant les données topographiques ne provenaient pas de la même source (USGS à l'époque) et l'objectif n'était pas ciblé sur les températures en montagne ce qui signifie que l'ensemble des traitements ont été à nouveau réalisés pour le travail présenté dans cet article selon une logique rigoureusement identique.

Enfin la dernière étape s'appuie sur la validation de la descente d'échelle pour la période de référence. Dans la mesure où les résultats sont jugés acceptables alors de nouveaux modèles de régression multiple sont établis pour les différents scénarios et pour les différentes périodes du XXI^e siècle en les alimentant cette fois-ci avec les simulations de températures des sorties climatiques futures du modèle ALADIN-Climat. L'établissement de nouveaux modèles de régression pour les deux périodes futures permet d'éviter de poser l'hypothèse, potentiellement fautive, de stationnarité des processus physiques implicitement contenus dans les équations de régression.

4. Validation et résultats

Une des complications majeures réside dans le fait qu'une partie des données d'entrée alimentant les modèles de régression sont déjà issues de modélisations. Ces chiffres sont donc potentiellement affectés de biais introduits lors des calculs dans ALADIN-Climat. Bien qu'ayant subi une correction ces modélisations sont marquées par un biais froid à haute altitude pour les T_n (Martin et al., 2013). Puisqu'en statistique les résultats sont étroitement dépendants de l'échantillon initial, il a été décidé de ne pas prendre en compte les pixels dont les T_n pour janvier étaient inférieures à -14°C (choix notamment justifié par confrontation avec les données des stations de Météo-France situées aux plus hautes altitudes dans les Alpes-Maritimes). De plus des pixels ne recouvrant pas directement la zone d'intérêt mais situés à plus basse altitude ont été sélectionnés pour atténuer le biais froid de l'étape de modélisation antérieure (70 pixels sélectionnés pour les T_n). En dépit de cette précaution, les résultats obtenus montrent systématiquement une sous-estimation des T_n de plusieurs degrés pour les postes montagneux. Une correction d'une intensité correspondant à l'erreur moyenne sera donc apportée (cf. tab.2). Par contre pour les T_x tous les pixels (36) surplombant le département ont été retenus pour élaborer les équations de régression.

SAGEO'2016 – Nice, 6-9 décembre 2016

Pour la période de référence 1961-1990 les résultats obtenus sont le fruit d'une multitude d'essais pour aboutir à la meilleure solution, c'est-à-dire celle pour laquelle l'ensemble des critères d'appréciation statistiques sont les plus élevés tout en maintenant un nombre de prédicteurs faible et pour lesquels les mécanismes physiques sous-jacents sont cohérents. Après une longue phase de tâtonnement les Tn sont modélisées statistiquement par 5 prédicteurs (altitude, pente calculée sur 100 m, logarithme de la distance à la mer, encaissement maximum à 100 m et encaissement maximum à 20 km) ; les modèles de régression des Tx contiennent 6 variables explicatives (altitude, pente calculée sur 100 m, logarithme de la distance à la mer, rugosité à 5 km, exposition nord/sud et exposition est/ouest).

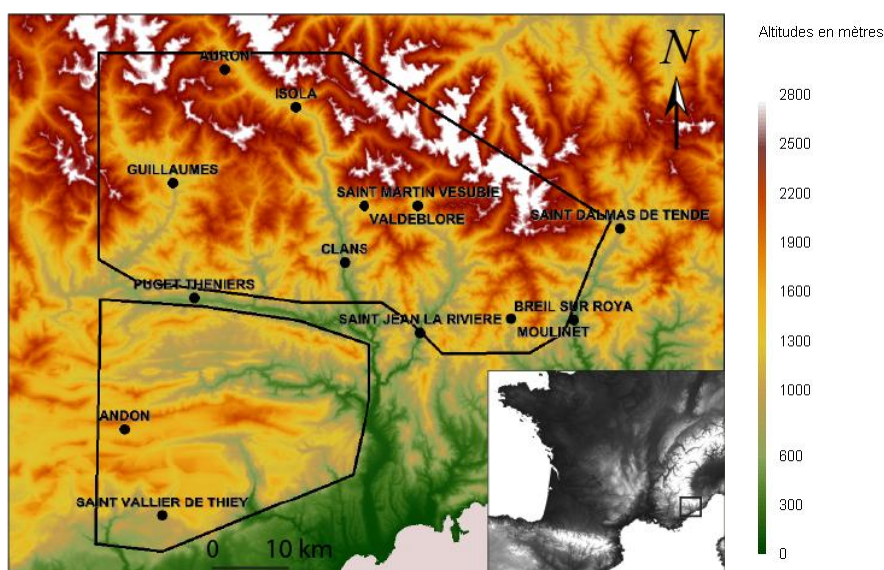


FIGURE 1. Postes de Météo-France utilisés pour valider les modèles de régression sur la période 1961-1990 et délimitations en noir des zones géographiques Massif du Mercantour (nord) et Préalpes de Grasse (sud-ouest)

Les coefficients de détermination des modèles de régression des Tn sont compris entre 0.88 et 0.91, les erreurs-type varient entre 0.69 et 1.03°C et les écarts moyens (RMCE pour racine de la moyenne du carré des écarts) s'échelonnent de 0.71 à 1.07°C. Concernant les Tx les paramètres statistiques sont encore meilleurs avec notamment des R^2 avoisinant les 0.98 (tab. 1). L'ensemble de ces chiffres attestent de résultats très satisfaisants pour les 5 mois étudiés de décembre à avril. Les valeurs de Tn ou de Tx des pixels d'ALADIN-Climat sont correctement restituées par nos modèles de régression. Cependant pour valider le "saut spatial" évoqué précédemment il est encore plus pertinent, après avoir construit les cartes des Tn et des Tx mensuelles à l'aide des équations de régression à 25 m de résolution spatiale, de mesurer les écarts entre les températures modélisées et les normales climatiques

Évolution de l'isotherme 0°C dans les Alpes du Sud durant le XXI^e siècle

enregistrées par Météo-France (13 postes, fig. 1) pour les pixels comprenant cette observation. Comme évoqué dans le paragraphe précédent, pour les Tn les cartes ont été corrigées de l'écart moyen constaté sur l'ensemble des 13 points d'observation. Après cet ajustement les écarts moyens observés pour les Tn sont compris entre 0.6 et 1.2°C, et pour les Tx entre 0.8 et 1.2°C selon les mois (tab. 2).

TABLE 1. Coefficients de détermination ajustés, erreurs-type et RMCE des modèles statistiques des Tn et Tx mensuelles de décembre à mars pour 1961-1990

	Décembre	Janvier	Février	Mars	Avril
Tx R ²	0.97	0.98	0.98	0.99	0.98
Tx erreur-type	0.44	0.41	0.42	0.31	0.42
Tx RMCE	0.48	0.45	0.46	0.35	0.47
Tn R ²	0.90	0.89	0.90	0.91	0.88
Tn erreur-type	0.69	0.70	0.72	0.79	1.03
Tn RMCE	0.71	0.73	0.75	0.82	1.07

Comme les résultats obtenus sur la période 1961-1990 sont jugés satisfaisants et qu'ils ont été validés, alors l'opération est renouvelée à partir des Tn et Tx futures simulées par ALADIN-Climat pour les 3 scénarios et pour les deux périodes trentenaires en utilisant les mêmes prédictors. Seules les valeurs de correction apportées aux cartes des Tn sont conservées et à nouveau introduites dans la génération des cartes des Tn futures puisqu'il est impossible d'établir de nouveaux ajustements pour 2021-2050 et 2071-2100 en l'absence de mesures déjà réalisées... Excepté cette opération de correction, l'ensemble des relations statistiques entre températures et variables topographiques/environnementales sont à nouveau calculées pour construire de nouveaux modèles de régression. Les paramètres statistiques obtenus sont du même ordre que pour la période de référence.

TABLE 2. Valeurs d'ajustement des modèles statistiques des Tn mensuelles de décembre à mars pour 1961-1990 et écarts moyens (RMCE) constatés entre les Tn et Tx modélisées et les mesures dans 13 postes de Météo-France

	Ajustement Tn	RMCE Tn	RMCE Tx
Décembre	+ 3.23°C	1.2	1.2
Janvier	+ 2.84°C	1.2	1
Février	+ 2.45°C	0.8	0.8
Mars	+ 1.92°C	0.6	1
Avril	+ 0.52°C	1.1	1.6

SAGEO'2016 – Nice, 6-9 décembre 2016

Pour mieux apprécier l'évolution des températures, les hausses des Tm (températures moyennes établies à partir de la moyenne des Tn et des Tx) ont été calculées pour deux zones géographiquement distinctes dans les Alpes-Maritimes, correspondant globalement aux Préalpes de Grasse, où les Tm hivernales sont moins basses car plus influencées par la proximité de la mer Méditerranée, et au Massif du Mercantour (cf. fig. 1). Globalement Les Tm augmenteront de 0.8°C à 4.8°C avec des différences sensibles en fonction des périodes, des scénarios et des mois (tab. 3). Géographiquement il semblerait que la hausse soit plus forte dans les Préalpes de Grasse que dans le Massif du Mercantour avec +1.4°C et +2.3°C contre +1.7°C +2.6°C tous mois et scénarios confondus respectivement pour 2021-2050 et 2071-2100 ; temporellement ce sont les mois de mars et d'avril qui connaîtraient les plus fortes progressions.

TABLE 3. Hausses moyennes des Tm par mois, par scénarios, par période et par zone géographique par rapport à 1961-1990

Hausses Tm en °C par rapport à 1961- 1990		Massif du Mercantour					Préalpes de Grasse				
		Dec.	Jan.	Fev.	Mars	Avr.	Dec.	Jan.	Fev.	Mars	Avr.
2021-2050	B1	1	1	1.1	1.7	1.6	1.2	1	1.2	2.7	1.5
	A1B	1.2	1.3	1	2.9	2.4	1.4	1.2	1.1	4.1	2.2
	A2	0.8	1.3	0.9	1.8	1.4	0.9	1.2	1.1	2.9	1.3
2071-2100	B1	1.3	1.4	1.4	2.5	2.3	1.4	1.4	1.6	3.7	2.1
	A1B	1	2.1	2.2	3.3	2.9	1.5	2.1	2.4	4.4	2.7
	A2	2.7	2.1	2.4	3.8	3.7	2.8	2.2	2.6	4.8	3.4

A partir des cartes des Tm mensuelles produites pour les 3 périodes de 30 ans et pour les 3 scénarios futurs l'altitude moyenne de l'isotherme 0°C a été extraite. Pour l'ensemble des mois une nette remontée en altitude de l'isotherme 0°C est simulée quel que soit le scénario envisagé par rapport à 1961-1990 (tab. 4). Pour 2021-2050 et en moyenne sur les 5 mois l'isotherme 0°C s'élèverait de 246 à 339 m sur les ubacs (zones davantage privilégiées par les domaines skiables) du Massif du Mercantour. Ces chiffres s'amplifient pour 2071-2100 pour rejoindre un intervalle de 352 à 599 m. Des disparités importantes sont simulées en fonction des scénarios et des mois, les évolutions les plus marquantes sont provoquées par les scénarios A1B et A2 et pour la fin du siècle. De plus, l'évolution de l'altitude de l'isotherme permet d'imaginer quelle station de ski, en fonction de la répartition altitudinale de son domaine skiable, pourrait rencontrer des difficultés (fig. 2 et 3). Naturellement la température moyenne n'est que l'un des facteurs explicatifs de la présence de neige au sol et aucun raccourci n'est fait ici pour dire que l'altitude de l'isotherme 0°C marquera la limite basse du manteau neigeux naturel (d'autant plus que sur les pistes la neige est tassée et de ce fait perdure bien plus longtemps) ; cette variable représente seulement une indication forte, et une sorte de seuil psychologique, pour

Évolution de l'isotherme 0°C dans les Alpes du Sud durant le XXI^e siècle

visualiser les changements climatiques attendus dans la zone montagneuse du département. Dans le Massif du Mercantour les principales stations de ski, Auron, Isola 2000 et Valberg voient respectivement leur domaine skiable s'étendre entre 1250 et 2450 m, 1800 et 2610 m, 1700 et 2066 m. Ainsi, et progressivement durant le XXI^e siècle, une proportion grandissante des pistes sera située sous l'isotherme 0°C, et même en plein cœur de l'hiver. La zone des Préalpes de Grasse, un peu moins froide car située à plus basse altitude (point culminant à 1778 m avec la Cime du Cheiron) ne connaîtrait plus de Tm mensuelles moyennes négatives selon nos résultats dès 2021-2050 quel que soit le scénario excepté pour le mois de janvier.

TABLE 4. *Elévation en mètres de l'altitude de l'isotherme 0°C pour les scénarios B1, A1B et A2 de décembre à mars en 2021-2050 / 2071-2100 par rapport à 1961-1990*

	2021-2050				2071-2100			
	B1	A1B	A2	Moy.	B1	A1B	A2	Moy.
Dec.	264	338	213	272	367	272	731	457
Jan.	244	303	315	287	348	532	518	466
Fev.	240	220	199	220	304	469	514	429
Mars	238	446	250	311	368	507	589	488
Avr.	269	387	253	303	373	498	641	504
Moy.	251	339	246	279	352	456	599	469

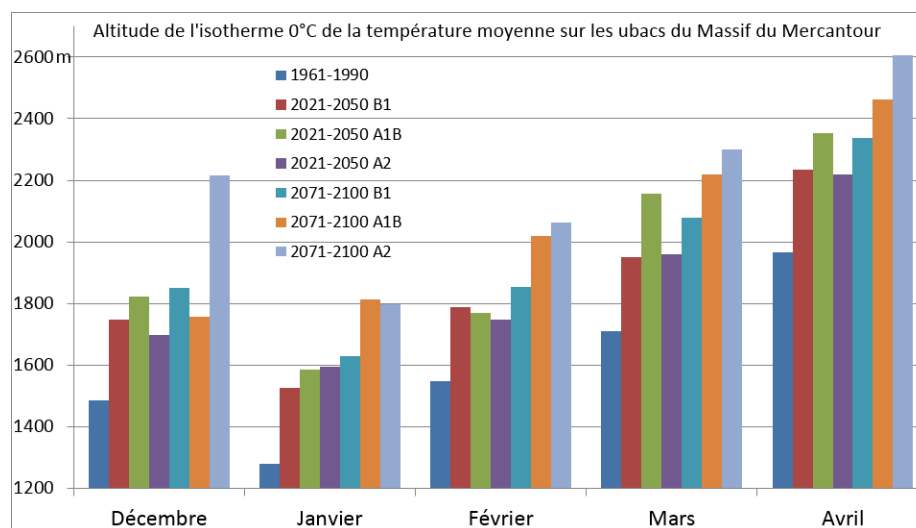


FIGURE 2. *Altitudes moyennes de l'isotherme 0°C des Tm mensuelles de décembre à mars sur les ubacs du Massif du Mercantour en fonction des périodes et des scénarios*

SAGEO'2016 – Nice, 6-9 décembre 2016

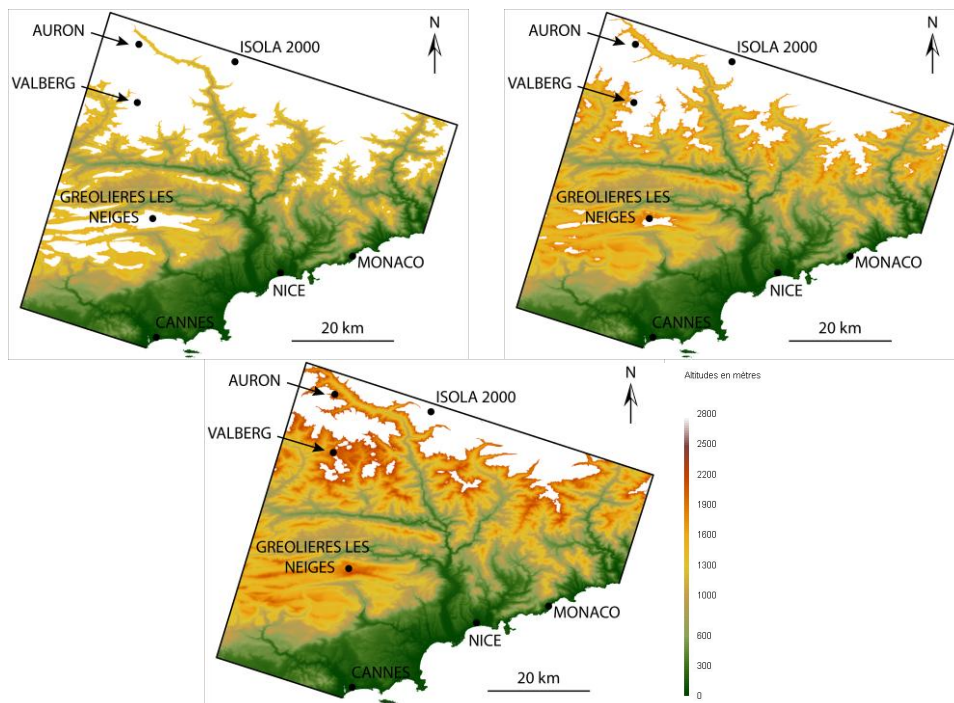


FIGURE 3. Évolution de la surface concernée par des T_m négatives (en blanc) au mois de février 1961-1990 (en haut à gauche), 2021-2050 (en haut à droite) et 2071-2100 (en bas) pour le scénario A1B

4. Conclusion

L'évolution du climat dans les Alpes du Sud françaises aura inéluctablement un impact sur l'activité économique de l'arrière-pays montagneux qui s'appuie aujourd'hui largement sur l'or blanc. L'étape de downscaling statistique appliquée aux températures modélisées par ALADIN-Climat a permis d'affiner considérablement la résolution spatiale des températures afin d'établir des cartes de l'isotherme 0°C des T_m qui ont une signification plus concrète que les pixels de 12 km de côté en sortie de modèle climatique. La superficie montagneuse concernée par des T_m négatives de décembre à mars connaîtra une décroissance marquée tout au long du XXI^e siècle. Seules les stations et domaines skiables situés aux plus hautes altitudes des Alpes-Maritimes pourront encore bénéficier de températures favorables aux activités de sports d'hiver pendant plusieurs mois.

Remerciements : Nous remercions Michel Déqué du CNRM pour la mise à disposition des modélisations d'ALADIN-Climat.

Évolution de l'isotherme 0°C dans les Alpes du Sud durant le XXI^e siècle

Bibliographie non numérotée et références

- Ahmed K.F., Wang G., Slinader J., Wilson A.M., Allen J.M., Horton R. et Anyah R., (2013). Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *Global and Planetary Change*, 100, p. 320-332.
- Bocchiola D. et Diolaiuti G. (2009). Evidence of climate change within the Adamello Glacier of Italy. *Theoretical and Applied Climatology*, 100, p. 351-369.
- Boé J., Terray L., Habets F. et Martin E., (2007). Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology*, 27, p. 1643-1655.
- Carrega P., (1982). *Les facteurs climatiques limitants dans le sud des Alpes Occidentales*. Thèse doctorat 3^e cycle. Revue d'Analyse Spatiale Quantitative et Appliquée. Nice, n°13, 221 p.
- Carrega P., (1994). *Topoclimatologie et habitat*. Analyse Spatiale Quantitative et Appliquée, Revue de Géographie du laboratoire d'analyse spatiale Raoul Blanchard, Thèse de doctorat, 408 pages.
- Carrega P., (1995). A method for the reconstruction of mountain air temperatures with automatic cartographic applications. *Theoretical and Applied Climatology*, 52, p. 69-84.
- Déqué M., (2012). *Projet ANR-SCAMPEI Compte-rendu de fin de projet*, 20 pages.
- Déqué M., (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, Vol 57, p. 16-26.
- Goyal M. K. et Ojha C.S.P., (2011). PLS regression-based pan evaporation and minimum-maximum temperature projections for an arid lake basin in India. *Theoretical and Applied Climatology*, 105, p. 403-415.
- Hantel M. et Hirtl-Wielke L.-M. (2007). Sensitivity of Alpine snow cover to European temperature. *International Journal of Climatology*, 27 (10), p. 1265-1275.
- Hengl T., (2009). *A practical guide to geostatistical mapping*, 293 pages.
- IPCC, (2007). *Changements climatiques 2007 : Rapport de Synthèse*. 114 pages.
- Joly D., (1987). L'interpolation supervisée. Une méthode de traitement destinée à la cartographie automatique présentée à l'aide d'un exemple de climatologie. *Actes du Symposium intern. sur la topo-climatologie et ses applications*. Liège (Belgique), mars 1985. pp. 135-148.
- Joly D., Brossard T., Cardot H., Cavaihes J., Hilal M. et Wavresky P., (2011). Temperature interpolation based on local information: the example of France. *International Journal of Climatology*, 31, p. 2141-2153.
- Khalili M., Nguyen V.T.V. et Gachon P., (2013). A statistical approach to multi-site multivariate downscaling of daily extreme temperature series. *International Journal of Climatology*, 33, p. 15-32.
- Laborde J-P., (1984). *Analyse des données et cartographie automatique en hydrologie. Eléments d'hydrologie lorraine*. Thèse de doctorat, Nancy, 484 pages.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Lhotellier R., (2005). *Spatialisation des températures en zone de montagne alpine*. Thèse de doctorat, Grenoble, 352 pages.
- Liu W., Fu G., Liu C. et Charles S.P., (2013). A comparison of three multi-site statistical downscaling models for daily rainfall in the North China Plain. *Theoretical and Applied Climatology*, 111, p. 585-600.
- Lutz K., Jacobeit J., Philipp A., Seubert S., Kunstmann H. et Laux P., (2012). Comparison and evaluation of statistical downscaling techniques for station-based precipitation in the Middle East. *International Journal of Climatology*, 32, p. 1579-1595.
- Madelin M., (2004). *L'aléa gélif printanier dans le vignoble marnais en Champagne : Modélisation spatiale aux échelles fines des températures minimales et des écoulements de l'air*. Thèse de doctorat, Paris, 353 pages
- Marty C., (2013). Climate change and snow cover in the european Alps. *The impacts of skiing on mountain environments*, p. 30-44
- Mullan D., Fealy R. et Favis-Mortlock D., (2012). Developing site-specific future temperature scenarios for Northern Ireland: addressing key issues employing a statistical downscaling approach. *International Journal of Climatology*, 32, p. 2007-2019.
- Sachindra D.A., Huang F., Barton A. et Perera B.J.C., (2013). Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *International Journal of Climatology*, 33, p. 1087-1106.
- Scherrer S.C. et Appenzeller C. (2006). Swiss Alpine snowpack variability: major patterns and links to local climate and large-scale flow. *Climate research*, 32 (3), p. 187-199.
- Serquet G., Marty, C., Dulex J-P. et Rebetez M., (2011). Seasonal trends and temperature dependence of the snowfall/precipitation-day ratio in Switzerland. *Geophysical Research Letters*, vol. 38, L07703, doi: 10.1029/2011GL046976
- Souvignet M., Gaese H., Ribbe L., Kretschmer N. et Oyarzun R., (2010). Statistical downscaling of precipitation and temperature in north-central Chile: an assessment of possible climate change impacts in an arid Andean watershed. *Hydrological Sciences Journal*, 55:1, p. 41-57.
- Valt M. et Cianfarra P. (2010). Recent snow cover variation and avalanche activity in the Southern Alps. *Cold Regions Science and Technology*, 64 (2), p. 146-157.
- Von Storch H. (1999). On the use of "inflation" in statistical downscaling. *Journal of Climate*, 12, p. 3505-3506.
- Wetterhall F., Bardossy A., Chen D., Halldin S. et Xu C.-Y., (2006). Daily precipitation-downscaling techniques in three Chinese regions. *Water Resources Research*, 42, W114223.
- Wilby R.L. et Wigley T.M.L., (2000). Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology*, 20, p. 641-661.
- Wilby R.L., Charles S.P., Zorita E., Timbal B., Whetton P. et Mearns L.O., (2004). *Guidelines for use of the climate scenarios developed from statistical downscaling methods*. Published on-line, supporting material to the Intergovernmental Panel on Climate Change, 27.

[ISBN : 2-910545-11-3]



Posters

Analyses des cycles saisonniers de la végétation en lien avec les variations saisonnières des précipitations à partir des données de Télédétection basses résolutions spatiales NDVI (MODIS) et de précipitation (TRMM)

G.B. Boussougou¹, Y.T.Brou^{1,2}, P. Valimba³

^{1,2} UMR ESPACE-DEV (UAG, UM2, UR) / SEAS-OI, Université de La Réunion, 40, Avenue de Soweto (IUT), 97410 Saint Pierre
guyboussougou@yahoo.fr

³ Université de Dar Es Salaam, *College Engineering and Technology (COET)*

RESUME : L'eau est le facteur principal de la croissance de la végétation des régions tropicales. Dans cette étude à partir des données de télédétections basses résolutions, nous nous proposons de montrer les corrélations spatiales entre les pluies et les végétaux en Tanzanie entre 2001 et 2013. Pour cela, on utilise les données satellites d'indice de végétation du capteur MODIS (Moderate Resolution Imaging Spectroradiometer) et des données pluviométriques obtenues par TRMM (Tropical Rainfall Measuring Mission). Dans un premier, des séries temporelles des NDVI et des précipitations ont été créées. Par la suite les calculs statistiques (moyennes annuelles coefficients de corrélations) ont permis de spatialiser les variations pluies et de la distribution de l'activité végétale. Les forts coefficients de corrélations (supérieur à 0.50) permettent de voir que le centre du pays est la région où se localisent les végétaux les plus sensibles (savanes, prairie) qui ont une réponse précoce aux pluies, contrairement aux zones de forêts du sud qui ont eux, une réponse tardive (plus d'un mois de décalage) faces aux précipitations.

MOTS-CLES : NDVI, précipitation, corrélation, Tanzanie, Télédétection

ABSTRACT: Water is the main factor in the growth vegetation in tropical regions. In this study from remote sensing low resolution data, we intend to show the spatial correlations between seasonal cycles of rainfall and plant in Tanzania between 2001 and 2013. For this, we use satellite data of vegetation index MODIS (Moderate Resolution Imaging Spectroradiometer) and rainfall data from TRMM (Tropical rainfall Measuring Mission). In a first time series of NDVI and rainfall have been created. Subsequently statistical calculations (annual, correlation coefficients) allowed spatialize seasonal variations in rainfall. The strong correlation coefficients (greater than 0.50) to see that the center of the country is the region or locate the most susceptible plants (savannah grassland) have an early response to rainfall, unlike the southern forest areas have a late response (more than one month lag) precipitation variations.

KEYWORDS: NDVI, PRECIPITATION, CORRELATION, TANZANIA, REMOTE SENSING

1. Introduction

Les précipitations représentent l'élément prépondérant de la diversification des paysages forestiers tropicaux (Meneses-Tovar, 2011 ; Dupuy et al, 1999 ; Brou et al, 2004). En Afrique, beaucoup d'études sur le lien pluie/végétation restent régionale (Nicholson et al. 1990 ; Fuller et Prince, 1996 ; Richard et Pocard, 1998) et n'ont pas permis de répondre à toutes les questions sur les relations entre ces deux paramètres, à une échelle nationale. Existe-t-il des différenciations spatiales dans la réponse végétale face aux précipitations?

L'objectif de cette étude est donc d'apporter quelques réponses à cette interrogation à l'échelle du territoire Tanzanien (pays d'Afrique de l'Est en bordure de l'océan indien, dans la partie tropical sud).

2. Données et méthodologie

-Les données de l'étude sont :

1) Données d'estimation de pluies par satellite de la Rainfall Monitoring Mission (TRMM), à une résolution spatiale de $0,25^\circ \times 0,25^\circ$, agrégées au pas de temps mensuelle (TRMM 3B43) ; 2) Données de l'indice de végétation normalisé (Normalized Difference Vegetation Index : NDVI) (Tucker, 1979) issus des produits MODIS de résolution spatiale 250m ; 3) La carte d'occupation de sol de Tanzanie à une résolution spatiale de 300m (Globcover ,ESA).

-La méthodologie est résumée comme suit :

A) : Archivages des données et création des séries temporelles de NDVI (276 images) et TRMM (156 images) ; B) Ré-échantillonnage des données NDVI à la résolution $0,25^\circ$ et agrégés mensuellement ; C) calcul de la valeur moyenne de NDVI et pluie sur la période 2001-2013; D) calcul du coefficient de corrélation (Goupe, 2001) pixel de NDVI par pixel de pluie TRMM. La moyenne mensuelle de chaque mois de NDVI et pluie sur la période sert pour le calcul de corrélation. La corrélation est estimée entre (i) les valeurs moyenne de NDVI pour tous les mois et (ii) les valeurs correspondantes de précipitations. On décale ensuite, le NDVI de 1,2 et 3 mois après le mois de précipitations pour observer le temps de réponse de la végétation.

SAGEO'2016 – Nice, 6-9 décembre 2016

3. Résultats

La cartographie moyenne annuelle des pluies et NDVI (2001-2013) permet de spatialiser les pluies et l'activité végétale. La zone de précipitations élevées (humide) au Nord et au Sud-ouest ($> 1128\text{mm}$) correspond aux valeurs de NDVI > 0.60 ; zone de précipitations modérées (sèche) (entre 772 et 962 mm) au Sud-est et Ouest correspond au NDVI moyen (0.46-0.60); zone de « de précipitations faibles ($< 800\text{mm}$) au centre correspond au NDVI < 0.46 .

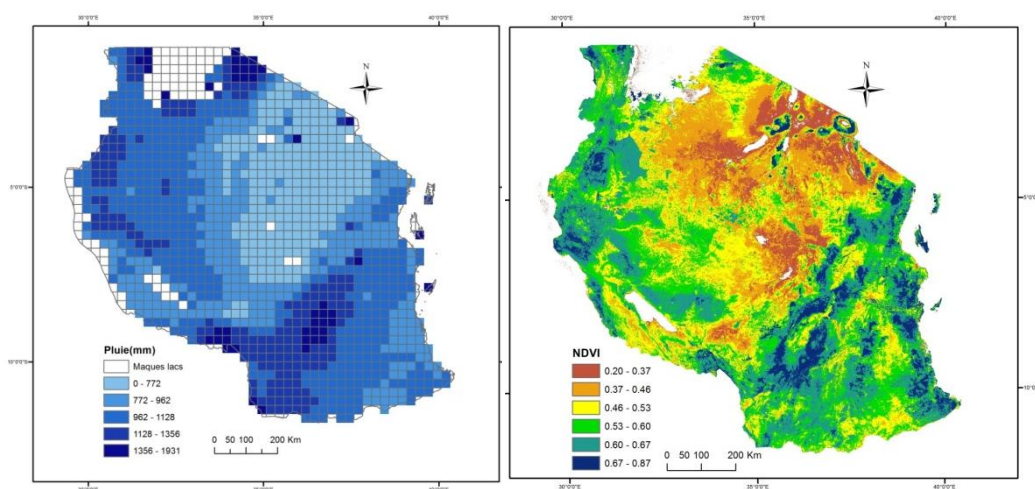


Figure 1: Moyenne annuelle du NDVI (résolution spatiale : 250 m) et pluie (résolution spatiale : 0.25°) sur la période 2001-2013

Le résultat du calcul des coefficients de corrélations est présenté dans la figure 2. Des variations importantes des coefficients sont observées (0.22-0.75) spatialement dans le pays. Les fortes corrélations donnent un seuil de significativité $> 5\%$.

On note, pour 0 mois de décalage, une réponse précoce de la végétation localisée au centre du pays: réaction végétale presque simultanée aux pluies (corrélations > 0.50). A partir de 1 mois de décalage, on observe une réponse tardive à l'Est, Ouest et Sud. Les fortes valeurs de corrélations se déplacent du centre vers le Sud du territoire en fonction du nombre de mois de décalage. On remarque que, les corrélations varient également en fonction du type de végétation (Figure 3, Annexe): La végétation de type 'savane,

prairie ‘ aura une réponse rapide (précoce) et la végétation type ‘forêt’ réagira tardivement aux précipitations. Les valeurs négatives de corrélation signifient un impact négatif des pluies sur cette région du Sud. En effet, dans cette région se trouve l’arc de chaine de montagne d’Afrique de l’Est qui se prolonge au Sud de Tanzanie. La variation végétale ici, peut être influencée par le relief au profil des précipitations.

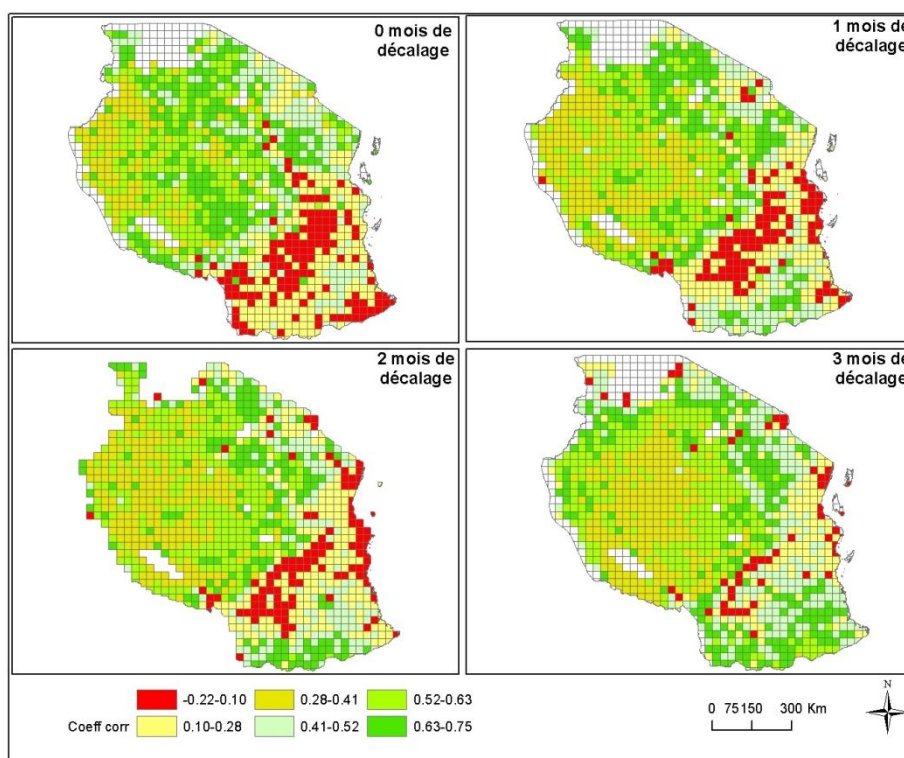


Figure 2: *Corrélation Pluie/NDVI et décalage temporel*

4. Conclusion

Cette étude a permis de montrer qu’il existe un lien étroit entre la distribution de la végétation et les précipitations à une échelle Nationale (Tanzanie). De même, on a vu qu’il existe un temps de réponse végétale aux pluies qui dépend du nombre de mois de décalage du NDVI, mais aussi du type de végétation. Les savanes et

SAGEO'2016 – Nice, 6-9 décembre 2016

prairies du centre ont une réponse précoce (même mois) tandis que les forêts de l'Ouest, Est et Sud ont une réponse tardive de 1,2 et 3 mois aux précipitations.

Cette étude peut être améliorée en utilisant l'indice EVI (enhanced vegetation index) à la place du NDVI pour s'affranchir des problèmes de saturation des forêts claires et denses des pays tropicaux.

Bibliographie

Brou Y. T., Akindès F., Bigot S. (2005). La variabilité climatique en Côte d'Ivoire : entre perceptions sociales et réponses agricoles. Cahiers d'études et de recherches francophones /Agriculture. Vol. 14, no 6, 530-40.

Dupuy B., Nasi R., Amsallem I., J. Loumeto J., Gérard C, Dupuy B. (1999). Gestion des écosystèmes forestiers denses d'Afrique tropicale humide: 2 : Congo.

Goupe Chadule (2001): Initiation aux pratiques statistiques en géographie. Collection U, série «Géographie», Edition Armand Colin, pp.103 et 109-110.

Meneses-Tovar, C. L. 2011. NDVI as indicator of degradation. Unasyuva, 8p.

Nicholson Se, Davenport MI, Malo Ar 1990. A Comparison of the Vegetation Response to Rainfall in the Sahel and East Africa, Using Normalized Difference Vegetation Index from NOAA AVHRR. Climatic Change 17, 209-241.

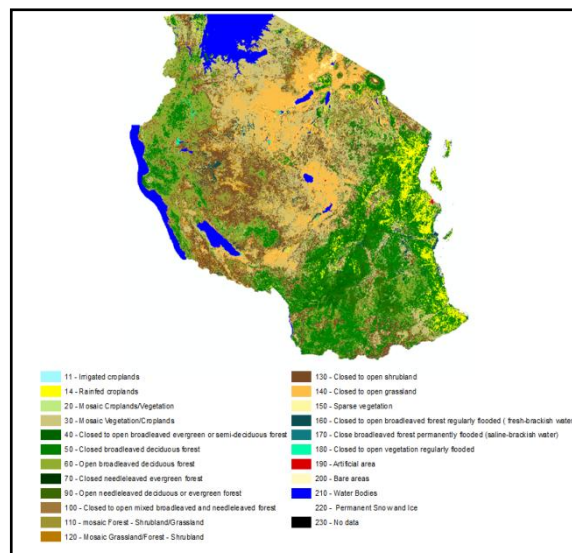
Richard Y, Pocard I. 1998. NDVI sensitivity to spatial and interannual rainfall variations in southern Africa. International Journal of Remote Sensing, 19: 29072920-2920.

Tucker, C. J. (1979). Red and photographic infrared linear combination for monitoring vegetation. Remote Sensing of Environment, 8,127–150.

http://www.glcn.org/databases/lc_gc-africa_en.jsp

Annexe:

Figure 3 : Carte d'occupation de sol Tanzanie (2005)



Apport de l'imagerie hyperspectrale pour la planification urbaine.

C Weber¹, S Gadal², X. Briottet³, C Mallet⁴

1. UMR 9000 TETIS, Montpellier 34 000 France christiane.weber@teledetection.fr;

2. Aix Marseille Université, CNRS ESPACE UMR 7300, Technopôle Arbois Méditerranée. Aix en Provence. sebastien.gadal@univ-amu.fr

3. ONERA DOTA Toulouse ; Xavier.Briottet@onera.fr

4. ING MATIS Paris ; Clement.Mallet@ign.fr

RESUME. L'ANR « Hyperspectral imagery for Environmental urban Planning – HYEP » est un projet dont l'objectif principal est de favoriser l'usage de l'imagerie hyperspectrale pour des applications de planification et de gestion urbaine. L'imagerie hyperspectrale diffère des moyens optiques usuels par la richesse des informations obtenues et les choix méthodologiques à réaliser. Le projet a identifié un certain nombre de verrous à lever permettant d'améliorer les traitements de classification ou de caractérisation des éléments urbains (toits, les infrastructures routières, les espaces de végétation etc.). Après avoir présenté les spécificités de ce type d'imagerie, le projet HYEP et ses avancées sont développés notamment les efforts réalisés pour la constitution d'un regroupement des bases de données spectrales en une base accessible dont les informations (spectres) puissent être complétées par des indicateurs morphologiques par exemple. La démarche générale proposée est résumé ainsi que l'articulation envisagée avec les acteurs de la Métropole de Toulouse (France) et la ville de Kaunas (Lituanie).

ABSTRACT. The HYEP project (Hyperspectral imagery for Environmental urban Planning) is dedicated to the use and the development of hyperspectral imagery for urban planning and management. Hyperspectral imagery differs from other optical products by the number of bands and so by the richness of information provided. HYEP has identified several locks that have to be locked up to promote hyperspectral imagery. This document presents the specificities of the sensor product and the methodological and technical advances we have realized. The spectral libraries package is thus presented and the general workframe of the project. The interaction with local stakeholders in Toulouse (France) and Kaunas (Lituanie) is also highlighted.

MOTS-CLES : imagerie hyperspectrale, planification, caractérisation spectrale, bases de données spectrales.

KEYWORDS: Hyperspectral imagery, planning, spectral recognition, spectral libraries

SAGEO'2016 – Nice, 6-9 décembre 2016

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

Les applications de l'imagerie hyperspectrale à la planification urbaine connaissent un développement important avec la miniaturisation des capteurs et l'utilisation croissante des drones et des plateformes aéroportées. Le grand nombre de bandes spectrales (entre 100 et 300), la finesse des intervalles spectraux sur l'ensemble d'un continuum spectral (visible proche infrarouge, infrarouge moyen, infrarouge thermique) permet une meilleure identification et caractérisation des objets urbains structurant et organisant les territoires urbains : bâtiments, réseaux, espaces verts ainsi que de leur caractéristiques : matériaux de toitures, types de revêtements, végétation etc., et de leur état: usure ou propreté par exemple.

L'objectif principal de l'ANR HYEP est le développement d'un continuum "méthodes/outils/services" se fondant sur les avancées obtenues en analyse d'imagerie hyperspectrale. Il s'agit de proposer un corpus de méthodes fiables à diverses résolutions spectrales et spatiales qui puisse être répliqué avec une définition et une stabilisation des meilleures approches de traitement.

Parmi les résultats majeurs attendus, on peut citer la production de chaînes de procédures adaptées à des applications cibles, et la démonstration de la valeur de l'imagerie hyperspectrale dans des approches urbaines.

Un des buts de l'ANR est de favoriser un transfert de technologie vers les concepteurs de capteurs, en effet HYEP a des liens fort avec le projet hyperspectral français HYPXIM.

Les territoires urbains sont constitués par un large panel d'objets géographiques qui caractérisent les usages socio-économiques, les structures urbaines, leurs évolutions et transformations. La complexité, l'hétérogénéité, et la multiplicité des objets géographiques résultent de choix de planification et de gestion urbaine. La caractérisation géométrique et spectrale des objets urbains détectés constitue la première étape nécessaire pour développer des outils de gestion urbaine dans ce contexte. Le catalogage et la génération d'une base de données spectrale participent à cette première phase. Une telle base de données spectrales est utilisée dans ces travaux pour caractériser précisément les objets urbains, en fonction de leur signature spectrale et leur morphologie à la fois dans les traitements de classification (comme base d'apprentissage) et comme élément de validation des résultats.

2. Imagerie hyperspectrale

L'imagerie hyperspectrale fournit des possibilités accrues de caractérisation des territoires à plus d'un titre et de manière plus spécifique car elle aborde les grandeurs physiques de la réalité observée à travers une gamme radiométrique plus large (0.4 μ m à 2.5 μ m) avec différentes largeurs de bandes (de l'ordre de 10 nm dans le domaine 0.4 à 2.5 μ m). Le grand nombre de celles-ci (entre 100 et 300) permet de discriminer plus aisément les caractéristiques d'absorption et d'émission des éléments. Depuis plusieurs années des capteurs hyperspectraux généralement aéroportés apportent des informations pertinentes sur des domaines variés (végétation, pollution des sols, cryosphère et urbain). Pour les territoires urbanisés la résolution spatiale de métrique à décimétrique des images obtenues permet d'obtenir des performances géométriques presque du même ordre que les satellites à très haute résolution spatiale (sentinel 2, Worldview etc.).

La compréhension des phénomènes radiatifs caractéristiques de la nature ou du comportement d'un objet d'intérêt est très importante afin de relier des grandeurs physiques recherchées par l'observateur (espèces végétales, des matériaux de

couverture, des zones imperméables, etc.) avec des grandeurs optiques de surface (émissivité et réflectance).

Des prétraitements doivent être appliqués aux images aéroportées hyperspectrales. En effet, en dehors de tous traitements liés au survol et à la prise de vue (correction géométrique), une correction atmosphérique est nécessaire pour s'affranchir des phénomènes de diffusion, d'absorption et/ou d'émission atmosphériques. Trois méthodes sont testées dans le cadre de HYPE pour diminuer l'impact de ces phénomènes (COCHISE, Méthode empirique et ICARE, tous trois développés par l'ONERA) en tenant compte des caractéristiques du milieu urbain (milieu hétérogène, élévation des bâtiments...) afin d'obtenir la meilleure estimation de la luminance diffuse ascendante via l'utilisation d'un couple de superpixels ombre/soleil.

Diverses approches peuvent ensuite être déployées pour traiter les images, certaines sont assez proches de ce qui se fait habituellement en observation de la terre, mais une difficulté spécifique réside dans la nécessité de trouver des modèles de mélange qui permettent d'approcher la composition des pixels et d'estimer conjointement les spectres des composants purs de l'image et leurs proportions respectives au sein de chaque pixel (Figure 1). Le problème devient ardu dans un milieu hétérogène comme le milieu urbain où la variété des éléments, la diversité des matériaux ainsi que la verticalité des objets complexifient cette étape.

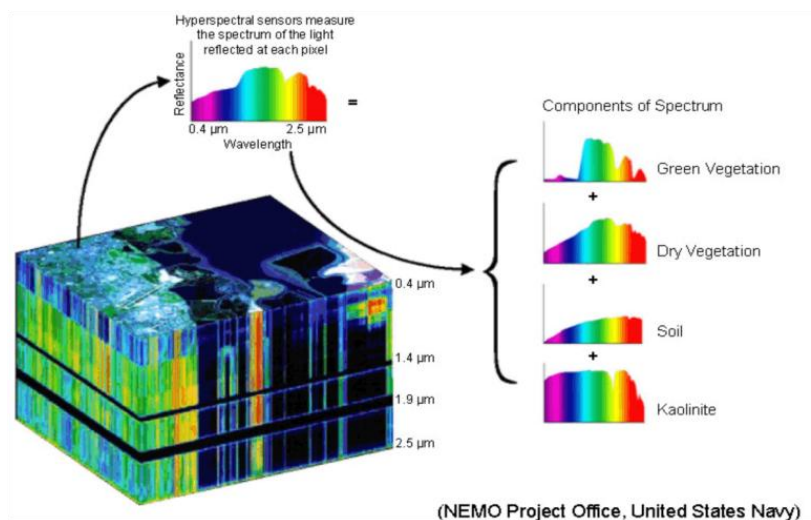


Figure 1 Composition d'un pixel hyperspectral (sourceUSNavy NEMO)

SAGEO'2016 – Nice, 6-9 décembre 2016

3. Les verrous

L'intérêt de l'imagerie hyperspectrale est indéniable en milieu urbain car elle facilite l'observation des éléments avec des grandeurs optiques plus riches à partir desquels non seulement l'occupation des sols est accessible mais aussi les spécificités liés à ces sols comme l'état de l'asphalte (propre ou non, usé ou non), des pavés ou des tuiles, ou encore la discrimination de la végétation par espèces et bien entendu par strates si des données 3D existent.

Cependant parmi les divers verrous généraux, nous pouvons noter :

- L'amélioration de la sensibilité radiométrique, de la résolution spatiale ou encore une plus grande gamme radiométrique allant jusqu'à l'infrarouge qui complexifie les traitements,
- Une certaine maturation des algorithmes les plus courants en hyperspectral est encore nécessaire pour que des non spécialistes puissent les utiliser (performances et ergonomie),
- La disponibilité de bases de données de spectres mesurés rassemblant des propriétés spectrales de manière efficace et qui soient représentatives ; celles qui existent sont dédiés à des milieux constitués le plus souvent pendant un projet de recherche et donc sont souvent fragmentaires ; elles sont dépendantes de la date, du lieu et des moyens d'acquisition et pas toujours bien documentées. Il s'agit de fournir des spectres validés permettant des analyses intra et intergroupes spectraux.
- L'accessibilité aux capteurs et la disponibilité des moyens d'exploitation.

3.1 Les vérités terrain ou les moyens de validation

Comme tout traitement en télédétection la validation des résultats est essentielle, cependant la possibilité de se référer à des « vérités terrain » n'est pas toujours aisée et ce, d'autant plus que les échelles spatiales associées à ces vérités terrains peuvent s'avérer contraignantes. La possibilité d'aller sur les toits afin de relever le matériau utilisé n'est pas commune, alors que le relevé par numérisation de polygone à partir d'une photographie aérienne reste l'approche la plus usitée. La caractérisation fine des différents matériaux de toitures obtenus par imagerie hyperspectrale couplée à des bases spectrales d'objets de références (mesurées en laboratoire ou sur le terrain) permet de valider sur des échelles locales voire régionales, les objets reconnus.

Cependant, le recours à des possibilités de validation à grande échelle notamment dans un cadre comparatif pose actuellement un certain nombre de questions : le faible nombre de polygones retenus et leur mise à disposition d'une part, ainsi que la représentativité statistique de ces vérités terrains d'autre part.

3.2 Les bases de données spectrales (BDS)

La constitution de bases de données de signatures spectrales BDS (en réflectance)

est aussi un moyen d'améliorer les classifications supervisées opérées sur les images en fournissant les moyens de discriminer au mieux les pixels proches d'une référence spectrale (figure 2). Plusieurs bases de données spectrales ont été constituées lors de diverses campagnes, que ce soit pour tester les capteurs (ASTER Library <http://speclib.jpl.nasa.gov/search-1> (Baldrige et al, 2009) ; <http://speclab.cr.usgs.gov/spectral-lib.html>, (Clark et al, 2007) ou pour s'appropriier une zone d'étude particulière comme la SLUM (Spectral Library of impervious Urban Materials) (Kotthaus et al, 2014). Il s'agit d'une librairie spectrale de matériaux urbains développée sur la ville de Londres (Grande-Bretagne), et contenant 74 spectres de réflectance visible - proche infrarouge (300- 2500nm) et infrarouge lointain (8 – 14 μ m) (Kotthaus et al., 2014). L'utilisation de ces BDS améliore le choix des bandes d'intérêt et permet d'identifier pour certains matériaux le nombre optimal de bandes nécessaires à la classification (Le Bris et al., 2016).

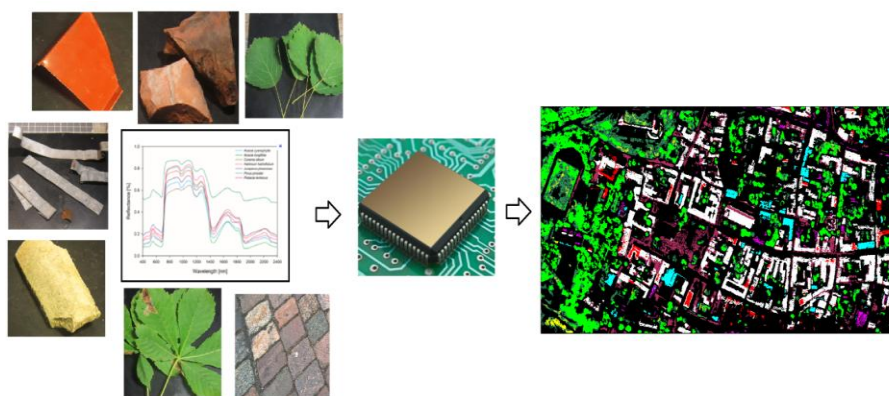


Figure 2: des matériaux à l'image hyperspectrale

Les matériaux sont mesurés à partir de spectromètres afin d'obtenir des courbes spectrales ; celles-ci sont ensuite utilisées pour la classification

4. La Démarche du projet « Hyperspectral imagery for Environmental urban Planning HYEP» (ANR Systèmes urbains durables (DS0601) <http://hyep.cnrs.fr/>)

Pour favoriser l'utilisation de l'imagerie hyperspectrale dans les études urbaines et fournir des moyens d'observation différents, plusieurs laboratoires et établissement de recherche se sont regroupés dans le cadre de l'ANR HYEP (CNRS, Onera, IGN, IRAP, GIPSA). Deux sites d'étude sont observés : la métropole de Toulouse (France) et la ville de Kaunas (Lituanie).

Sur ces deux villes plusieurs survols ont fourni des séries d'images (figure 3) l'une à partir d'un capteur HYPEX (Onera), l'autre à partir d'un capteur hyperspectral Vis-NIR RICOLA (RICOLA LTD).

SAGEO'2016 – Nice, 6-9 décembre 2016

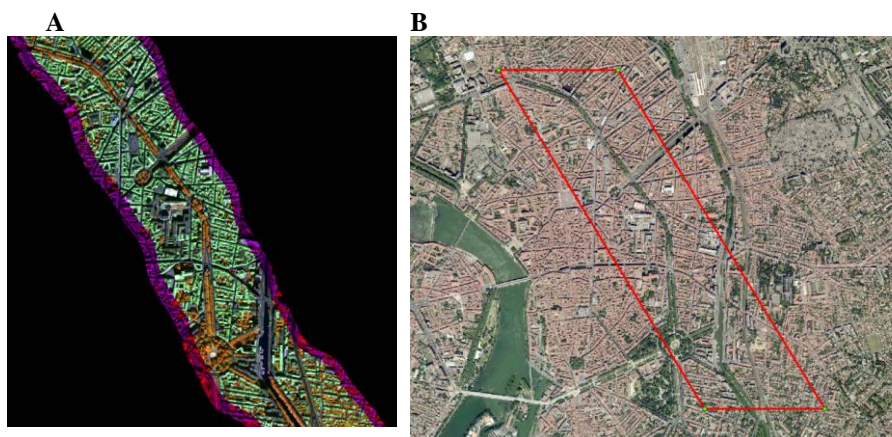


Figure 3. Image hyperspectrale campagne UMBRA ONERA/IGN : (R=830 nm; G=1650 nm; B=660 nm) (A) et en couleurs naturelles (B)

Sur chacune de ces villes les prises de vue sont accompagnées si possible de mesure de terrain selon un protocole défini.

4.1. Les prétraitements

Des corrections atmosphériques sur les images aéroportées ont été réalisées avec des comparaisons de résultats selon qu'on utilise ou non les ombres ou la 3D dans l'algorithme (Onera, UMR ESPACE). (Roussel et al. 2016 ; Gadal et al. 2016).

D'autres traitements de géoréférencement ont été nécessaires pour caler les images par rapport à des repères topographiques et corriger si possible les problèmes de prise de vue (IGN).

4.2 Les Bases de Données Spectrales

Un gros effort a été réalisé sur les bases de données spectrales après un état de l'art et un inventaire, afin d'harmoniser les bases existantes et fournir un protocole générique garantissant une cohérence des métadonnées avec la directive Inspire. Quelques 6000 spectres ont été rassemblés (plus de 1000 sur le milieu urbain) afin de structurer les signatures existantes et donner à la communauté un outil de travail interactif : d'une part par une interface de requête fondée sur une ontologie urbaine (CLC, 2006 2009) interrogeant la base de données rassemblant les signatures spectrales (Figure 4) ; d'autre part selon un protocole permettant de standardiser les informations collectées sur le terrain et d'enrichir la base facilement (relevés terrain: photo, mesures spectrales, croquis...). La localisation par GPS facilite aussi le repérage des sites sur d'autres sources (GoogleStreetmap, BD TOPO IGN etc.).

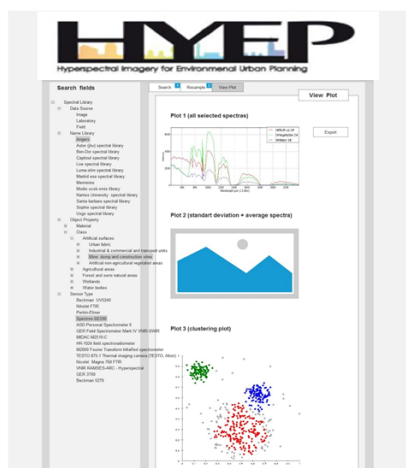


Figure 4 Prototype de la BDS
(source: rapport intermédiaire de l'ANR HYEP)

4.3 Les choix méthodologiques du projet

Les choix méthodologiques se posent à différentes étapes de la démarche générale, il est important de valider les choix opérés pour permettre l'établissement de spécifications utiles à la définition d'un capteur hyperspectral .

Pour les prétraitements on notera : la correction atmosphérique (quels algorithmes selon que l'on veut ou non tenir compte de l'ombre portée ou d'un modèle numérique d'élévation) et le géoréférencement pour corriger les décalages géométriques (quelles méthodes préservent au mieux la qualité du signal ?),

La constitution des bibliothèques spectrales permet d'améliorer les classifications et notamment de préciser le choix des bandes à retenir et leur largeur. Il s'agit d'harmoniser les bases de données collectées dans la communauté scientifique et de créer le prototype permettant à l'ensemble de la communauté une utilisation efficace et performante.

Actuellement, les méthodes de classification supervisées ou non, sont associés à des méthodes de démixage qui fournissent des cartes d'abondance de telles ou telles sources spectrales permettant de discriminer au mieux la composition des pixels. La fusion avec des données panchromatiques ou des données Lidar est aussi testée. Les résultats seront confrontés aux résultats obtenus avec les capteurs optiques en cours actuellement (Sentinel2, Pleiades) à différentes résolutions spatiales afin de statuer sur l'apport de l'imagerie hyperspectrale par rapport aux images optiques multispectrales.

Les résultats doivent servir à la spécification d'un futur capteur hyperspectral dans le cadre de la mission HYPXIM.

Enfin la représentation des résultats reste un enjeu incontournable, la traduction en grandeurs physiques des résultats doit être adaptée aux interrogations des décideurs locaux.

SAGEO'2016 – Nice, 6-9 décembre 2016

4.4 Les interactions avec les acteurs

Ces représentations font l'objet d'un module spécifique avec la Métropole de Toulouse et celle de Kaunas. En effet il nous a semblé qu'il était important de déterminer dès le début les éléments de restitution des résultats avec les services intéressés (environnement, planification). Pour cela nous avons échangé à la fois sur les zones d'intérêt de ces collectivités et sur les supports potentiellement intéressants et utilisables : comme la répartition par matériaux ou une carte d'imperméabilisation (Figure 5).. Pour la Métropole de Toulouse la zone de Montaudran, correspond à un enjeu local important, il s'agit d'une zone industrielle en reconversion, sur laquelle une partie de nos travaux se focalisent.

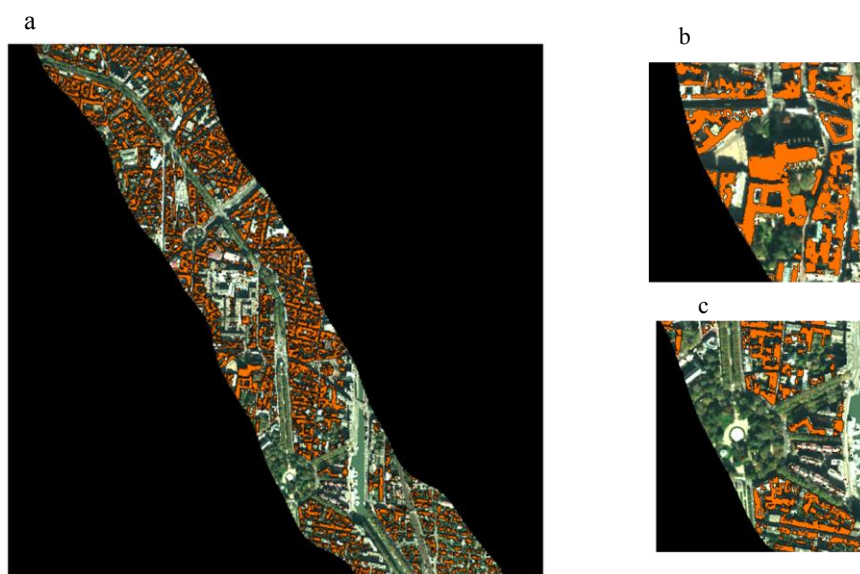


Figure 5 : Répartition des toits en tuile (a) la zone de survol (centre Toulouse)
b et c des extractions

4. Conclusion

L'ANR HYEP est l'occasion d'investiguer le milieu urbain avec une source d'image particulière. Comme tout domaine en constitution elle est le reflet des questionnements que se posent la communauté et tout particulièrement le groupe Hyperspectrale de la Société Française de Photogrammétrie et de Télédétection (SFPT). Les avancées des travaux permettront de baliser quelques développements concernant les algorithmes de correction atmosphériques, de fusion LIDAR et multispectral, de tests de variabilité des résultats selon l'approche de démixage retenue.

Références

- Baldridge A.M., S.J. Hook, C.I. Grove, G. Rivera. The ASTER spectral library version 2.0. *Remote Sensing of Environment*, 113 (2009), 711–715
- Clark, R.N., Swayze, G.A., Wise, R., Livo, E., Hoefen, T., Kokaly, R., Sutley, S.J., 2007, USGS digital spectral library splib06a: U.S. Geological Survey, Digital Data Series 231.
- Gadal S., Ouerghemmi W. (2016). Morpho-spectral objects classification by hyperspectral airborne imagery, Whisper, 21-24 August 2016, UCLA, USA.
- LeBris A., N Chahata, X Briottet, N Paparoditis, 2016, Spectral band selection for urban material classification using hyperspectral libraries. ISPRS. DOI: 10.5194/isprs-annals-III-7-33-2016.
- Kotthaus, S., Smith, T., Wooster, M. J. and Grimmond, C., 2014. Derivation of an urban materials spectral library through emittance and reflectance spectroscopy. *ISPRS Journal of Photogrammetry and Remote Sensing* 94, pp. 194–212.
- Roussel G., C Weber et X Briottet, 2016, Comparaison de deux méthodes de correction atmosphérique à différentes résolutions pour des applications de classification d'images hyperspectrales. SFPT Section Hyperspectrale Grenoble mai 2016.

Annexe(s) éventuellement

TETIS UMR 9000 Montpellier (WP0 et WP4)

ESPACE UMR 7300 Aix-Marseille Université (WP1)

UMR LETG Université de Rennes (WP3)

ONERA DOTA Toulouse (WP1 et WP3)

IGN Matis Paris (WP2)

IRAP Toulouse (WP2)

GIPSA Grenoble (WP2)

De l'intérêt de la morphologie mathématique pour la définition de la taille de la fenêtre de convolution lors de la réalisation d'une carte de densité

Hadrien Fouillade Orsini¹

*1. CNRS, UMR 7300 E.S.P.A.C.E., Université Nice Sophia Antipolis
98 boulevard Edouard Herriot, BP 3209-06204 Nice Cedex 3, France
hadrien.fouillade-orsini@unice.fr*

RESUME. La cartographie des points chauds permet d'observer la répartition spatiale de la criminalité. L'estimation de la densité par la méthode du noyau est une des méthodes les plus efficaces pour illustrer la distribution spatiale de la criminalité. La taille des cellules et de la fenêtre de convolution sont les deux paramètres principaux dont les valeurs doivent être fixées pour produire des résultats avec la technique KDE. Si la variation de la taille des cellules semble peu affecter les cartes des points chauds, ce n'est pas le cas de la taille de la fenêtre de convolution. Il est donc nécessaire de réfléchir à une méthode permettant de fixer de manière scientifique le rayon de recherche. Les opérations basiques de la morphologie mathématique peuvent être utilisées pour connaître la zone d'influence du phénomène étudié, en l'occurrence les crimes.

ABSTRACT. Hotspot mapping is used in order to distinguish the spatial pattern of crime. Kernel density estimation is one of the most efficient method to represent the spatial distribution of crime. If the variation of the cell size doesn't affects the crime hotspot map, the bandwidth size does. So, it seems necessary to find a method which can determine the search radius. Basic operations of mathematical morphology can be used to know the influence zone of the phenomenon analysed, which are the crimes in this case.

MOTS-CLES : géocriminologie, analyse spatiale, densité par la méthode du noyau, rayon de recherche, morphologie mathématique.

KEYWORDS: geocriminology, spatial analysis, kernel density estimation, bandwidth, mathematical morphology

1. Introduction

Les recherches dans le domaine de la géocriminologie ont démontré la concentration de la moitié des actes de délinquance recensés, dans des aires limitées de l'espace urbain, désignées comme « hotspots » ou points chauds (Sherman *et al.*, 1989 ; Weisburd et Green, 1995).

L'intérêt de l'analyse des points chauds est de pouvoir prédire dans quels secteurs des crimes semblables à ceux déjà cartographiés sont susceptibles d'apparaître. La méthode la plus utilisée pour identifier la distribution spatiale du crime est la méthode du noyau (Kernel Density Estimation) dont les représentations cartographiques ont qui plus est le mérite d'être esthétiques (Chainey, 2013).

Les systèmes d'information géographique permettent à partir d'un fichier de points géoréférencés de réaliser des cartes de points chauds en laissant l'utilisateur libre de paramétrer la taille des cellules et de la fenêtre de convolution. La fenêtre de convolution est également désignée comme le rayon de recherche.

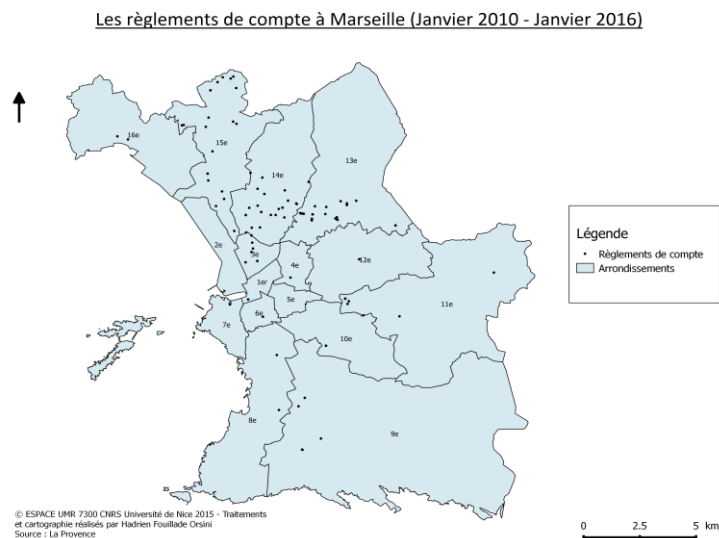
Or Spencer Chainey (2013) a constaté que si la variation de la taille des cellules a peu de conséquence sur les cartes de densité obtenues, il n'en est pas de même pour le paramétrage de la taille de la fenêtre de convolution. La capacité prédictive des cartes de points chauds diminue à mesure que les valeurs du rayon de recherche augmentent.

Il est par conséquent nécessaire de réfléchir à la mise en place d'une méthode permettant de fixer de manière scientifiquement justifiée la taille de la fenêtre de convolution. La morphologie mathématique parce qu'elle permet de déterminer la structure spatiale de la distribution du phénomène et connaître sa zone d'influence peut aider à définir le rayon de recherche lors de la réalisation d'une carte de points chauds.

2. Données

Chaque homicide recensé dans la commune de Marseille par la préfecture de police des Bouches-du-Rhône entre le 1^{er} janvier 2010 et le 1^{er} janvier 2016 et porté à la connaissance du grand public par les médias régionaux et nationaux a été géolocalisé à l'échelle de la rue sous forme de point grâce au recoupement des différentes sources d'information : articles de presse, photographie de la scène du crime, reportage vidéo (Figure 1).

SAGEO'2016 – Nice, 6-9 décembre 2016

FIGURE 1. *Cartographie des homicides à Marseille*

Au total pour la période considérée ce sont 116 homicides et tentatives de meurtre qui ont été géoréférencés sous environnement quantumgis 2.6.0 au moyen de l'extension « go2streetview » dont l'intérêt pour les études de criminologie a fait l'objet d'un article scientifique publié dans la revue *Crime Science* (Vandeviver, 2014). L'échelle d'utilisation de la base de données qui couvre l'ensemble de la commune de Marseille est le niveau infra-urbain. Les niveaux de granularité des données sont à la fois temporels et spatiaux. En termes spatiaux le niveau le plus fin est celui de la rue avec le renseignement des coordonnées x et y et le niveau le plus agrégé est celui de la commune. En termes temporels le niveau le plus fin correspond au jour et à l'heure du crime et le niveau agrégé correspond à la période comprise entre janvier 2010 et janvier 2016.

La transformation du fichier vecteur d'objets ponctuels représentant les crimes en fichier raster a été effectuée sous environnement Idrisi. L'échelle de la nouvelle image créée a été renseignée de telle sorte que la surface d'un pixel corresponde à 25 mètres carrés dans la réalité. La taille de l'image correspondait elle à la superficie des 16 arrondissements de Marseille à l'échelle 1 : 110 000. Puis la nouvelle image a été enregistrée en mode binaire afin de procéder à une transformation morphologique élémentaire : la dilatation, qui a été réalisée dans le logiciel libre ImageJ.

3. Méthode

La morphologie mathématique est l'une des méthodes appropriées pour repérer la structure spatiale d'un phénomène géographique au sein d'une image (Voiron, 1995). La dilatation qui est une des transformations élémentaires (Schmitt, Mattioli, 2013) devait permettre de connaître la zone d'influence du phénomène étudié afin de pouvoir paramétrer la taille de la fenêtre de convolution de manière rigoureuse en fonction des résultats obtenus. Il s'agit ainsi de dilater les pixels contenant de l'information afin de dénombrer les points restants à chaque pas de dilatation pour repérer les potentiels effets de palier. Les dilatations ont été opérées dans le logiciel libre ImageJ en ayant recours à l'élément structurant 3*3 préprogrammé (Figure 2).

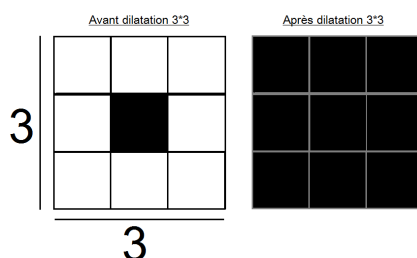


FIGURE 2. Effet de la dilatation 3*3 dans un logiciel de traitement d'images

4. Résultats

Le nombre de composantes restantes après chaque dilatation fut reporté dans un tableur afin de créer un graphique permettant de déterminer le seuil de rupture de pente de la courbe granulométrique. Le calcul effectué indique qu'au-delà du quarante-sixième pas, équivalant à une distance réelle de 230 mètres, la significativité apportée par la distance n'est plus probante (Figure 3).

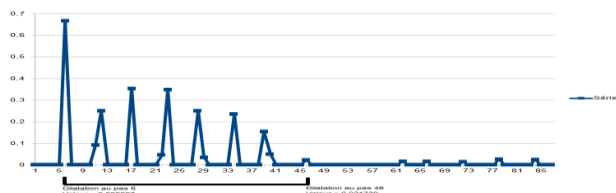


FIGURE 3. Zone d'influence des crimes à Marseille

SAGEO'2016 – Nice, 6-9 décembre 2016

Par conséquent la taille de la fenêtre de convolution dont le paramétrage est nécessaire pour réaliser la cartographie des points chauds est fixée à 230 mètres. L'intérêt du lissage est qu'il permet de remplir les vides et de s'affranchir des limites administratives, à savoir dans le cas présent le contour des arrondissements (Figure 4).

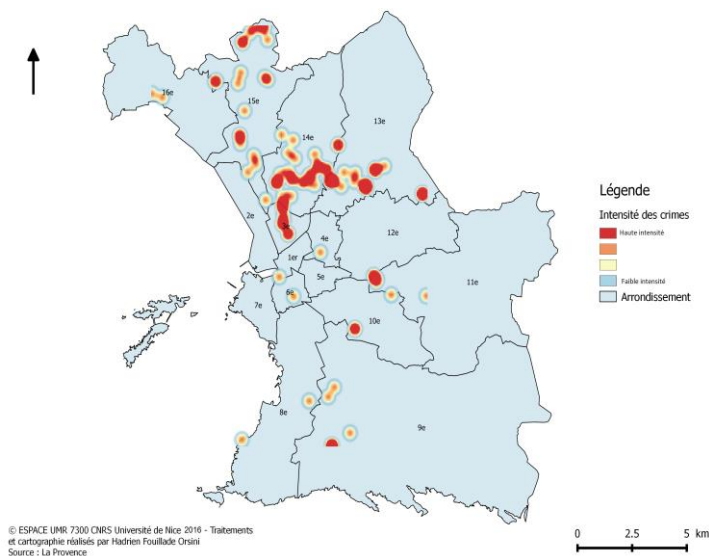


FIGURE 4. *Cartographies des points chauds de crimes à Marseille avec la technique KDE*

La carte des points chauds renseigne sur la concentration des crimes dans les 3^{ème}, 13^{ème}, 14^{ème} et 15^{ème} arrondissements où sont localisés la plupart des grands ensembles de logements sociaux que compte la ville et où sévit le trafic de drogue.

Bibliographie

- Chainey S. (2013). Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime. *Bulletin of the Geographical Society of Liege*, n°60, p. 7-19.
- Schmitt M., Mattioli J. (2013). *Morphologie mathématique*, Presses des Mines, Paris.
- Vandeviver C. (2014). Applying google maps and google street view in criminological research. *Crime Science*, vol 3, n°13, p.1-16.
- Voiron-Canicio C. (1995). *Analyse spatiale et analyse d'images par la morphologie mathématique*, GIP RECLUS, Montpellier.
- Weisburd D., Green L., (1995) Policing drug hotspots : the Jersey city drug market analysis experiment. *Justice quarterly*, vol 12, n° 4, p. 711-735.

Démarche d'analyse de l'intégrité d'un système de localisation de navires

Clément IPHAR¹, Aldo NAPOLI¹, Cyril RAY²

1. CRC, MINES ParisTech, PSL Research University
Rue Claude Daunesse, F-06560 SOPHIA ANTIPOLIS, FRANCE
{clement.iphar ; aldo.napoli}@mines-paristech.fr

2. Institut de Recherche de l'Ecole Navale (IRENav)
BRCM Brest, F-29240 BREST, FRANCE
cyril.ray@ecole-navale.fr

RESUME. Le système d'identification automatique (AIS) est un système électronique permettant aux navires d'envoyer et de recevoir des messages de positionnement. Ces messages sont utilisés à des fins diverses, telles que la prévention des abordages ou la surveillance du trafic. Cependant les messages envoyés via ce système contiennent des erreurs et subissent des falsifications. Cet article propose une méthode d'analyse des données transmises via ce système, méthode basée sur l'étude de l'intégrité des données contenues au sein même des messages, et ce à plusieurs niveaux de profondeur dans un message et entre les messages, le but étant d'attribuer à chaque message un coefficient de confiance relatif à aux déterminations de l'intégrité des données en son sein.

ABSTRACT. The Automatic Identification System (AIS) is an electronic system which allows the vessels to send and receive positioning messages. Those messages are used for several purposes such as boarding prevention or traffic monitoring, amongst others. However, the messages are subject to errors and they undergo falsifications. This paper proposes a methodology for the analysis of data broadcasted via this system, based on the integrity of data contained within the messages themselves, this being done at several levels of depth in one lone message and between different messages, the purpose being to assign to each message a coefficient revealing the degree of confidence computed by the proposed integrity-based method.

MOTS-CLES : AIS ; falsification de données ; intégrité des données ; détection d'anomalies.

KEYWORDS: AIS ; data falsification ; data integrity ; anomaly detection.

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

Carrefour des enjeux internationaux, l'espace maritime est soumis à une activité de plus en plus intense. Cette croissance des mobilités maritimes a conduit à la généralisation de systèmes de suivi de navigation. Parmi ces systèmes, on peut citer les radars marins, des satellites optiques ou radars à synthèse d'ouverture. L'un de ces systèmes est le Système d'Identification Automatique (AIS).

Des mécanismes de malversation et des pratiques de navigation sont inévitablement apparues récemment pour contourner, falsifier, exploiter de tels systèmes de surveillance dans l'intérêt des contrevenants. C'est la découverte de ces malversations, notamment au travers de l'étude des données et de leur intégrité au sein des messages AIS qui constitue le cœur de notre démarche de recherche.

2. Principes et limite du système d'identification automatique

2.1. Un système pour améliorer la sécurité de la navigation

L'utilisation du système AIS a été instaurée en 2000 par l'Organisation Maritime Internationale, pour "Tous les navires de jauge brute supérieure ou égale à 300 engagés dans des voyages internationaux, les navires de charge de jauge brute supérieure ou égale à 500 et les navires à passagers, quel que soit leur tonnage" (IMO, 2004). Le système utilise des transpondeurs pour envoyer des messages par ondes VHF (Very High Frequency) aux navires alentours et aux stations côtières.

2.2. Les défauts du système

Trois principaux types d'actions menant à une perte de qualité des données peuvent être recensés : l'erreur, la falsification et le piratage. Les informations erronées transmises peuvent être fausses, incomplètes, impossibles selon la norme ou impossibles physiquement. La falsification est le fait de dégrader volontairement un message par la modification d'une valeur authentique ou l'arrêt volontaire et non motivé de l'envoi des messages, fait dans le but de tromper l'extérieur. D'après (Harati-Mokhtari et al., 2007), environ 50% des messages contiendraient des données erronées et 1% des navires diffusent des données falsifiées. Par ailleurs, le piratage de message est effectué par un acteur extérieur par la création de faux messages et leur transmission sur les fréquences AIS (Balduzzi et al., 2014).

3. Une méthode basée sur l'intégrité des données pour la découverte de falsifications

Notre recherche se concentre sur les problématiques basées sur le message, en portant un intérêt tout particulier à l'information présente au sein même de chacun

des messages AIS et à sa qualité. Ainsi, la notion d'intégrité de la donnée est utilisée à des fins d'analyse (Iphar et al., 2015).

Les messages AIS sont divisés en champs, et ces champs peuvent prendre différentes valeurs en accord avec les spécifications techniques du système. De plus, ces données peuvent être de types différents : numérique représentant un identifiant (Numéro OMI), numérique représentant une quantité ou une grandeur physique (Latitude), numérique représentant un choix dans une liste (statut de navigation), textuel (nom du navire), de type date (Date d'arrivée estimée) ou de type binaire.

3.1. Détermination de l'intégrité des messages AIS

En prenant en considération les données en provenance de tous les messages AIS, notre proposition de démarche se base sur quatre moyens de tester l'intégrité de ces données. La première voie consiste en le contrôle de l'intégrité interne à un message : la conformité aux spécifications techniques de chaque champ de chaque message est étudiée, ainsi que l'intégrité au sein des champs de ce seul message. La deuxième voie consiste en le contrôle d'intégrité de l'information entre plusieurs messages : les messages de même type ont les mêmes champs, et il est donc possible de comparer les valeurs de ces champs et de déterminer leur intégrité entre différents messages, par ailleurs la comparaison et la détermination de l'intégrité de la donnée AIS entre champs de différents messages est également possible, en effet, même si les données proviennent de différents messages, du fait de l'existence de champs similaires et liés dans différents messages, l'intégrité de certaines informations est vérifiable. La troisième voie est une méthode statistique pour évaluer la similarité d'un message avec les messages d'un jeu de données (en partant du principe que dans un groupe de messages, issus d'un même navire ou d'une même aire géographique, les valeurs de certains champs auront tendance à se ressembler). La quatrième voie consiste en l'utilisation de bases de données externes afin de recouper des informations (par exemple la base des données des navires de pêches de l'Union Européenne étant disponible librement en ligne, il est aisé de vérifier certaines informations envoyées par les navires, telles que nom ou longueur).

3.2. Cas d'études

Les méthodes de détermination d'intégrité et les méthodes statistiques vues précédemment nécessitent un jeu de données qui doit être sélectionné afin de pourvoir à une étude donnée, telles que proposées dans cette partie. Nous avons sélectionné deux études principales et deux études secondaires, qui sont basées pour les premières sur la station et le MMSI (numéro unique d'identifiant de navire), et pour les secondes sur une zone et la route. Les hypothèses proposées ici sont que des messages sélectionnés selon l'une de leurs caractéristiques (même station de réception, même émetteur) auront tendance à se ressembler, et donc que certaines anomalies, dans ce contexte, auront tendance à émerger.

SAGEO'2016 – Nice, 6-9 décembre 2016

Par leur nature, les études basées sur la station et sur le MMSI peuvent être effectuées aussi bien à la volée qu'avec des données archivées alors que les études basées sur une zone et la route ne peuvent s'effectuer que basées sur des données archivées à des fins d'analyse.

Afin de réaliser cette étude, nous avons développé un parser AIS opérationnel, adapté du parser AISMESSAGE disponible en ligne. Par ailleurs, l'architecture de base de données a été réalisée, construite avec une base de données relationnelle utilisant une architecture PostgreSQL/PostGIS. Le processus d'analyse en cours de développement utilise la base de données à l'aide de scripts Python qui extraient les informations issues des traitements présentés dans la sous-partie précédente.

4. Conclusion

Les vulnérabilités de l'AIS, qui conduisent à la présence d'erreurs et de falsifications au sein des messages que le système transmet aux autres navires et aux systèmes de surveillance, sont présentées dans ce papier. C'est l'intégrité en tant que dimension caractéristique de la qualité de la donnée que nous vérifions par une approche basée sur le message et plus particulièrement sur les informations contenues au sein des champs de données des messages. Les travaux en cours se concentrent sur le développement d'un système fondé sur une architecture de base de données, et l'usage de bases de données contextuelles (de type météorologiques, ou registres de pêche). Ces travaux pourront en outre se concentrer sur la définition d'alertes en fonction de l'analyse des mouvements, des risques potentiels engendrés et des éléments d'intégrité issus des analyses proposées dans cet article.

Remerciements :

Ces travaux sont financés par l'ANR sous la référence ANR-14-CE28-0028 dans le cadre du projet DéAIS, cofinancés par la DGA, et labélisés par le Pôle Mer Bretagne Atlantique et le Pôle Mer Méditerranée.

Références

- Balduzzi M, Pasta A., Wilhoit K. (2014). A security evaluation of AIS automated identification system. *Actes de la 30th annual computer security applications conference 2014*, La Nouvelle-Orléans, USA.
- Harati-Mokhtari A., Wall A., Brooks P., Wang J. (2007). Automatic Identification System (AIS): a human factors approach. *Journal of Navigation*. Vol 60(3).
- IMO (2004). International Maritime Organization. *International convention for the safety of life at sea*.
- Iphar C, Napoli A, Ray C. (2015). Detection of false AIS messages for the improvement of maritime situational awareness. *Actes de la conférence OCEANS'15 Washington, USA*

Distance discrète de Fréchet optimisée

Devogele Thomas¹, Esnault Maxence¹, Etienne Laurent¹

1. Laboratoire d'informatique de l'Université François Rabelais
64 avenue Jean Portalis
37200 Tours France
prenom.nom@univ-tours.fr

RESUME. Le calcul de distance entre des polygones est essentiel pour mesurer la similarité entre les géométries de deux objets linéaires. Cet article décrit une nouvelle version de la distance de Fréchet discrète. Cette version optimise grandement le temps de calcul et améliore la précision du résultat. Ces propriétés sont indispensables pour manipuler de gros volumes de polygones ayant des nombres d'extrémités de segments importants.

ABSTRACT. Distance computation between polylines is a key point to assess similarity between geometrical objects. This paper describes a new optimized algorithm to compute discrete Fréchet distance which lower computation time and improved precision. Thanks to this algorithm, big data polyline repositories can be mined.

MOTS-CLES : Distance de Fréchet, distance linéaire, appariement, mesure de similarité

KEYWORDS: Fréchet distance, linear distance, data matching process, similarity measure

1. Introduction

Les Systèmes d'Information Géographique (SIG) disposent de fonctionnalités de comparaison d'objets géométriques. Ces fonctionnalités s'appuient sur différents types de mesures de similarité. Parmi ces dernières, les mesures de distances entre deux géométries tiennent une place prépondérante. Pour la comparaison de points, des mesures de distances exactes existent. En ce qui concerne les objets linéaires et plus particulièrement les polygones, la **distance de Fréchet** (Alt et Godau 1992) est couramment employée. Elle renvoie une distance maximale entre des points en tenant compte de l'ordre des points. Une illustration intuitive de cette distance est la suivante : un maître et son chien suivent deux chemins. Ils avancent ou s'arrêtent à volonté, indépendamment l'un de l'autre, mais ils ne peuvent pas revenir sur leurs pas. La distance de Fréchet entre ces deux chemins est la longueur minimale de la laisse qui permet de réaliser un cheminement de concert satisfaisant ces conditions. La version discrète (Devogele 2002) plus facile à calculer ne prend en compte que les points extrémités des segments des polygones. Elle s'appuie sur deux matrices. La première matrice (matrice des distances : MD) stocke les distances entre les points des deux polygones L_1 et L_2 . La deuxième (matrice de Fréchet : MF) calcule de manière dynamique la distance de Fréchet.

SAGEO'2016 – Nice, 6-9 décembre 2016

La valeur de la cellule MF_{ij} est calculée de la manière suivante : $MF_{ij} = \max(MD_{ij}, \min(MF_{i-1,j}, MF_{i,j-1}, MF_{i-1,j-1}))$. Pour des polygones de taille N, M , la distance de Fréchet discrète (D_{dF}) est la valeur de la cellule $MF_{n,m}$. La figure 1(a) fournit les deux matrices du couple de lignes de 4 points de la figure 1(b). Pour cet exemple la D_{dF} est égale à 0.9 et est portée par le couple de points $L_{1,3}$ et $L_{2,4}$. Cette distance est approximative. En prenant en compte en plus des points sur les segments, une valeur plus faible et plus précise peut être obtenue. Pour ces matrices, un cheminement de concert est une suite de cellules de la matrice partant du coin haut gauche et arrivant au coin bas droit. Les cellules grisées de la matrice de distance de la figure 1 forment un cheminement possible respectant une laisse de longueur 0.9.

1. Les trois optimisations proposées

Cet article propose des optimisations de cette mesure. Les objectifs de cette version optimisée sont d'améliorer la précision du résultat et de réduire drastiquement le temps de calcul pour des lignes composées d'un nombre élevé de points. Trois étapes d'optimisations sont proposées. Premièrement, une étape de **filtrage** à l'aide de l'algorithme de Douglas-Peucker (1973) est définie. Le temps de calcul de la distance de Fréchet étant proportionnel aux nombres de points extrémités des segments de deux polygones, la diminution du nombre de points est donc fondamentale. Cet algorithme supprime les points non significatifs tout en conservant la forme de la polygone. Par exemple, pour les polygones issus de relèvement GPS, les points sur des parties quasi droites des polygones sont négligeables.

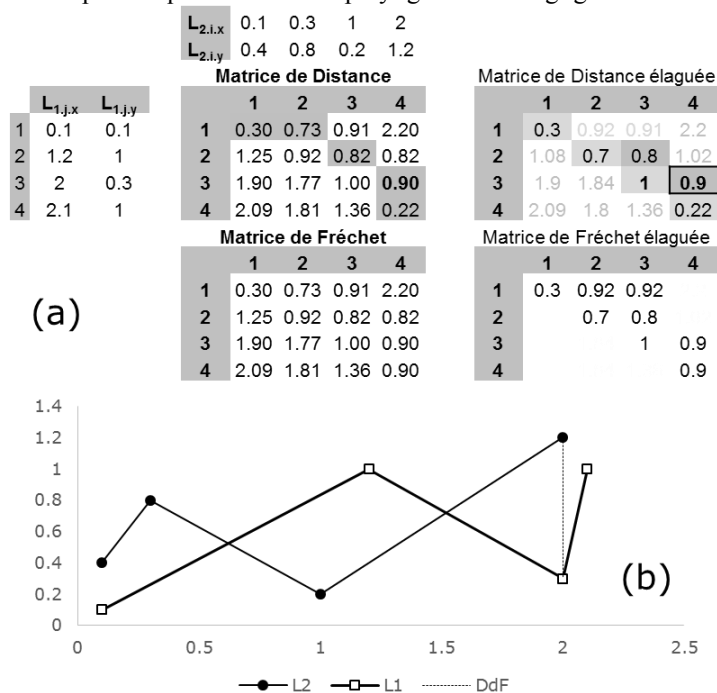


Figure 1 : Exemple de calcul de la distance de Fréchet discrète

La deuxième optimisation est l'**élagage** des deux matrices. En effet, il n'est pas utile de calculer les valeurs des cellules qui ne font pas partie potentiellement d'un cheminement optimal. Pour cela, dans un premier temps, les valeurs de cellules formant un cheminement entre les deux angles de la MD sont calculées. Pour notre exemple de la figure 1, le cheminement est la diagonale formée par les cellules en gris clair sur la matrice de distance élaguée. Le maximum de ces valeurs définit une borne. Pour l'exemple, cette borne vaut 0.9. Les cellules au-dessus et en-dessous de ce cheminement sont alors calculées jusqu'à ce qu'elles dépassent cette borne. Seules les cellules ayant une valeur inférieure à cette borne sont stockées. Pour l'exemple, seules deux nouvelles cellules en gris foncé sont stockées en plus de la diagonale. La matrice de Fréchet élaguée ne va s'appuyer que sur les valeurs de la matrice de distance élaguée.

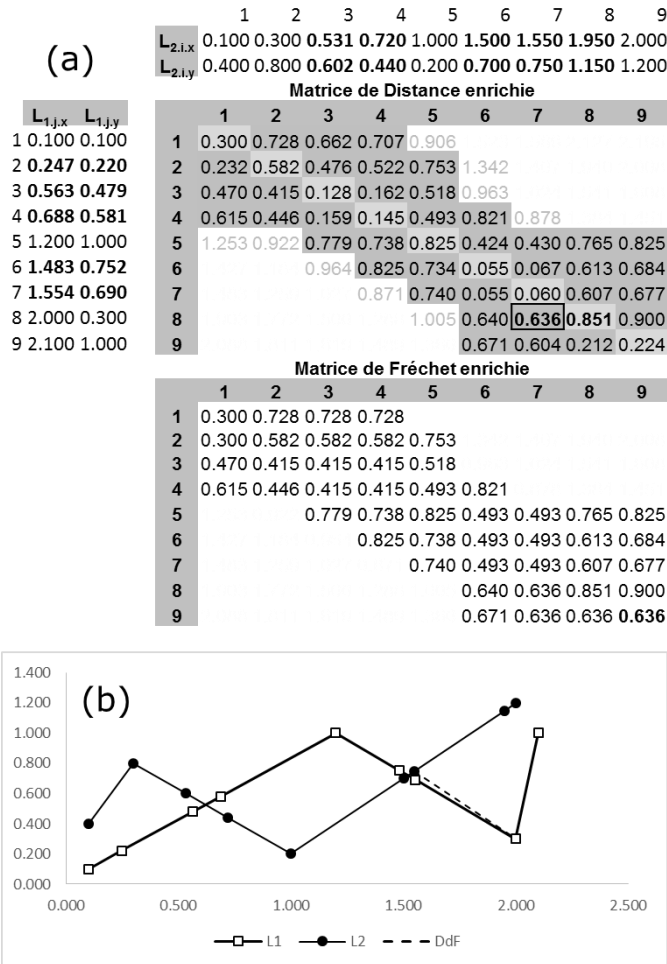


Figure 2 : Matrices et lignes enrichies de l'exemple de la figure 1

SAGEO'2016 – Nice, 6-9 décembre 2016

Après ces deux premières optimisations du temps de calcul, la troisième optimisation améliore la précision. Elle consiste à **enrichir** les matrices en ajoutant les points les plus saillants des segments : les projetés orthogonaux des extrémités sur les segments de l'autre polyligne. Pour un point i , seul sont considérés les segments dont au moins un des deux points j définit une cellule $[i,j]$ des matrices élaguées. La figure 2(a) présente les matrices enrichies de l'exemple. 5 points projetés (coordonnées en gras) sont ajoutés à chacune des lignes (figure 2(b)). Les matrices élaguées sont alors à nouveau calculées. La D_{af} est maintenant de 0.636 et est portée par les points $L_{1.8}$ et $L_{2.7}$.

2. Expériences sur un large jeu de données

Une expérience a été menée sur un jeu de données 495 trajectoires de pigeons (Freeman et al. 2011) représentant 1 533 521 points GPS (4 m en moyenne entre deux points consécutifs). Ce jeu de données est conséquent. Il faut 5h 22 min pour calculer la distance de Fréchet discrète pour les 122 265 couples de trajectoires. En filtrant à 5 m, ce qui correspond à la précision de ces GPS, 70 714 points (4.61%) sont conservés. A l'aide de la méthode optimisée, il ne faut plus que 9min 39sec pour calculer les distances. En ce qui concerne la précision, l'erreur médiane n'est que de 55cm et de 7m au pire. L'étape d'enrichissement est indispensable afin de compenser la suppression de 95.39% des points. Sans cette étape, l'erreur peut atteindre 213m.

3. Conclusion

Un algorithme optimisé de mesure de la distance de Fréchet discrète a été présenté dans cet article. Le temps de calcul est réduit de manière significative. Cette réduction est indispensable pour traiter les gros volumes de lignes issues entre autre de capteurs GPS. La phase d'enrichissement améliore aussi la précision de la mesure. Cependant, il reste à démontrer que le résultat obtenu est égal à la distance de Fréchet (non discrète) pour des polygones et à le tester pour d'autres jeux de données volumineux.

4. Références

- Alt, H., & Godau, M. (1992). Measuring the resemblance of polygonal curves. In Proceedings of the eighth annual symposium on Computational geometry (pp. 102-109). ACM.
- Devogele, T. (2002). A new Merging process for data integration based on the discrete Fréchet distance. In Advances in spatial data handling (pp. 167-181). Springer Berlin Heidelberg.
- Douglas, D. & Peucker, T. (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, The Canadian Cartographer 10(2), 112–122
- Freeman, R., Mann, R., Guilford, T., & Biro, D. (2011). Group decisions and individual differences: route fidelity predicts flight leadership in homing pigeons (*Columba livia*). *Biology letters*, 7(1), 63-66.

Dynamiques paysagères et dysfonctionnement (SLC-off) du satellite Landsat.

La morphologie mathématique comme outil d'interpolation de l'occupation du sol.

Matthieu VIGNAL¹

*1. UMR ESPACE, Université Nice Sophia Antipolis
98, bd E. Herriot, 06204 Nice Cedex
matthieu.VIGNAL@unice.fr*

RESUME. Les analyses des changements d'utilisation et d'occupation du sol sont un élément essentiel dans l'étude des relations Nature - Société. Depuis une trentaine d'années, le suivi de ces changements, notamment des dynamiques paysagères, est effectué à partir de données satellites en raison de la résolution spatiale, spectrale et temporelle de ces données. Peu après son lancement, le satellite Landsat 7 subit un dommage technique (SLC-off). Les prises de vue comportent alors des « zébrures » sans information qui empêchent une cartographie complète et donc, une estimation correcte de l'ampleur et de la forme des phénomènes observés. Ce poster a pour objectif de présenter une méthode de reconstruction de ces « zébrures » sans information. La méthode repose sur l'utilisation des outils qu'offre la morphologie mathématique. Elle pose le postulat que le phénomène observé suit une progression dans le temps et l'espace. La méthode de reconstruction mise en place donne des résultats satisfaisants d'une part, dans un objectif de quantification de surface et, d'autre part, dans une optique d'analyse morphologique.

ABSTRACT. Studies of land-use and land-cover change are an essential element in nature - society relationship research. For thirty years, tracking these changes, including landscape dynamics, is made from satellite data due to spatial, spectral and temporal resolution of this data. Shortly after its launch, the Landsat 7 satellite suffered technical damage (SLC-off). The shots then comprise "stripes" without information that prevents a complete map and, a correct estimation of the size and shape of the observed phenomena. The aim of this poster is to present a reconstruction method of this "stripes" without information. It postulates that the observed phenomenon follows a progression in time and space. This reconstruction method provides satisfactory results in surface quantification goal and, secondly, in view of morphological analysis.

MOTS-CLES : Dynamiques paysagères - Morphologie mathématique - Reconstruction morphologique – Landsat - SLC-off

KEYWORDS: Landscape dynamics - Mathematical morphology - Morphological reconstruction – Landsat - SLC-off

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction.

Les transformations économiques, l'urbanisation, le changement climatique et les risques naturels sont autant de facteurs qui façonnent la planète (Kuenzer et al, 2015). Les forêts primaires et secondaires sont défrichées au profit de monoculture de rente ou de pâtures (Oswald, 2005). *De facto*, les analyses des changements d'utilisation et d'occupation du sol sont devenues des éléments essentiels (Kuenzer et al, 2015) de grands projets internationaux qui visent à acquérir de nouvelles connaissances sur les relations Nature - Société (*Intergovernmental Panel on Climate Change – IPCC*).

Depuis une trentaine d'années, le suivi de ces changements est effectué à partir de données satellites en raison de la résolution spatiale, spectrale et temporelle de ces données (Andrieu, 2008). Elles forment des données particulièrement efficaces pour caractériser l'utilisation et l'occupation du sol à différentes échelles spatio-temporelles, et cela, dans des domaines d'étude variés notamment dans le suivi des dynamiques paysagères (Andrieu, 2008). Dans l'étude des changements, les images Landsat sont les plus utilisées par la communauté scientifique, en raison d'une part, de leur résolution spatiale, spectrale et temporelle (Andrieu, 2008) et, d'autre part, de la gratuité de ces données, qui permet leurs utilisations par l'entièreté de la communauté scientifique. Cependant, peu après le lancement du satellite Landsat 7, une défaillance technique apparaît sur le système optique *Scan Line Corrector* (SLC) qui corrige le parallélisme des images. Les prises de vue comportent alors des « zébrures » sans information. Pour remédier à cela, des méthodes efficaces s'appliquant aux images brutes ont été élaborées (Zhu et al, 2011). Or, la littérature sur des méthodes post-traitements des images est quasi-inexistante. Ces défauts techniques ont ainsi éloigné la communauté scientifique de ces images en laissant bien souvent de côté cette période (2003-2013) dans les analyses diachroniques de l'occupation du sol. Or, certains phénomènes se déroulent essentiellement pendant cette période, dès lors, il convient de mettre en place une méthode, post-traitement des images, de reconstruction de ces « zébrures » sans information.

La méthode proposée ici est appliquée à la Forêt Classée du Haut-Sassandra (FCHS), située au centre-ouest de la Côte d'Ivoire. Cette forêt a subi un important processus de défrichement pendant la période de crises politico-militaires des années 2000. Ce poster s'inscrit dans le projet de recherche D2PCPCI qui consiste à étudier des dynamiques paysagères en période de crises politico-militaires en Côte d'Ivoire.

2. Données.

D'abord, il est essentiel de se procurer des cartes d'occupation du sol à des dates qui ne présentent pas de « zébrures » sans information. Ainsi, deux cartes d'occupation du sol en 2002 et 2013 forment les données de base de cette étude (Barima et al, 2016). Ces cartes, d'une résolution de 30 mètres, comportent quatre classes d'occupation du sol : Forêt, Jachères et agriculture, Sol nu et Eau. Ensuite, afin de cartographier l'occupation du sol à des dates intermédiaires, des images Landsat ont été mobilisées. Il s'agit d'images Landsat 7 aux dates suivantes : 19-01-2005 ; 11-12-2007 ; 15-09-2009 ; 04-01-2011 et 01-07-2012 (USGS).

3. Méthode.

L'occupation du sol est, en 2002 dans la FCBS, une grande matrice forestière composée de quelques taches agricoles. Or, entre 2002 et 2013, ces taches ont augmenté en nombre et en taille. Pour cette raison, la méthode pour reconstruire l'information manquante repose sur la morphologie des taches agricoles à travers l'utilisation des outils qu'offre la morphologie mathématique (Voiron-Canicio, 1995). La méthode pose le postulat que le phénomène observé suit une progression dans le temps et l'espace donc, que le phénomène à une date T résulte du phénomène à une date T - 1.

La méthode (cf. figure 1) suit les étapes suivantes : les zones agricoles de la date T - 1 (2012) sont dilatées d'un pixel puis intersectées avec les zébrures sans information de cette même date. Il en résulte alors une couche avec des pixels qui représentent les espaces agricoles présents au contact des zébrures sans information. Ces pixels vont être dilatés géodésiquement selon le masque des zones agricoles de la date T (2013) selon un pas de dilatation nécessaire pour compléter les zébrures dans leurs largeurs, soit au troisième pas de dilatation. Le résultat est alors associé aux zones agricoles de la date T - 1 (2012). Ce qui n'est pas repris en espaces agricoles peut alors être considéré comme des espaces forestiers.

Cette méthode est itérative, le résultat de la date T - 1 (2012) peut alors servir de données complètes pour la reconstruction des espaces agricoles à la date T - 2 (2011). Ce processus itératif a permis la reconstruction des cartes d'occupation du sol pour les dates 2005, 2007, 2009, 2011 et 2012 (cf. figure 2). Bien entendu, l'observation de ces résultats amène nécessairement à aborder la question de leur validation. La validation s'est déroulée en deux étapes : par la comparaison des pourcentages d'occupation du sol d'une part, et l'analyse de la forme des taches agricoles d'autre part. Cette dernière est réalisée par comparaison des indices de forme, calculés par le logiciel Fragstat, entre les taches reconstruites et les taches observées.

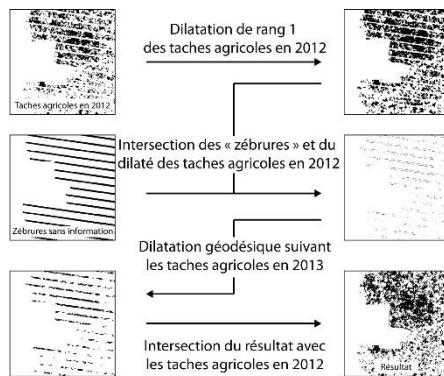


FIGURE 1. Organigramme méthodologique.

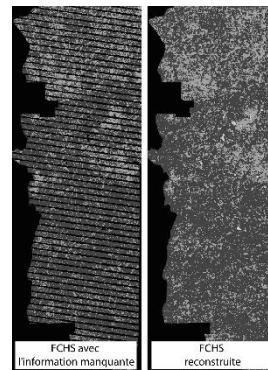


FIGURE 2. Résultat de la reconstruction pour 2012.

SAGEO'2016 – Nice, 6-9 décembre 2016

4. Résultats.

Les pourcentages d'espaces agricoles des données brutes et transformées sont relativement semblables. Des différences existent pour les dates 2007, 2009 et 2011, mais celles-ci n'excèdent pas les 3 % pour les différences les plus marquées. La marge d'erreur de la méthode d'un point de vue quantitatif semble être de 3 %. Cette méthode apparaît cohérente au regard de l'évolution de la forme des taches agricoles. Les résultats montrent des formes légèrement plus complexes que celles observées par télédétection, la marge d'erreur est de 6 %, soit des formes 1,06 fois plus complexes. Cette méthode semble donc valide. Les cartes d'occupation du sol issues de cette reconstruction morphologique peuvent désormais être utilisées afin d'analyser les changements paysagers.

5. Discussion et conclusion.

Dans l'étude des dynamiques paysagères, la période de dysfonctionnement du satellite Landsat est souvent ignorée de la communauté scientifique en raison des « zébrures » sans information que comportent les données. Or, il semble que des outils comme la morphologie mathématique permettent de récupérer ces données. La méthode de reconstruction mise en place donne des résultats satisfaisants d'une part, dans un objectif de quantification de surface et, d'autre part, dans une optique d'analyse morphologique. Bien que les différentes méthodes de validation donnent satisfaction, il convient de se demander dans quelle mesure peut-on analyser la forme d'une image reconstruite d'une part, et d'autre part, dans quelle mesure peut-on analyser la forme des taches reconstruites à partir des taches d'une autre date. Afin d'apporter des éléments de réponses supplémentaires à ces interrogations, il serait particulièrement intéressant de tester cette méthode sur différents environnements et différents phénomènes afin de lui accorder un crédit plus important.

Bibliographie.

Andrieu J. (2008). *Dynamique des paysages dans les régions septentrionales des Rivières-du-Sud (Sénégal, Gambie, Guinée-Bissau)*. Thèse de doctorat de Géographie, Université de Paris Diderot - Paris 7.

Barima Y. S. S., Kouakou A. T. M., Bamba I., Sangne Y. C., Andrieu J. (2016). Cocoa crops are destroying the forest reserves of the classified forest of Haut-Sassandra (Ivory Coast). *Global Ecology and Conservation*, n° 8, p. 85-98.

Chen J., Zhu X., Vogelmann J. E., Gao F., Jin S. (2011). A simple and effective method of filling gaps in Landsat ETM+ Slc-off images. *Remote Sensing of Environment*, Vol. 15, n° 4, p. 1053-1064.

Kuenzer C., Dech S., Wagner W. (2015). *Remote sensing time series: revealing land surface dynamics*, Springer.

Oswald J. (2005). *Dynamique des formations agroforestières en Côte d'Ivoire (des années 1980 aux années 2000)*. Thèse de doctorat de Géographie. Université des Sciences et Technologies de Lille.

Voiron-Canicio C. (1995). *Analyse spatiale et analyse d'images*, Reclus.

Etude du paramètre turbidité par télédétection (Landsat, et Spot5-Take5) : indicateur « d'aléa sanitaire » dans les eaux de surface en Afrique de l'Ouest

Robert Elodie¹, Grippa Manuela², Kergoat Laurent³, Martinez Jean-Michel⁴, Pinet Sylvain⁵, Somdecoste Tom⁶, Gosset Cindy⁷, Gal Laetitia⁸

¹ GET (CNRS, IRD, Université de Toulouse 3),
14 avenue Edouard Belin, 31400 Toulouse, France.
elodie.robert@get.obs-mip.fr

²GET (CNRS, IRD, Université de Toulouse 3), manuela.grippa@get.obs-mip.fr

³GET (CNRS, IRD, Université de Toulouse 3), laurent.kergoat@get.obs-mip.fr

⁴GET (CNRS, IRD, Université de Toulouse 3), martinez@ird.fr

⁵GET (CNRS, IRD, Université de Toulouse 3), sylvain.pinet@get.obs-mip.fr

⁶Tom.somdecoste@gmail.com

⁷GET (CNRS, IRD, Université de Toulouse 3), cindy.gosset@get.obs-mip.fr

⁸GET (CNRS, IRD, Université de Toulouse 3), laetitia.gal@get.obs-mip.fr

Les maladies hydriques, 5^{ème} cause de mortalité à travers le monde et 2^{ème} dans les pays à faible revenu, constituent toujours au XXI^{ème} siècle un enjeu sanitaire majeur (OMS, 2008). Le risque sanitaire premier demeure d'origine microbiologique. Il cause notamment des **maladies diarrhéiques**, qui touchent chaque année 1,7 milliard de personnes et causent 2 millions de décès dont majoritairement des enfants de moins de 5 ans, essentiellement dans les Pays du Sud (OMS, 2013). Les pathogènes responsables de ces maladies dans ces pays sont principalement les rotavirus et des bactéries d'origine fécales. A partir de la fin des années 1950, plusieurs études (Dennis 1959 ; Tombes et al. 1979 ; Hoff et Geldreich 1981) ont montré qu'il existe une importante corrélation entre une forte concentration de parasites, de bactéries et certains paramètres de qualité de l'eau dont la turbidité (Santé Canada, 2004). Les matières en suspension peuvent véhiculer des virus et des bactéries pathogènes pour la santé humaine (Brock, 1996 ; Stotzky 1966) et favoriser leur développement (adsorption des agents pathogènes sur les matières en suspension, présence de nutriment propice à leur croissance). Ainsi une forte turbidité (synonyme de processus d'adsorption, de « blocage » du rayonnement solaire et marqueur d'érosion hydrique qui est une possible source de bactéries d'origine fécales) peut être considérée comme un vecteur de contaminants microbiologiques, donc comme un « aléa sanitaire » à l'origine notamment de maladies diarrhéiques (la valeur limite de qualité de l'eau en France est 1 NFU – Nephelometric Formazine Unit - proche de 1 NTU – Nephelometric Turbidity Unit).

La **turbidité** correspond à la « *propriété optique de l'eau permettant à la lumière incidente d'être déviée ou absorbée par des particules plutôt que transmise en ligne droite* » (INSPQ, 2003). Les données de télédétection dans le domaine du visible et du proche infrarouge (PIR) permettent le suivi de différentes caractéristiques des eaux de surface, notamment leur turbidité (Martinez *et al.*, 2009 ; Kaba *et al.*, 2014, Robert *et al.*, 2016). Ainsi, Martinez *et al.* (2009) ont trouvé des relations empiriques solides entre les matières en suspension (MES) et la réflectance de surface en faisant correspondre des images Modis (250 m) et des prélèvements de terrain. Espinoza Villar *et al.* (2012) ont intégré des données satellites et des prélèvements de terrain pour suivre des affluents du fleuve Amazone au Pérou. Ils ont constaté que ces données pouvaient être utilisées pour étudier les concentrations de matières en suspension. Toutefois, peu d'études portent sur ce paramètre en Afrique Subsaharienne, exception faite des travaux de Kaba *et al.* (2014) sur le lac de Tana en Ethiopie, alors que les niveaux de turbidité dans les eaux continentales africaines sont généralement très élevés et présentent de fortes variabilités spatio-temporelles causées par la variabilité intra et inter-annuelle de la mousson Africaine. Cette forte dynamique spatio-temporelle des mares et des lacs en zone tropicale, plus particulièrement en Afrique de l'Ouest, requièrent alors une très bonne résolution spatiale et une bonne répétitivité temporelle pour le suivi de ces objets, comme celle fournie par les données **Landsat** (couplage L7 et L8) et plus particulièrement **Spot5-Take5** et **Sentinel-2**. Ce n'est donc que très récemment, 2013 pour Landsat 8 et 2015 pour Sentinel-2, que les satellites disposent de la résolution spatiale (10

à 30 m) et le temps de revisite¹ qui permettent de suivre les eaux de surfaces (lacs, mares, rivières et réservoirs) sur de vastes territoires. Ces données sont maintenant accessibles librement, ce qui en fait un outil très intéressant pour les suivis environnementaux dans les pays du Sud où les réseaux de mesures sont de plus en plus présents. Une étape critique est d'établir si ces données, qui sont des mesures de rayonnement réfléchi, dans différentes longueurs d'onde, permettent une quantification suffisamment précise des paramètres des eaux comme la turbidité et les matières en suspension. Un certain nombre d'obstacles sont susceptibles de compliquer cette tâche et d'empêcher de dériver des relations robustes entre observations satellite et turbidité : la nature des matières en suspension et des pigments présents dans l'eau, les gammes de variations très larges de la turbidité typiques des milieux tropicaux semi-arides ou subhumides, les perturbations liées à l'atmosphère (aérosols, feux de biomasse) pour n'en citer que trois.

L'objectif de cette étude est donc d'étudier l'apport des images satellites à haute résolution (Landsat 7 et 8, et Spot5-Take5) pour le suivi et la cartographie d'un « **aléa sanitaire** », la turbidité, en milieu tropical.

Zone d'étude

Trois terrains d'étude ont été retenus : la mare d'Agoufou au Mali (Service National d'observation – SNO AMMA-CATCH), le fleuve Niger Moyen et le lac de Bagré au Burkina Faso. Ils permettent d'étudier 3 unités hydrologiques (un réservoir sur un grand fleuve – le Nakambé, une mare permanente en zone endoréique et le fleuve majeur de la région) ainsi que 2 régimes climatiques (sahélien et nord soudanien) différents. Ces régions ont connu des changements de régimes hydrologiques (comme le « paradoxe sahélien » en zone non cultivée, Gardelle *et al.*, 2010) qui interrogent l'évolution du paramètre turbidité, et pointent une augmentation généralisée des ruissellements, comme conséquence des sécheresses sahéliennes dramatiques des années 70 et 80. De plus, les vulnérabilités sanitaires dans ces régions sont très fortes : de nombreuses populations dépendent encore **d'eaux de surface insalubres** pour leurs usages domestiques (consommer l'eau, se laver, nettoyer les légumes, laver les ustensiles de cuisine) et récréatifs (jouer).

Données

Des **relevés d'eau (turbidité – turbidimètre, MES – filtrage et pesée)** sont réalisés par des partenaires locaux sur les différents sites :

- 153 mesures (turbidité et MES) sur la mare d'Agoufou (15.36°N - 1.47°W) depuis septembre 2014 par le SNO AMMA-CATCH (hebdomadaire) ;
- 97 mesures (turbidité et MES) au niveau de Niamey, sur le fleuve Niger Moyen, depuis juin 2015 (hebdomadaire) ;
- sur le lac de Bagré : tous les 5 j du 16 avril au 13 septembre 2015 pendant l'acquisition Spot5-Take5, et actuellement hebdomadaires (total de 159 relevés pour la turbidité et 131 pour les MES).

Des prélèvements d'eau (turbidité et matières en suspension) ont aussi été effectués lors d'une **mission terrain** sur différentes zones du lac de Bagré du 20 juillet au 6 août 2015 parallèlement à des mesures radiométriques (3 radiomètres TriOs RAMSES) pour calculer des réflectances de surface *in situ* et ainsi évaluer les relations turbidité / radiométrie *in situ* puis à partir des réflectances de surface des images satellites (Landsat et Spot5-Take5).

Des images Landsat (L7 et L8 Land Surface Reflectance) sans nuage sur les zones de prélèvement d'échantillon d'eau sont acquises pour chacun des sites d'étude sur le site de l'USGS (United States Geological Survey). Enfin pour le site de Bagré, 16 images satellites Spot5-Take5 ont été également utilisées. Il s'agit de l'usage du satellite SPOT en fin de vie sur une orbite basse pour obtenir, sur un certain nombre de sites sélectionnés, une répétitivité de 5 jours avec une résolution de 10 m. Cette résolution spatio-temporelle est équivalente au futur Sentinel-2A et B. Le satellite SPOT a été utilisé comme un simulateur des futures séries temporelles Sentinel-2.

Méthodes

Les algorithmes proposés par la littérature pour les eaux continentales (R, PIR, PIR/R, PIR + R, PIR/B, PIR/V, R-PIR/R+PIR) sont appliqués aux images des différents capteurs en s'appuyant sur la comparaison avec les données terrain (turbidité, MES) sur les 3 sites, afin de caractériser les niveaux de turbidité, leurs évolutions temporelles et d'effectuer des comparaisons entre sites. Les mesures radiométriques collectées sur le lac de Bagré

¹ La résolution temporelle est de 10 jours pour les images Sentinel-2A (disponibles avec corrections atmosphériques à partir de décembre 2016 pour l'Afrique) à 16 jours pour Landsat 7 et 8. Toutefois, le « couplage » L7 et L8 peut permettre l'acquisition d'image tous les 8 jours. Enfin, lorsque le satellite Sentinel-2B sera lancé la résolution temporelle de Sentinel (A et B) sera de 5 jours.

sont aussi exploitées pour l'évaluation des algorithmes sur ce site. La méthode s'appuie sur un modèle empirique MES/réflectance existant (Martinez et al. 2009, Espinoza Villar et al. 2012). Elle a été appliquée en Afrique de l'Ouest par Robert et al. (2016) et repose sur 3 étapes : la mise en regard des données de MES et de turbidité ; l'étude des corrélations entre les bandes ou ratios de bandes d'une part des données radiométriques terrain et d'autre part des données Landsat ou Spot avec les mesures de turbidité et de MES ; l'étude des variations spatio-temporelles (inter et intra-annuelle) de la turbidité et des MES sur les différents sites. Ainsi, 96 images Landsat² ont été retenues pour être comparées aux prélèvements d'eau (MES et de turbidité). L'écart maximal retenu est de 2 jours par rapport à la date de prélèvement.

L'algorithme retenu permet également de réaliser des cartes d'aléas suite à l'inversion des valeurs.

Résultats

Les matières en suspension et la turbidité présentent assez logiquement une grande variabilité : de 9 mg/l à 2784 mg/l et de 9.9 NTU à plus de 1000 NTU (limite du turbidimètre). Les trois types de surface d'eau (fleuve, mare et lac) montrent un maximum saisonnier marqué, l'eau la plus turbide étant celle de la mare d'Agoufou. Malgré ces valeurs extrêmement élevée, cette eau est régulièrement consommée par les populations locales et responsables de maladies.

De bonnes à très bonnes relations sont observées entre les mesures de turbidité et de MES d'une part pour chaque site ($R^2=0.72$ pour Agoufou, $R^2=0.82$ et 0.97 pour Bagré respectivement pour les données Landsat et Spot5-Take5, $R^2=0.98$ pour Niamey) et d'autre part pour l'ensemble des sites ($R^2=0.83$, Figure 1). Les meilleures corrélations observées avec les images satellites reposent sur l'usage du PIR associé au rouge : ratio (pour Spot5-Take5, $R^2=0.88$) ou addition (pour Landsat, $R^2=0.85$, Figure 2). Enfin pour le site de Bagré, la meilleure corrélation observée entre les mesures radiométriques terrain et les MES concerne le ratio PIR/R ($R^2=0.97$).

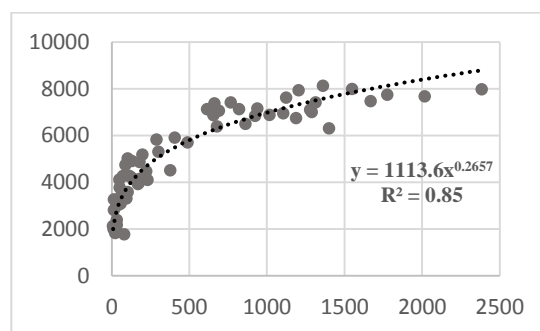
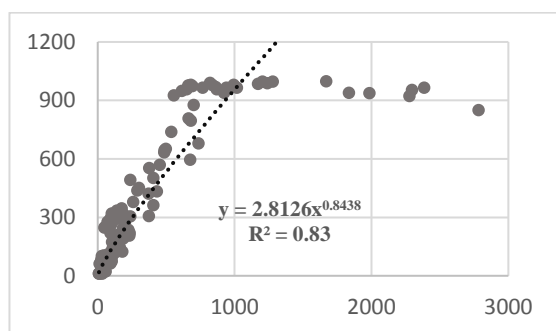


FIGURE 1. Turbidité fonction des MES à partir de mesures in-situ : Lac de Bagré, Mare d'Agoufou et Niamey

FIGURE 2. Ratio PIR+R des images Landsat fonction des MES pour les 3 sites : Bagré, Agoufou et Niamey

Ces ratios sont alors utilisés pour étudier les variations saisonnières de la turbidité. Sur le lac de Bagré, les premiers événements pluvieux (mars, avril) ont pour conséquence une augmentation simultanée des valeurs du PIR/R (Spot5-Take5) et traduisent donc une augmentation de la turbidité (Figure 3). A cette période de l'année, de nombreux espaces du bassin versant peuvent contribuer aux matières en suspension : les sols nus et les sols destinés à la mise en culture peu protégés par la végétation. Ces valeurs continuent d'augmenter jusqu'au milieu de la saison des pluies (août), puis diminuent. L'analyse du cycle saisonnier moyen met également en évidence des **variations spatio-temporelles** de la turbidité au sein même du lac, surtout durant la saison des pluies. Il y a une contribution importante depuis l'amont, à partir de juin, puis un transport des particules de l'amont vers l'aval entre juin et septembre (pic en juillet-août pour les zones centres, pic en septembre pour les zones aval), et finalement une diminution régulière à la fin de la saison des pluies (septembre, octobre) indiquant une dilution et des phénomènes de dépôt.

Le site d'Agoufou présente aussi d'importantes évolutions saisonnières, mais avec des valeurs de turbidité et de MES plus élevées. Les valeurs de turbidité augmentent dès les premières pluies en juin avec un maxima observé de juillet à fin août, puis une diminution à partir d'octobre. Toutefois une nette différence est observée entre la saison sèche de 2015 et 2016, avec un maximum secondaire observé en février 2016 causé probablement par la remobilisation de sédiments par les vents ou par le piétinement du bétail.

² La mise en correspondance de ces données s'étend sur toutes les périodes de l'année.

Le Niger au niveau de Niamey présente une augmentation plus rapide de la turbidité entre mi-juin et fin août, associée à la 'cruie rouge' (couleur de la charge sédimentaire), avec un maxima fin juillet-début août, puis une forte baisse associée à la 'cruie noire' ('Cruie Guinéenne').

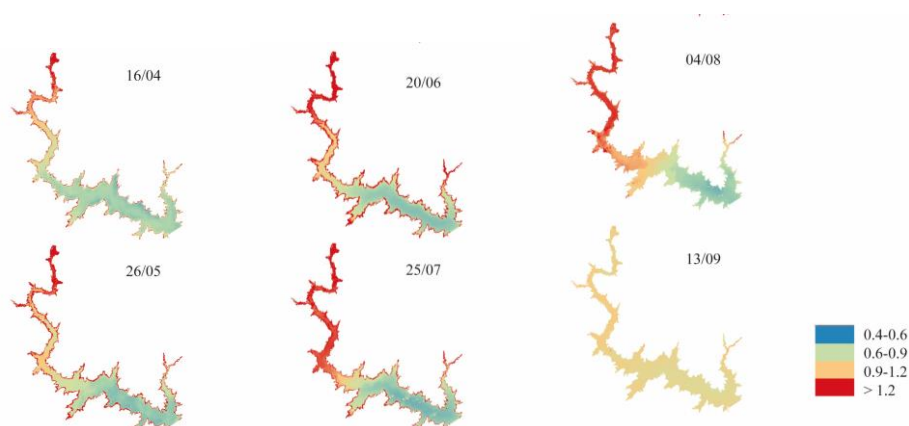


FIGURE 3. Evolution spatiotemporelle de la turbidité dans le Lac de Bagré (PIR/R, images Spot5-Take5)

Ces bonnes corrélations entre le ratio PIR/visible ou PIR + visible et la turbidité conduisent à la réalisation de cartes exposant l'évolution spatio-temporelle de la turbidité au niveau de ces différentes unités hydrographiques. Cette cartographie permet la réalisation de cartes d'aléa (variation spatio-temporelles intra-annuelles). La mare d'Agoufou présente tout au long de l'année une forte turbidité. Le fleuve Niger à Niamey est plus turbide pendant la saison des pluies (en particulier lors de la première crue, appelée 'cruie rouge'), mais la seconde crue présente des eaux moins turbides. Le lac de Bagré est en situation intermédiaire, avec une variation au sein du lac : valeurs plus élevées dans les zones amont et pic plus précoce. Toutefois, le minimum saisonnier de turbidité reste au-dessus des normes de consommation d'eau, synonyme de risque sanitaire.

Conclusion

Les données satellitaires fournissent donc des informations sur l'exposition des populations et sur la propagation, l'évolution de l'aléa. Combinées à des données épidémiologiques et à la prise en compte des pratiques et des usages, il peut être alors possible de cartographier les risques pour la santé des populations. Les nouvelles images Sentinel-2 associant une bonne résolution spatiale et temporelle (10 m et 5 jours) représentent donc un fort potentiel pour ce suivi et une occasion unique pour la télé-épidémiologie des eaux continentales en Afrique de l'Ouest.

Bibliographie

- Brock T.D. - 1966 - *Principles of microbial ecology*. Prentice-Hall, NJ, 72 p.
- Dennis J.M. - 1959 - Infectious hepatitis epidemic in Delhi, India, *J. Am. Water Works Assoc.*, 51 : 1288.
- Espinoza Villar, R., Martinez, J.-M., Guyot, J.-L., Fraizy, P., Armijos, E., Crave, A., Bazan, H., Vauchel, P., Lavado, W., 2012. The integration of field measurements and satellite observations to determine river solid loads in poorly monitored basins. *J. Hydrol.* 444, 221–228, <http://dx.doi.org/10.1016/j.jhydrol.2012.04.024>.
- Gardelle J., Hiernaux P., Kergoat L., Grippa M. - 2010 - Less rain, more water in ponds: a remote sensing study of the dynamics of surface waters from 1950 to present in pastoral Sahel (Gourma region, Mali), *Hydrol. Earth Syst. Sci.* 14, pp. 309-324.
- Hoff J.C., Geldreich E.E. - 1981 - Comparaison of the biocidal efficiency of alternative disinfectants, *J. Am. Water Works Assoc.*, pp. 40-44.
- INSPQ - 2003 - Turbidité, fiche technique turbidité, par le Groupe scientifique, juin 2003.
- Kaba E., Philpot W., Steenhuis T. - 2014 - Evaluating suitability of MODIS-Terra images for reproducing historic sediment concentrations in water bodies: Lake Tana, Ethiopia, *International journal of applied earth observation and geoinformation*, 26, pp. 286-297.

Martinez J-M., Guyot J-L., Filizola N., Sondag F. - 2009 - Increase in suspended sediment discharge of the Amazon River assessed by monitoring network and satellite data, *Catena*, pp. 257-264.

OMS - 2008 - Directives de qualité pour l'eau de boisson, 2^{ème} édition, Critère d'hygiène et documentation à l'appui, vol. 2.

OMS - 2013 - Statistiques sanitaires mondiales, 172 p.

Robert E., Grippa M., Kergoat L., Martinez J-M., Pinet S., Gal L. - 2016 - Monitoring water turbidity and surface suspended sediment concentration of the Bagre Lake (Burkina Faso) using MODIS and field reflectance data, *International Journal of Applied Earth Observation and Geoinformation*.

Tombes A.S., Abernathy A.R., Welch D.M., Lewis S.A. - 1979 - The relationship between rainfall and nematode density in drinking water, *Water Res.*, 13.

Stotzky G. - 1966 - Influence of clay minerals on microorganisms. III. Effect of particle size, cation exchange capacity, and surface area on bacteria. *Can. J. Microbiol.*, vol. 12, n°4, pp. 831-848.

Évaluation de la sensibilité littorale à l'érosion en région PACA dans un contexte de changement climatique.

Mavromatidi Asimina¹, Briche Elodie¹

1. Aix Marseille Univ., IRD, LPED

MC3- Mediterranean Cities and Climate Change, project financé par les Investissements d' Avenir A*MIDEX

3, place Victor Hugo, 13331 – Marseille Cedex 3, France

semimavr@gmail.com, elodie.BRICHE@univ-amu.fr

RÉSUMÉ. *Un tiers de la population française se concentre sur les zones côtières méditerranéennes. Le GIEC (2013) prévoit une hausse du niveau marin d'environ 0,53 à 0,98 m à l'horizon 2100 et ces zones seront d'autant plus sujettes à l'érosion. L'objectif principal est de déterminer les côtes de la région PACA les plus sensibles à l'érosion à partir d'un indice de sensibilité côtière (ISC) qui repose sur 6 variables telles que la géomorphologie, la pente côtière, l'élévation relative du niveau de mer, le changement du littoral, l'amplitude moyenne de la marée et la hauteur de vague significative. Cette étude est réalisée à l'échelle régionale à partir de données gratuites et/ou «open data». La cartographie montre que 17,66 % de la zone d'étude appartient à la catégorie de sensibilité très élevée. Cette étude donne un aperçu de la sensibilité littorale à l'érosion en région PACA afin de mieux anticiper les changements à venir et les intégrer dans les aménagements des villes méditerranéennes.*

ABSTRACT. *A third of the French population is concentrated on the Mediterranean coastal areas. The IPCC (2013) predicts an increase of about 0.53 to 0.98 m in sea-level rise in 2100 and these areas are particularly prone to erosion. The main objective is to determine the most susceptible to erosion coasts of the region PACA, from a coastal sensitivity index (CSI) based on the following 6 variables: geomorphology, coastal slope, sea-level rise, shoreline changes, mean tidal range and significant wave height. This study is held at a regional scale with free data and / or «open data». The result shows that 17.66% of the study area, belongs to the very high sensibility category. This study gives an overview of coastal erosion sensitivity in the region of PACA, in order to anticipate better in future changes and integrate them into the amenities of Mediterranean cities.*

MOTS-CLES : *Indice de Sensibilité Côtière, érosion, SIG, élévation du niveau de la mer, données «open data»*

KEYWORDS: *Coastal Sensitivity Index, Erosion, GIS, sea-level rise, open data*

1. Introduction

Les villes méditerranéennes, par leurs configurations spatiales et géographiques (proximité du littoral, forte dépendance à la ressource en eau, climat favorisant les sécheresses estivales et les événements météorologiques extrêmes, comme les pluies intenses) et par leur croissance, sont considérées comme des milieux particulièrement vulnérables, “*hot-spot*” du changement climatique. La région PACA est très exposée au risque d'érosion côtière (Lambert *et al.*, 2007) et les bordures littorales et fronts de mer (notamment des zones urbaines) sont aménagés avec des digues ou enrochement visant à protéger les constructions des villes littorales. Déterminer les franges littorales sensibles à l'érosion, dans le cas d'une élévation potentielle du niveau de la mer, permet d'identifier les zones où il faudra probablement penser ou repenser les aménagements urbains littoraux. Cette étude a pour objectif l'évaluation de la sensibilité littorale à l'érosion en région PACA, en utilisant au maximum des données gratuites voire libres (« open data »). Elle est réalisée à partir de l'indice de sensibilité côtière (ISC) (Shaw *et al.*, 1998). La cartographie résultante permet de déterminer quel paramètre de l'indice influe le plus sur la sensibilité à l'érosion mais également de régionaliser la frange littorale PACA en cinq catégories de sensibilité (de très faible à très élevée). Cet indice permet d'amorcer des réflexions sur l'adaptation des zones côtières au changement climatique en reconnaissant l'ampleur et la cause du problème.

2. Données et méthodes

Les indicateurs de sensibilité côtière ont été développés comme une méthode rapide et cohérente pour caractériser la sensibilité relative des différentes côtes. L'indice de sensibilité côtière (ISC) proposé par Shaw *et al.* (1998) a été appliqué sur la base de données gratuites et disponibles de campagnes de mesures de terrain mais aussi de données topographiques et géomorphologiques.

L'indice est calculé selon la formule ci-dessous:

$$(ISC) = \sqrt{\frac{(a*b*c*d*e*f)}{6}} \quad (1)$$

Où : (a) est la géomorphologie, (b) : la pente côtière (%), (c) : le taux de l'élévation relative du niveau de la mer (mm/an), (d) : le taux de changement du littoral (m/an) (e) : l'amplitude moyenne de la marée (m), et (f) : la hauteur de vague significative (m).

Les paramètres mentionnés ci-dessus sont classés sur une échelle linéaire de 1 : très faible sensibilité à l'érosion à 5 : très élevée sensibilité à l'érosion selon Pendleton *et al.*, 2004. La classification comprend donc 5 catégories/classes de sensibilité à l'érosion avec des critères spécifiques, pour chaque paramètre de l'indice (catégories non montrées).

SAGEO'2016 – Nice, 6-9 décembre 2016

TABLE 1. Sources des données pour chaque paramètre.

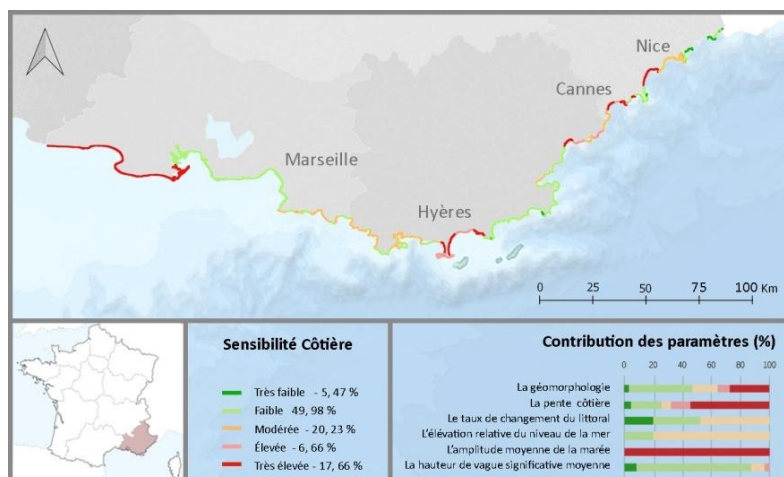
Paramètres (a)	Google Earth / Cartes géologiques nationales, http://portal.onegeology.org/ , (1: 50.000), 2011
Paramètres (b)	SRTM de Consortium pour l'information spatiale (CGIAR-CSI) résolution 90 m, 2004
Paramètres (c)	Mesures satellitaires (1992-2013) par Agence européenne pour l'environnement (EEA), 2014
Paramètres (d)	Geoportail / Google Earth, (1: 68,244), 1977 and 2015
Paramètres (e)	Service Hydrographique National (SHOM), 2016
Paramètres (f)	Base de données national de mesures des vagues (CANDHIS), 2002-2016

La démarche de la classification et du traitement des données va présenter au poster.

3. Résultats

Les valeurs calculées de l'indice en région PACA varient entre 2,58 et 23,72. Un pourcentage de 17,66 % de la zone d'étude incluant la Camargue, Hyères, Fréjus, Villeneuve-Loubet, Cannes et Cagnes-sur-Mer est classé comme ayant une très grande sensibilité à l'érosion (figure 1). De plus, 6,66 %, 20,23 %, 49,98 %, et 5,47 % sont classés respectivement comme ayant une sensibilité grande, modérée, faible et très faible.

Figure 1: Indice de Sensibilité Côtière (ISC) de la Région PACA



Source des données topographiques: INSEE, Amante et Eakins NGDC ETOPO1 (2009), modifié par les auteurs

Toutes les variables intégrées dans l'évaluation de l'indice peuvent être considérées comme des facteurs qui contribuent au changement côtier. Cependant, certaines contribuent davantage que d'autres : l'amplitude moyenne de la marée provoque une très grande sensibilité sur toute la zone d'étude. Pour la Camargue, Hyères et Cannes, la géomorphologie et la pente côtière influencent davantage le résultat final de l'indice. Fréjus, Villeneuve-Loubet et Cagnes-sur-Mer sont quant à elles des villes affectées en priorité par la pente côtière. Des six variables, l'indice peut être expliqué surtout par l'amplitude moyenne de la marée, la géomorphologie et la pente côtière en région PACA.

3. Conclusion et Discussion

Dans la présente étude, l'analyse a été volontairement menée à l'échelle régionale en raison de la disponibilité des données nécessaires à l'analyse. Bien que l'utilisation de données ouvertes (*open data*) permet de nouvelles approches, il faut toutefois toujours faire preuve de prudence et de recul quand à ces données d'un nouveau genre en matière de précision, d'exhaustivité ou de formalisme. De plus, tous les facteurs n'étaient pas disponibles en « *open data* » et un travail de cohérence spatiale et temporelle a été réalisé au préalable afin de pouvoir diffuser les résultats dans le cadre du projet MC3 « villes méditerranéennes et changement climatique ». Cette étude contribue en effet à la réflexion sur les villes littorales méditerranéennes dans le but de prioriser les efforts visant à améliorer la résilience naturelle de la côte, mais aussi à s'adapter au changement climatique en réfléchissant à des aménagements durables prenant en compte les variables qui contribuent le plus à la sensibilité côtière. Cependant, il est nécessaire de noter que l'indice a été calculé en utilisant les données relatives à l'élévation du niveau de la mer comme prévu à partir de l'EEA 2014. Ainsi, tout changement dans les valeurs des paramètres pourrait changer le résultat final de l'indice.

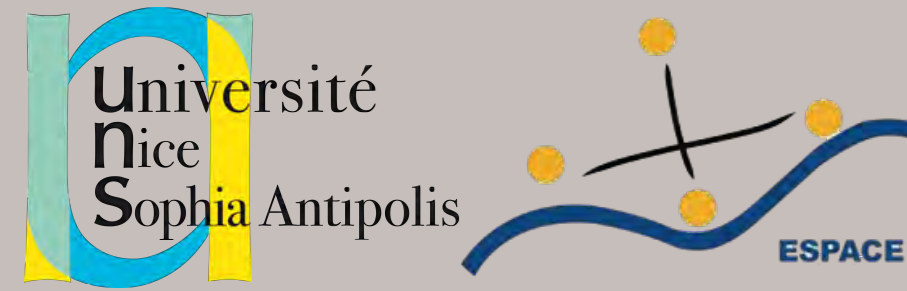
5. Bibliography

- GIEC (2013) Summary for policymakers. In Climate change 2013: The physical science basis. http://www.climatechange2013.org/images/report/WG1AR5_SPM_FINAL.pdf
- Lambert A., Rey V., Provansal M., Samat O. and Sabatier F. (2007). Lutte contre l'érosion littorale : efficacité des méthodes de stabilisation par drainage de plage, le cas de la baie d'Agay, Var. Méditerranée [Online], 108 | 2007, Online since 01 January 2009, connection on 09 June 2016. URL : <http://mediterranee.revues.org/184>
- Letetrel C. (2009) The Marseille tide gauge : recovery and analysis of high frequency sea level data from 1885. Final report for MedCLIVAR - Exchange Grant 1928.
- Pendleton, C., Thieler, E.R. and Williams, S.J., (2010), Importance of coastal change variables in determining vulnerability to sea-level change, Journal of Coastal Research, 26(1), 176-183.
- Shaw J., Taylor R.B., Forbes D.L., Ruz M.H., Solomon S., (1998). Sensitivity of the Canadian Coast to Sea-Level Rise. Geological Survey of Canada Bulletin, 505, 114

Expérimentation d'une cartographie collaborative pour le développement d'une procédure participative dans le cadre de la gestion durable des espaces verts.

L'exemple du quartier Ouest Résidentiel à Antibes Juan-les-Pins

Rondelli Cynthia, UMR 7 300 ESPACE, Université Nice Sophia-Antipolis



Résumé :

A partir de l'exemple des espaces verts du quartier Ouest Résidentiel de la commune d'Antibes Juan-les-Pins, nous illustrons la mise en place d'une procédure participative grâce à la cartographie. L'objectif est d'établir une gouvernance des espaces verts dans le cadre de leur gestion. A l'aide de la plateforme cartographique collaborative numérique OpenStreetMap (OSM) et de la méthode de la Cartopartie un atelier de groupe est expérimenté comme cas d'étude pilote : VERT (Voir et Recenser sur mon Territoire). OSM est une organisation collaborative en ligne qui collecte des données afin de créer des cartographies libres et détaillées du monde entier. La Cartopartie consiste à réunir divers individus sur le terrain dans le but de compléter les informations d'une zone géographique. Associée à un protocole collaboratif, nous visons à renforcer les liens sociaux et à augmenter le pouvoir de participation des individus tout en valorisant le travail du gestionnaire et le développement d'une politique durable des espaces verts.

I. Introduction :

Cet atelier pilote cartographique est réalisé avec la collaboration de 10 volontaires : représentants d'associations, citoyens et élus. Nous souhaitons **construire collectivement une cartographie des espaces verts** de l'entrée de ville du quartier Ouest Résidentiel d'Antibes Juan-les-Pins. Trois catégories d'informations sont recensées :

- les types d'espaces verts (jardin d'enfant, parc, bois, prairie, rond-point...),
- les systèmes d'arrosage (système par aspersion, goutte-à-goutte, tuyau d'arrosage, présence d'eau potable ou non, présence de robinets ou de fontaines...),
- les caractéristiques des massifs (arbuste, arbre, fleur, haie, gazon, végétation méditerranéenne ou non...).



Espaces verts de l'entrée de ville Parc Exflora, Square Henri Delaunay et Ambroisie Lauvert

C'est par la création d'informations géolocalisées que la connaissance et la compréhension des espaces verts, leurs modes de gestion et d'intervention tendraient à être améliorés.
L'enjeu est d'étudier les espaces verts d'un **point de vue spatio-organisationnel** afin de **hiérarchiser les priorités d'action pour le gestionnaire.**

La méthodologie de notre travail consiste à utiliser un outil libre, ludique et convivial : OpenStreetMap. **L'objectif est de contribuer à l'amélioration de la qualité des informations** sur ce secteur en mettant en avant le rôle de la participation citoyenne et celui du gestionnaire. VERT permettra ainsi de **produire des renseignements**. L'atelier pilote est divisé en **trois phases** : une sortie en groupe dans le secteur choisi, appelée « **Cartopartie** », suivie d'une séance en salle où les **données recueillies sur le terrain sont entrées dans la plateforme OSM**. Une base de données est réalisée ainsi qu'une **représentation cartographique** sur le thème de la végétation et systèmes d'arrosage des espaces verts.

L'hypothèse de ce travail est que le **développement des nouvelles technologies de l'information**, telles que les smartphones ou les GPS et les outils du numérique **constituent des clés indispensables au transfert de connaissance entre acteurs du territoire**. Ils permettraient d'une part au citoyen de faire remonter ses lectures de l'espace et ses attentes spatiales et d'autre part au géographe de fournir le sens d'un espace (Dubus et al., 2010).

II. La procédure participative :

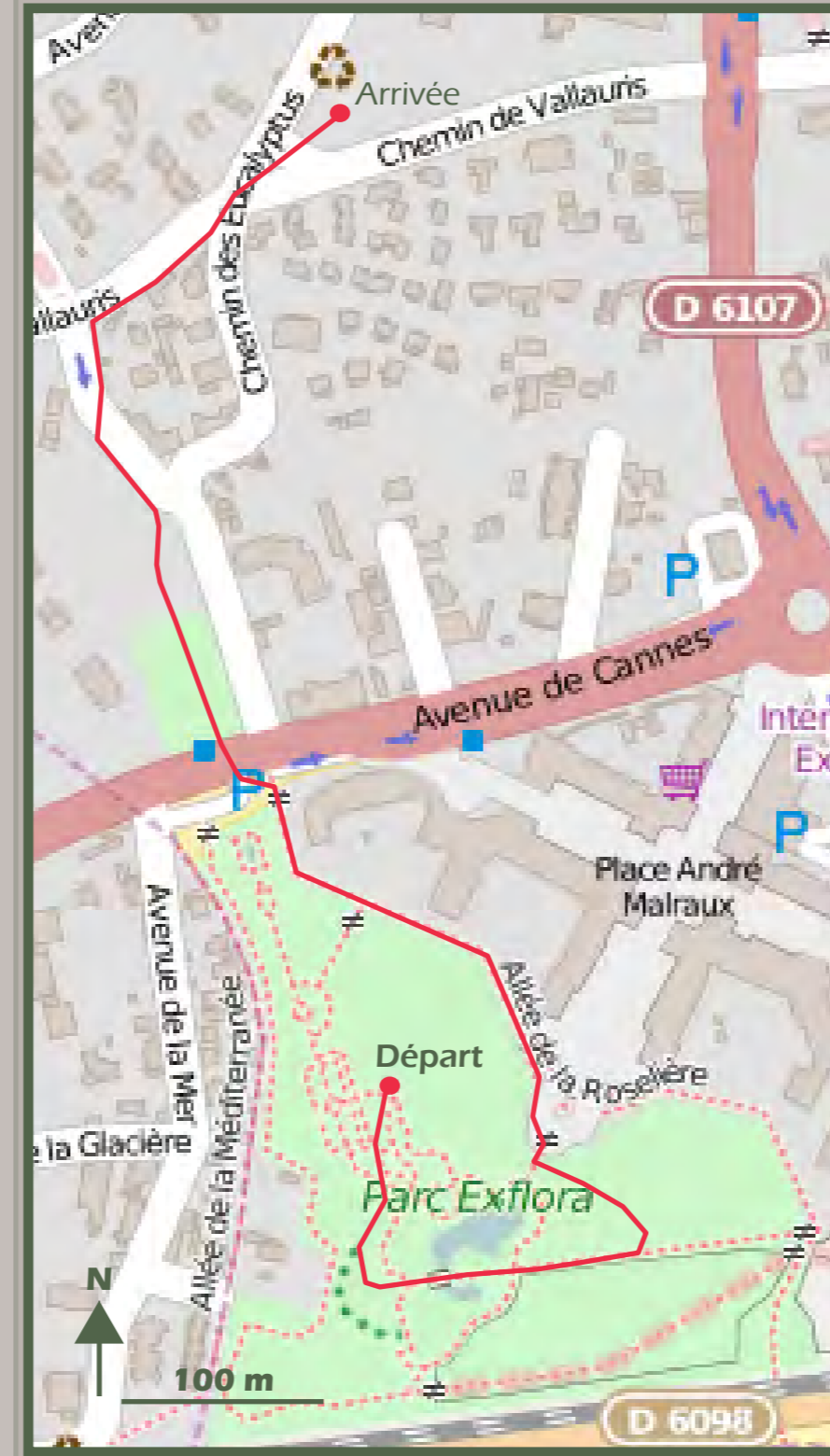
PHASE 1 : La Cartopartie



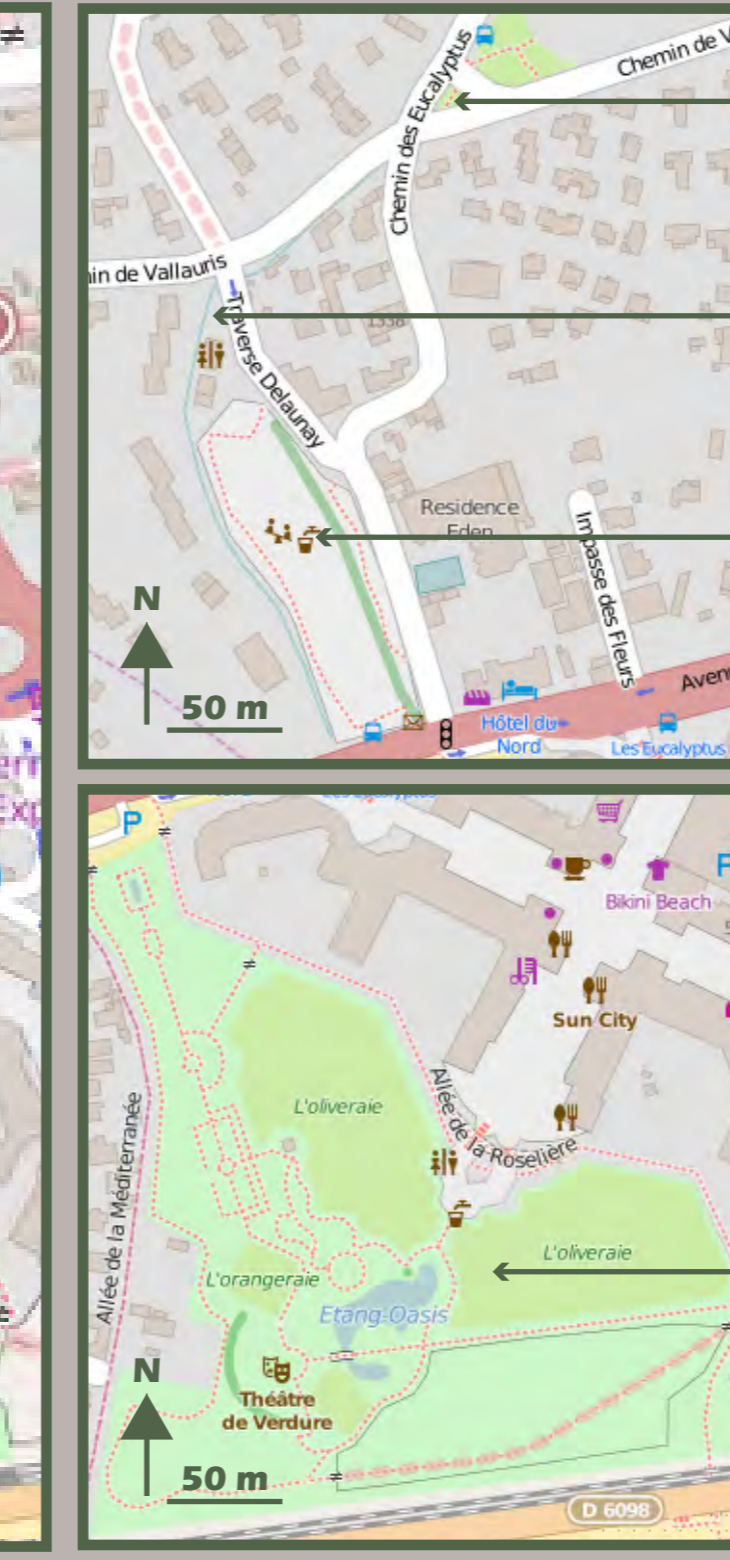
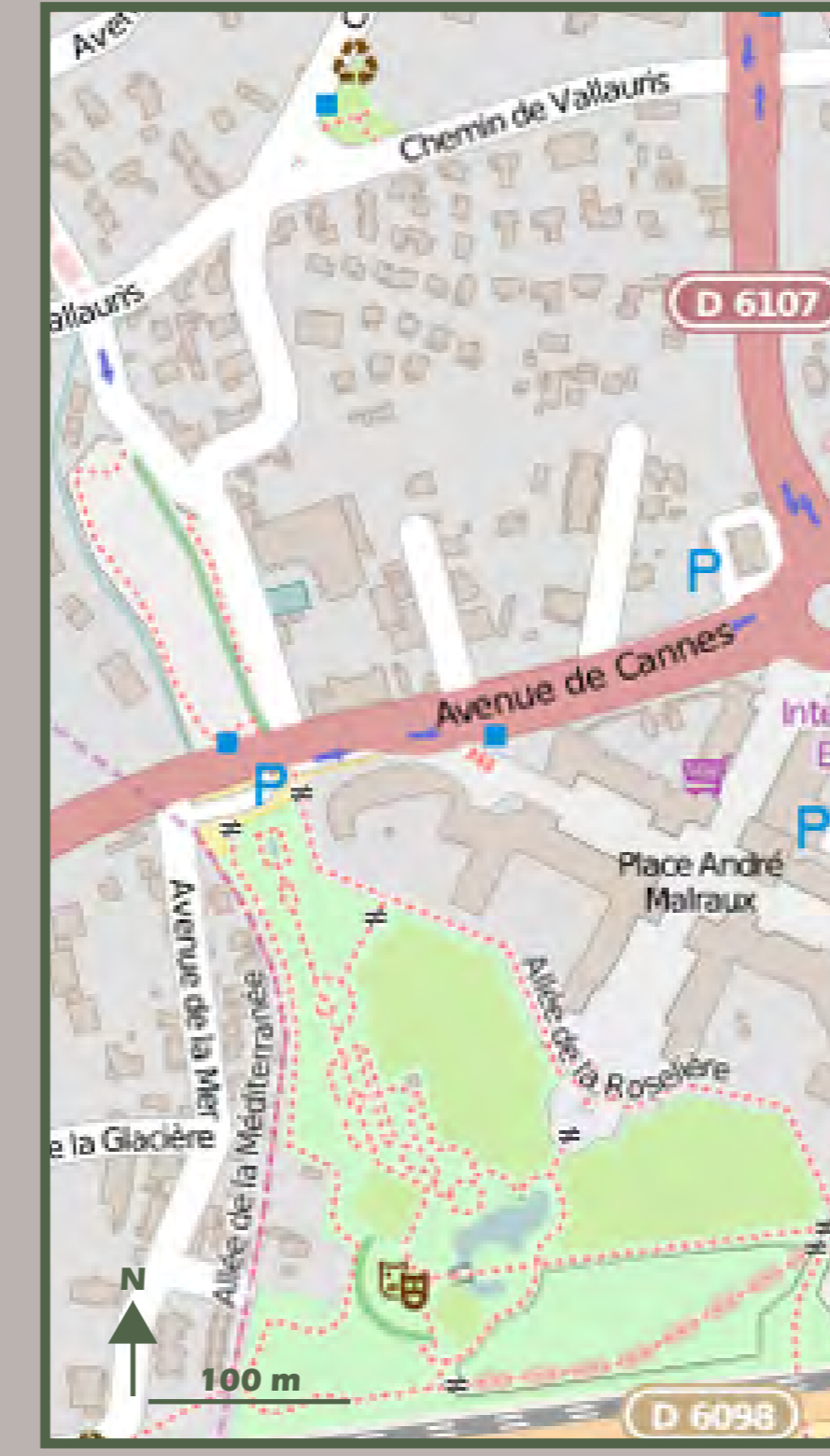
Plan du secteur pour noter les informations Découverte du parc d'Exflora Exploration du square Delaunay Fin du parcours Ambroisie Lauvert et collation

PHASE 2 : Cartographier les informations sur OpenStreetMap

Cartographie OpenStreetMap avant la sortie terrain :



Cartographie OpenStreetMap après la contribution des participants :



- Contour espace vert
- Arrêt de bus
- Sentier piéton
- Canal d'évacuation
- Correction contour
- Toilette
- Jardin d'enfant
- Eau potable
- Contour jardins à thèmes
- Théâtre
- Eau potable

PHASE 3 : Réalisation d'une carte thématique : Végétation et systèmes d'arrosage des espaces verts



Un clic sur un élément pour faire apparaître les informations dans la fenêtre latérale : descriptions, photographies, liens...

Carte web :

Cartographie thématique réalisée grâce à Umap, outil d'OpenStreetMap.

Il s'agit d'un logiciel open source qui permet de créer des cartes personnalisées et interactives sur le web qu'il est possible de partager.

Lien pour visionner la carte sur Umap :

http://umap.openstreetmap.fr/fr/map/vert_37541#16/43.5729/7.0887

III. De l'expérimentation à l'intérêt de la cartographie collaborative :

Acteur du territoire	Observation phase 1	Observation phase 2 et 3	Perception de l'information spatiale produite
Elu	Resté à l'écart soucieux des échanges entre les acteurs		Vecteur de communication et d'argumentation
Gestionnaire	Sensible face à l'approche participative développée	La totalité des informations n'ont pas été entrées dans la plateforme OSM par manque de temps car la durée accordée à l'enseignement des logiciels a été sous-estimée.	Moyen pour expliquer et comprendre des phénomènes en lien avec la gestion durable des espaces verts
Association	Echanges d'informations avec les participants Intérêt porté à la méthodologie	Niveau technologique élevé entraînant des difficultés d'utilisation chez certains participants	Travail produit par un logiciel professionnel lui conférant du crédit et de la fiabilité
Administré	Curieux des échanges sur les expériences et connaissances de chacun		

IV. Conclusion :

Participation = Conception d'ensemble dans le but d'améliorer la qualité des processus de gestion

La cartographie collaborative numérique = Spatialisation des informations par le géoréférencement des éléments du terrain + Outil d'aide à la gestion par la mutualisation des compétences et connaissances des acteurs du territoire

VERT = Tremplin pour développer une stratégie de l'innovation numérique + Tremplin pour former des individus à la pratique de la participation + Tremplin pour valoriser la gestion durable du territoire

V. Perspectives :

Ambition de l'atelier pilote :

- impluser une dynamique participative au sein du service gestionnaire,
- développer la mobilisation des administrés pour ainsi ouvrir à une pratique plus généralisée,
- assurer un accès aux données pour favoriser l'implication des acteurs du territoire aux débats locaux.

FLOOD AR : utiliser des modèles 3D temporels interactifs pour la sensibilisation au risque d'inondation

Premiers retours sur l'utilisation de dispositifs de simulation située auprès de riverains du Rhône

Florence Jacquinod¹, Jérémy Edert², Gilles Gesquière², Marcel Vogt³, Julia Bonaccorsi³, Maryline Cerqueira Da Cunha¹

1. Univ. Lyon, UJM Saint-Etienne, CNRS UMR 5600 Environnement-Ville-Société
6 rue basse des rives, F-42023 St-Etienne, France
jacquinod@gmail.com

2. Univ. Lyon, Université Lumière Lyon 2, CNRS UMR 5205 LIRIS
gilles.gesquiere@univ-lyon2.fr

3. Univ. Lyon, Université Lumière Lyon 2,
14 avenue Berthelot, 69007 Lyon, France
Julia.bonaccorsi@univ-lyon2.fr

RESUME. Le projet FLOOD AR est un projet exploratoire réalisé par un collectif de recherche pluridisciplinaire en collaboration avec des praticiens. Il a pour objectif de développer des dispositifs de sensibilisation au risque d'inondation sous forme de simulation située et d'évaluer leur efficacité en contexte, lors d'expérimentations publiques. Ce projet a conduit au développement de méthodes et d'outils pour produire des modèles 3D de territoire temporels interactifs ainsi que des dispositifs de consultation sur tablettes et smartphones. Des dispositifs ont pu être testés avec des riverains du Rhône en juin 2016, permettant de produire une première analyse. D'autres expérimentations seront réalisées afin d'affiner les modes de production et les choix d'interface et de déterminer les apports possibles des dispositifs de consultation de modèles 3D de territoire temporels interactifs à la sensibilisation au risque du grand public.

ABSTRACT. FLOOD AR is an ongoing exploratory research project led by a pluridisciplinary research group together with practitioners. It aims at developing devices to raise citizens awareness about flood risk and to evaluate their efficiency. To date, the project has led to the development of methodologies and tools to produce 3D temporal interactive models and interfaces to interact with 3D temporal geovisualizations on tablets and smartphones. Devices were experimented with citizens in June 2016 and preliminary results are available. Other

SAGEO'2016 – Nice, 6-9 décembre 2016

SAGEO'2016 – Nice, 6-9 décembre 2016

experiments will be conducted so as to complete an analysis of those situated simulation devices and to determine if and how they can raise citizens awareness about flood risk.

MOTS-CLES : géovisualisation 3D interactive, modèle 3D temporel, CityGML, risque d'inondation, sensibilisation, réalité augmentée, simulation située, réalité virtuelle immersive.

KEYWORDS: 3D interactive geovisualization, 3D temporal models, CityGML, flood risk, public awareness, augmented reality, situated simulation, immersive virtual reality.

1. Introduction

La sensibilisation du grand public au risque d'inondation est un pan important de la politique de prévention du risque d'inondation. Il s'agit de faire en sorte que les personnes pouvant être affectées par une inondation soient conscientes de ce qui peut se produire en cas de crue et comprennent comment elles peuvent s'en protéger. En France, dans le cas du fleuve Rhône, les sondages grand public réalisés ces dernières années montrent le déficit de conscience et de connaissance du risque d'inondation des riverains du fleuve. Dans ce contexte, les services en charge de la prévention du risque inondation ont exprimé le besoin de disposer d'outils de médiation par la visualisation, et, en particulier, de modèles 3D géoréférencés, ayant déjà eu l'occasion de mobiliser des géovisualisations 3D (Jacquinod, 2013).

Sont présentés ici les premiers développements et résultats du projet FLOOD AR, en cours de réalisation. Ce projet a pour objectif de développer des méthodes de production de modèles 3D temporels interactifs répondant aux besoins des praticiens. Le projet a également pour objet l'évaluation, lors d'expérimentations publiques, de ces dispositifs en termes de compréhension de l'information sur le risque. Les questionnements de recherche issus des demandes des praticiens sont d'abord présentés (2), avant de décrire les méthodes et outils développés à ce jour (3), puis de donner les premiers retours sur leur utilisation auprès du public (4).

2. Contexte et verrous scientifiques

Les sondages réalisés par l'institut BVA parmi les habitants des zones inondables le long du fleuve Rhône en 2009 puis 2013¹ ont révélé que seulement 1/5 des habitants savent qu'ils vivent dans une zone à risque et que moins de 10% d'entre eux ont déjà consulté les documents sur le risque produits par l'Etat et les collectivités. Le sondage de 2013 a également mis en évidence que 31% des riverains demandent à ce que les autorités publiques leur fournissent des

¹ Les résultats des sondages sont disponibles sur le site du Plan Rhône, qui les a financés : <http://www.planrhone.fr/Inondations/Thematiques/Savoir-mieux-vivre-avec-le-risque/Developper-la-culture-du-risque/Sondages-sur-la-perception-du-risque-d-inondations-des-populations-riveraines>

informations sur le risque, plaçant cette demande en tête des moyens cités par les sondés pour faciliter la vie des habitants des zones inondables. Devant l'absence de diffusion des documents d'information, les gestionnaires réfléchissent à d'autres modes de communication de l'information. A partir d'études existantes (Jacquinod et Langumier 2010 ; Goepfert *et al.*, 2012 ; CEPRI, 2013) et de plusieurs échanges entre les chercheurs du projet et les praticiens impliqués, des critères pour la création de nouveaux dispositifs numériques de sensibilisation à destination du grand public ont été établis :

- Les informations techniques sur le risque, en particulier les études d'aléas, doivent être représentées sur des modèles 3D de territoire, ces représentations ayant montré, par le passé, leur efficacité pour faire comprendre le risque d'inondation en situation de concertation avec les élus et les habitants ;

- Le phénomène de crue doit être représenté dans l'espace et dans le temps, grâce à un modèle dynamique, afin de donner à voir l'origine spatiale des débordements et la vitesse de propagation de la crue de même que sa durée ;

- Les modèles 3D spatio-temporels ainsi produits doivent pouvoir être consultés de manière interactive, en particulier via des dispositifs mobiles et tactiles et, si possible, sur site sous la forme de réalité augmentée.

D'un point de vue scientifique, le développement de tels dispositifs nécessitent que des recherches soient conduites à la fois dans le champ de l'informatique, de la géomatique et des sciences de l'information et de la communication. En effet, si de nombreux outils et technologies sont disponibles, la production de dispositifs pouvant être mobilisés lors d'événements grands publics nécessite de lever des verrous scientifiques et techniques, afin de permettre, d'une part, l'implémentation d'un modèle de données 3D géoréférencé permettant le stockage d'informations temporelles (Chaturvedi *et al.*, 2016) et, d'autre part, la navigation spatiotemporelle. De plus, la littérature scientifique sur les implications communicationnelles de tels dispositifs est quasi-inexistante dans le champ de la prévention du risque. Il existe des travaux, peu nombreux, sur l'évaluation des effets des géovisualisations 3D dans des contextes opérationnels et collectifs (Jacquinod, 2014), mais pas d'études du rôle joué par ces dispositifs en matière de sensibilisation, ces dispositifs étant plus souvent imaginés pour la gestion de crise ou pour être utilisés par des experts (Lonergan et Hedley, 2015 ; Nishio et Mori, 2015 ; Saher *et al.*, 2015).

3. Méthodes et outils développés dans le cadre du projet

Un collectif pluridisciplinaire a donc été constitué autour de ce projet, composé de chercheurs en aménagement et urbanisme, en géomatique, en informatique et en sciences de l'information et de la communication. La collaboration avec des gestionnaires impliqués dans la politique de prévention du risque d'inondation a été mise en place afin de produire des dispositifs adaptés à la sensibilisation du grand public pour les expérimenter sur des territoires rhodaniens et en réaliser une première évaluation. La production des dispositifs a été réalisée de manière itérative, prenant en compte les besoins exprimés par les praticiens et les retours sur les

SAGEO'2016 – Nice, 6-9 décembre 2016

dispositifs donnés par les praticiens et l'ensemble des chercheurs impliqués. Sur le plan technique et informatique, un premier résultat du projet est la production d'un outil automatisé pour la création de modèles 3D interactifs en CityGML intégrant les données hydrauliques temporelles montrant le déroulement de la crue sur le territoire choisi pour l'expérimentation. De plus, une première expérimentation publique a été réalisée pour évaluer deux des modes d'interaction possible avec ces visuels 3D sur site, sur le mode de la simulation située.

Le format CityGML est le seul format standard reconnu par l'OGC qui permette de stocker des données 3D temporelles. Il est néanmoins difficile de mobiliser ce format avec les outils SIG du marché. Une part importante des développements du projet FLOOD AR a donc porté sur la création d'un plugin pour un outil 3D Open Source. Ce plugin permet de créer de manière entièrement automatisée une base de données 3D temporelle en CityGML comprenant les données de base décrivant le territoire² ainsi que des données hydrauliques temporelles. Il permet également de consulter interactivement les données temporelles grâce à un curseur temporel.

4. Première expérimentation et premiers résultats

Un travail méthodologique sur l'export de visuels 3D consultables sur tablettes et smartphones avec casque de réalité virtuelle est en cours. Deux dispositifs ont été produits pour la première expérimentation publique. Deux types d'interfaces donnant accès aux visuels 3D du déroulement temporel de la crue ont été testés sur une journée, auprès de riverains du Rhône³. Le premier était dédié à la consultation sur site du déroulement spatiotemporel de la crue sur l'ensemble de la zone d'expérimentation. Ce dispositif était présenté par un médiateur expliquant le fonctionnement hydraulique de la zone, pendant que les participants pouvaient explorer avec une tablette les géovisualisations 3D montrant l'inondation de la commune depuis le même point de vue que celui où ils se tenaient (point de vue sur un coteau donnant sur la commune). Le second dispositif était destiné à la visualisation de l'évolution de l'inondation et de la hauteur d'eau sur la place centrale de la commune, où était installé le stand de sensibilisation. Ainsi, les personnes qui passaient sur la place pouvaient visualiser l'évolution de la crue sur la place à l'aide d'un smartphone et d'un casque de réalité virtuelle leur permettant de consulter un modèle 3D de la place avec une représentation volumique de l'évolution des hauteurs d'eau lors de l'inondation (cf. annexe photographique).

Les premières analyses des expérimentations, menées sous forme d'observation directe et de questionnaire auprès des participants donnent des premières indications sur l'efficacité des dispositifs. Nous entendons par *efficacité* le fait que les usagers portent leur attention sur le contenu (représentation du territoire et repérage, données mobilisées...) et la thématique de la crue (lien avec savoirs antérieurs,

² Modèle numérique de terrain, orthophotographie géoréférencée, bâtiments 2,5D - par exemple issues de la BD TOPO de l'IGN, ouvrages hydrauliques 3D géoréférencés.

³ 23 personnes ont testées le dispositif sur tablette et environ 50 les smartphones avec casque de réalité virtuelle.

reformulation...) et non pas uniquement sur les dimensions technique et logicielle du dispositif. Les résultats sont de deux ordres principaux :

- d'abord, en matière de prise en main et de manipulation des dispositifs, les smartphones associés au casque de réalité virtuelle se sont révélés plus aisés à utiliser⁴ que les tablettes nécessitant la maîtrise de l'interface tactile, et ce, alors que l'hypothèse avait été formulée que l'immersion ne serait probablement pas adaptée à tous les publics⁵. Il semble que le succès du dispositif smartphone soit aussi lié à la connaissance préalable du territoire, relativisant le pouvoir que l'on pourrait intrinsèquement attribuer au dispositif de réalité immersive en soi. Néanmoins, la sensibilisation est étayée par le dispositif 3D en tant qu'il est engagé dans un scénario de médiation plus large et le programme d'activité des usagers, par exemple, en donnant une fonction au corps impliqué dans l'activité de consultation des images immersives, ou encore en provoquant la recherche d'informations dans les panoramiques.

- ensuite, l'analyse des interprétations des représentations par les testeurs a produit des enseignements sur la « fidélité » des modélisations du territoire et son opérativité symbolique dans l'appropriation cognitive des visualisations. La recherche de lieux connus et familiers dans les visualisations engage une activité de discrimination des informations qui peut être accompagnée sémiotiquement afin d'explicitier les options prises et les différents statuts et natures des informations visualisées.

5. Conclusion (provisoire) et perspectives

Le projet FLOOD AR est en cours et des expérimentations sont en préparation pour affiner les dispositifs et pour compléter l'évaluation des modèles 3D temporels et des dispositifs de consultation interactive sous forme de simulation située. Les développements réalisés permettent néanmoins déjà d'envisager la production de modèles de territoire 3D temporels et interactifs sur l'ensemble des zones inondables du Rhône, grâce à l'outil Open Source développé. Cet outil pourrait d'ailleurs être utilisé pour produire un modèle 3D simple de n'importe quel territoire à partir des données BD TOPO de l'IGN ou de données géographiques semblables. En matière d'analyse des dispositifs de simulation située reposant sur des visuels 3D du déroulement de la crue, les évaluations des interfaces tablettes et smartphones restent à confirmer grâce aux expérimentations à venir. Ces évaluations permettront de guider la réutilisation des méthodes et outils développés sur d'autres territoires. Ces premiers résultats d'évaluation pourront également nourrir des réflexions sur les usages et les apports de l'information géographique, notamment les recherches sur son apport à la sensibilisation des populations au risque d'inondation, domaine où

⁴ Une seule personne sur les cinquante n'a pas pu utiliser le casque à cause d'une forte myopie et a consulté les visuels de la place sur une tablette tactile avec l'aide d'un médiateur.

⁵ Notamment en fonction de l'âge et les risque de vertige lié aux retours des premiers bêta-testeurs.

SAGEO'2016 – Nice, 6-9 décembre 2016

les technologies géonumériques 3D restent encore peu mobilisées, malgré les demandes exprimées par les praticiens.

Remerciements

Ce projet est co-financé par l'Union européenne dans le cadre du Programme Opérationnel Interrégional FEDER Rhône-Saône 2014-2020 et dans le cadre du volet « Inondations » du Plan Rhône. Il est également financé par la Compagnie nationale du Rhône et le Programme d'avenir Lyon –Saint-Etienne dans le cadre du dispositif Images et perceptions embarquées (PALSE-IPEm).

Bibliographie

- Chaturvedi K., Smyth C. S., Gesquière G., Kutzner T., Kolbe T. (2016). Managing versions and history within semantic 3D city models for the next generation of CityGML, *Advances in 3D Geoinformation, Lecture Notes in Cartography and Geoinformation*. Springer.
- Centre européen de prévention du risque d'inondation (CEPRI). (2013). Sensibiliser les populations exposées au risque d'inondation. Les guides du CEPRI, avril 2013, 60 p.
- Goepfert E-M., Martinais E., Le Naour G. (2012). *L'information sur les risques industriels. Quelles attentes ? Quels besoins ?* Rapport d'étude, 58 p., <http://www.pprtrhonealpes.com/medias/telecharger/2221/02221.pdf>
- Jacquinod F., Langumier J. (2010). Géovisualisation 3D en action dans l'aménagement du territoire. Stratégies et usages de l'outil à l'occasion de l'étude d'un Plan de Prévention des Risques Inondations. *Revue Géocarrefour*, vol. 85, n° 4, p. 303-312, <http://geocarrefour.revues.org/8163>
- Jacquinod F. (2013). *Géovisualisations 3D pour l'analyse et l'aide à la décision - Etude menée au sein de l'Université de Saint-Etienne*. Rapport de l'étude menée entre 2011 et 2013 par le laboratoire UMR EVS-ISTHME, subventionnée par l'Union Européenne et le volet « Inondations » du Plan Rhône, octobre 2013, 31 p.
- Jacquinod F. (2014). *Production, pratique et usages des géovisualisations 3D dans l'aménagement du territoire*. Thèse en géographie et aménagement, Université de Saint-Etienne.
- Lonergan, C., Hedley, N. (2015). Navigating the future of tsunami risk communication: using dimensionality, interactivity and situatedness to interface with society. *Natural Hazards*, vol. 78, n°1, p. 179–201.
- Nishio, M. and Mori, M., 2015. Hydrologic analysis of a flood based on a new Digital Elevation Model. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-7/W4, 127–134.
- Saher, F.N., Ali, N.B.M., Bin Wan Ishak, W.M.F., and Bin Bidin, M.A.N., 2015. Development of an Intelligent Decision Support System for Flood Mitigation in the Pahang River. *International Journal of Sustainable Land Use and Urban Planning (IJSLUP)*, 2 (2).

Annexe photographique

Ci-dessous : photos de l'expérimentations du 12 juin 2016 (tablettes, consultation de puis un point haut donnant une vue générale des zones inondables) – Utilisations de la tablette et exemple d'une des géovisualisations 3D consultables via la tablette

© Photos E. Robert



SAGEO'2016 – Nice, 6-9 décembre 2016

Ci-dessous : photos de l'expérimentation du 12 juin 2016 (smartphones et casque de réalité virtuelle, utilisé sur la place centrale de la commune) – Utilisations et exemples de géovisualisations 3D consultées via les smartphones

© Photos E. Robert



Génération d'itinéraires pour le Vélo en Libre Service : Vers la prise en compte de la multi-modalité

Ines Jguirim¹, Anny Vivas Abril², David Brosset¹,
Thierry Dubois², Céline Mayousse²

1. Institut de Recherche de l'Ecole Navale

BCRM Brest

CC 600

29240 BREST Cedex 9

ines.jguirim@ecole-navale.fr, david.brosset@ecole-navale.fr

2. Altran France – Région Ouest

ZAC de La Courrouze – Bâtiment Eolios

3 rue Louis Braille

35136 Saint Jacques De La Lande

anny.vivasabril@altran.com, thierry.dubois@altran.com, celine.mayousse@altran.com

RÉSUMÉ. La génération d'itinéraires d'activités urbaines avec prise en compte de la multi-modalité des transports reste encore un problème complexe. Nous présentons dans cet article les premières étapes d'une méthode de génération d'itinéraires. Cette méthode repose sur un modèle de transport multi-modal utilisant le concept de super réseau et l'utilisation d'un algorithme génétique pour l'optimisation multicritères des itinéraires.

ABSTRACT. Route planning including urban activities with multi-modality transport is still a complex problem. We present in this article the first steps of a route planning method. This method is based on a multi-modal transport model using the concept of "super-network", and on the use of a genetic algorithm for multi-criteria route planning.

MOTS-CLÉS : Transport multi-modal, algorithme génétique, itinéraire, super réseau

KEYWORDS: Multi-modal transport, genetic algorithm, routing, super-network

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

La recommandation d'itinéraires urbains en fonction de l'activité choisie par l'utilisateur demande de considérer l'ensemble des modes de transport possibles. Ces modes de transport peuvent être combinés au sein d'un même trajet. L'augmentation des lignes de transport en commun (tramway, métro) incite les citoyens à utiliser des parkings relais et à mélanger les différents modes : transport en commun, véhicule personnel et piéton. L'émergence de nouveaux modes de transport comme le Vélo en Libre Service (VLS), et bientôt le VLS à assistance électrique, tend à modifier la mobilité dans la ville.

L'objectif de ce papier est de présenter deux étapes de la génération d'itinéraires multi-modaux d'activités urbaines : la modélisation de la multi-modalité ainsi qu'une optimisation d'itinéraires pour le VLS à assistance électrique. La section 2 présente une modélisation de la ville par des super réseaux afin d'intégrer la multi-modalité et également une méthode pour obtenir des points d'intérêts pertinents. La section 3 décrit la génération d'itinéraires à base d'algorithme génétique en ne considérant que le nouveau mode de transport VLS à assistance électrique, en optimisant plusieurs critères liés à son usage.

2. Modélisation de la ville

Afin de pouvoir recommander des itinéraires, la ville doit être modélisée en termes de réseaux de transport et de points d'intérêt disponibles. De nombreux travaux ont modélisé les transports urbains dans un but de simulation ou d'analyse de la mobilité. Les graphes sont les outils privilégiés pour cette modélisation (Jiang, Claramunt, 2004; Jguirim *et al.*, 2015; Sauvanet, 2011). Néanmoins, la structure classique des graphes ne permet pas de considérer le changement de type de transport.

La première étape est de prendre en considération la possibilité de changer de mode de transport. L'approche choisie est celle des super-réseaux (Sheffi, 1985). Les super-réseaux, ou «super-networks», sont classifiés comme des graphes complexes permettant d'interconnecter plusieurs couches (Buldyrev *et al.*, 2010). Les couches représentant les graphes sont connectées à travers des nœuds particuliers de chaque graphe, permettant ainsi de connecter les nœuds de l'ensemble des graphes. Un super-réseau est formalisé par M couches qui représentent les différents modes de transport (Bianconi, Dorogovtsev, 2014). La figure 1 représente le chemin d'un super réseau avec 3 couches représentant 3 réseaux de transport différents : les pistes cyclables, le réseau piéton et les lignes de tramway. Pour appliquer cette modélisation à un cas d'étude précis, nous utilisons les couches de transport fournies par le projet OpenStreetMap (<http://openstreetmap.fr/>). Ce projet communautaire de type VGI (Volunteered Geographical Information) couvre de façon satisfaisante les réseaux de transport en France mais le nombre de points d'intérêt (POI) est souvent insuffisant.

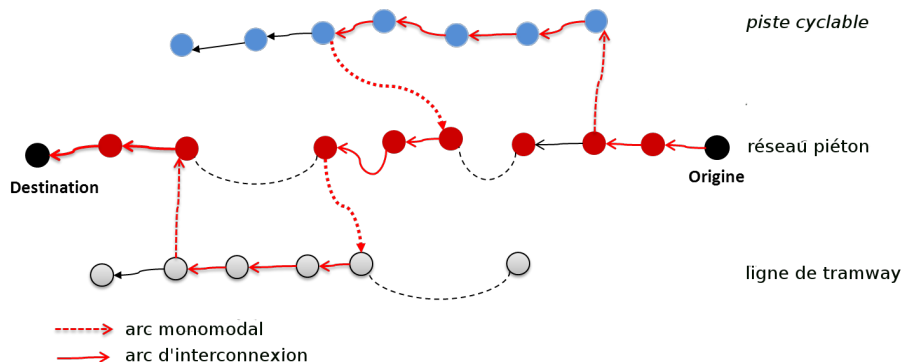


FIGURE 1. Réseau multimodal

Les POI sont des lieux offrant un service, et permettent d'intégrer les activités dans les itinéraires. Ils sont définis principalement par leur position géographique et un type (restaurant, station service, gare ...). Ce type dépend du fournisseur utilisé car plusieurs catégorisations plus ou moins détaillées existent. Nous prévoyons d'utiliser Factual (www.factual.com) pour la quantité de POI disponibles sur la France, près de 5 millions en 2016, et la mise à jour régulière des informations. L'accès aux données se fait par le biais d'une API (Application Programming Interface) utilisable avec plusieurs langages de programmation et permettant de réaliser des requêtes géographiques dans la base de données de POI.

3. Génération d'itinéraires

Dans le cadre du projet Connect conduit par Altran Research (Mayousse, Gilbert, 2016), l'algorithme génétique développé est principalement destiné aux Vélos en Libre Service (VLS) à assistance électrique permettant d'intégrer des POI, comme points de passage d'un itinéraire touristique. Il permet à l'utilisateur de calculer un itinéraire prenant en compte les critères suivants :

- La distance la plus courte
- La sécurité du cycliste en prenant en compte la configuration de la voirie
- la charge de la batterie afin de garantir à l'utilisateur une assistance électrique durant l'ensemble du parcours

L'évaluation des itinéraires générés est réalisée par la méthode de classement de Pareto (Deb *et al.*, 2002). Les premiers résultats obtenus sur la ville de Rennes montrent que l'algorithme est satisfaisant. En effet, les itinéraires produits sont calculés en moins de 3 secondes avec un score équivalent à celui obtenu par méthodes exactes, prenant plusieurs minutes d'exécution.

SAGEO'2016 – Nice, 6-9 décembre 2016

Le passage à l'échelle avec des itinéraires plus complexes sera néanmoins une étape difficile qui nécessitera de nouvelles optimisations de l'algorithme. L'utilisation du réseau multi-modal nous permettra d'ajouter de nouvelles contraintes à optimiser, notamment la diminution du temps d'attente lors de l'utilisation des transports publics.

4. Conclusion

Générer des itinéraires touristiques en considérant différents modes de transport dont le VLS à assistance électrique est un problème complexe. Dans ce papier, nous avons présenté deux étapes pour arriver à cette génération : la modélisation de la ville avec ces différents modes de transport et leurs interconnexions ainsi qu'un algorithme de calcul d'itinéraire multi-critères adapté au VLS à assistance électrique.

Ces deux étapes à présent réalisées, nous allons nous concentrer sur l'application de la méthode globale sur la ville de Rennes pour des itinéraires de type touristique avec la visite de plusieurs POI de la catégorie tourisme sur une période de temps d'une journée.

La définition formelle d'activités (durée, nombre et type de POI, enchaînement privilégié des modalités de transport ...) reste également une étape importante pour obtenir une application complète de génération d'itinéraires multimodaux d'activités urbaines.

Bibliographie

- Bianconi G., Dorogovtsev S. N. (2014). Multiple percolation transitions in a configuration model of a network of networks. *Physical Review E*, vol. 89, n° 6, p. 062814.
- Buldyrev S. V., Parshani R., Paul G., Stanley H. E., Havlin S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, vol. 464, n° 7291, p. 1025–1028.
- Deb K., Pratap A., Agarwal S., Meyarivan T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, vol. 6, n° 2, p. 182–197.
- Jguirim I., Brosset D., Claramunt C. (2015). Un système de génération d'itinéraires des activités mobiles dans la ville. In *International Conference on Spatial Analysis and GEOmatic (SAGEO)*.
- Jiang B., Claramunt C. (2004). A structural approach to the model generalization of an urban street network. *GeoInformatica*, vol. 8, n° 2, p. 157–171.
- Mayousse C., Gilbert E. (2016). The new generation of smart power station for electric vehicles. In *Les rencontres de la mobilité intelligente*.
- Sauvanet G. (2011). *Recherche de chemins multiobjectifs pour la conception et la réalisation d'une centrale de mobilité destinées aux cyclistes*. Thèse de doctorat non publiée.

Sheffi Y. (1985). *Urban transportation networks: Equilibrium analysis with mathematical programming methods*. Prentice-Hall, Englewood Cliffs, NJ.

Identifier des espaces préservés de la pollution de l'air sur le chemin de l'école

Julie Prud'homme^{1,2,3}, Malika Madelin², Sophie Baudet-Michel³

1. *LabexDynamite*

21 rue Broca, 75005 Paris, France
jul.prudhomme@gmail.com

2. *CNRS UMR 8586 PRODIG, Sorbonne Paris Cité, Université Paris Diderot, UFR GHSS (cc 7001)*

5 rue Thomas Mann, 75205 Paris Cedex 13, France
malika.madelin@univ-paris-diderot.fr

3. *CNRS UMR 8504 Géographie-Cités, Université Panthéon-Sorbonne Paris 1*

13 rue du Four, 75006 Paris, France
sophie.baudet-michel@univ-paris-diderot.fr

RESUME. Nous proposons de présenter les premiers résultats d'un protocole de mesure mis en place dans le 14^{ème} arrondissement de Paris pour estimer l'exposition des enfants à la pollution de l'air (PM10) sur le chemin de l'école. L'objectif est de construire une hiérarchisation des espaces autour des écoles en termes de pollution de l'air pour in fine être en capacité d'identifier des espaces moins exposés à la pollution de l'air.

ABSTRACT. We propose to present first results of a measurement protocol developed in the 14th district of Paris to estimate personal exposure of children on their way to school to air pollution (PM10). The aim is to establish a ranking of spaces around schools in air pollution terms.

MOTS-CLES : Pollution de l'air, mesures, cartographie, hiérarchisation des espaces, PM10

KEYWORDS: Air pollution, measurement, cartography, ranking of spaces, PM10

1. Introduction

À Paris, la pollution de l'air est une problématique environnementale récurrente. La pollution aux microparticules (PM – *Particulate Matter*) fait périodiquement la une des journaux, incriminant les véhicules diesel en circulation notamment. Les impacts de leur forte concentration sur la santé respiratoire de la population sont désormais bien connus (Host, 2013 ; Annesi-Maesano *et al.*, 2004 ; Anderson *et al.*, 2001).

Pour estimer l'exposition des populations à la pollution de l'air (dont les PM10), on distingue deux approches majeures : des estimations basées sur les niveaux de polluants sur le territoire et des estimations basées sur l'intensité et la distance au trafic routier (Host, 2013). En effet, depuis la fin des années 1990, une méthode basée sur des modèles de régression d'utilisation des sols (*land-use regression models*) a été développée pour estimer les concentrations de polluants sur les territoires. Ils sont classiquement basés sur des SIG, utilisant, ou non, des méthodes géostatistiques d'interpolation spatiale. Ils sont articulés avec des modèles de chimie-transport des polluants dans l'atmosphère, de trafic, parfois enrichis avec des mesures de stations fixes (Merbitzet *et al.*, 2012 ; Elbiret *et al.*, 2010 ; Jerrett *et al.*, 2005 ; Puliafito *et al.*, 2003 ; Lin & Lin, 2002 ; Jensen *et al.*, 2001). Les chaînes de modélisation mises en œuvre dans les méthodes d'estimation ont pour objectif de proposer des estimations de l'exposition de la population à différents polluants, à une échelle fine. Avec les évolutions techniques, les travaux en expologie tendent à mesurer les expositions individuelles, en couplant les mesures de pollution à un GPS (Chaix *et al.*, 2013) pour mettre en relation environnement et concentrations de polluants. Les SIG sont ainsi largement utilisés pour l'exploitation de ces données. Une limite forte de ces différentes méthodologies est le coût financier et temporel des campagnes de mesures, idéalement individuelles à effectuer pour préciser et valider les modèles (Hoek *et al.*, 2008).

L'exposition à la pollution de proximité (dont les PM10) des enfants lors du trajet entre le domicile et l'école est très peu documentée. Pourtant, les enfants se rendent à l'école à l'heure où le trafic est le plus important sur les routes (heure de pointe, surtout le matin) et où les niveaux de concentration sont les plus élevés de la journée, d'après les mesures des stations de pollution trafic routier du réseau AirParif (dans le cas de Paris).

2. Protocole de mesure mis en œuvre

Nous avons choisi d'approfondir la question de l'exposition des enfants sur le chemin de l'école dans deux secteurs scolaires du 14^{ème} arrondissement de Paris, à proximité d'une station de mesure de PM10 « trafic » d'AirParif. Le projet avait pour objectif d'élaborer un protocole reproductible s'appuyant sur une campagne de mesure permettant de s'approcher au plus près de la réalité de l'exposition des enfants.

SAGEO'2016 – Nice, 6-9 décembre 2016

Nous proposons un protocole de mesure le long d'un circuit de 3,34 km qui parcourt des rues de chaque secteur scolaire, dans des types de rues variés. Il constitue un échantillon de rues potentiellement parcourues par des enfants pour se rendre à l'école. Pour effectuer les mesures, on utilise un analyseur portable homologué de particules (*Dustmate* de TurnKey Instrument Ltd.) qui fournit les concentrations de PM10, PM2.5 et PM1 de manière horodatée. On utilise une application de tracking sur Android (« Mes parcours » v2.0.9 ©2016 Google) pour faire les relevés GPS en continu de chaque circuit (un peu moins d'une heure) et on met en correspondance l'heure de chacune des données pour cartographier les mesures. Nous effectuons une mesure toutes les 5 secondes, un compromis entre les contraintes matérielles (mémoire de l'analyseur), durée du circuit de mesure établi, et la précision spatiale des mesures souhaitée. Les analyses qui suivent sont basées sur hauteurs de mesures, soit huit matinées. La Table 1 présente les caractéristiques statistiques de chacune.

TABLE 1. Description statistique des huit séries de mesures de PM10

	Min.	25%	Méd.	Moy.	75%	Max.	Coef.Var.
08/04/2016	83,4	142,7	157,5	163,3	176,4	562,2	0,24
11/04/2016	7,1	27,78	35,5	37,04	43,8	104,4	0,36
14/04/2016	9,3	28,42	36,55	39,84	48,3	145,1	0,43
05/09/2016	31,4	63,22	73,85	75,99	86,48	262,8	0,27
08/09/2016	19,5	39,7	49,9	57,57	62,2	135,8	1,45
13/09/2016	11,1	44,08	55,2	59,83	68,78	457,7	0,51
14/09/2016	4,2	23,63	33,25	37,9	46,15	177,1	0,57
20/09/2016	27,4	45,92	54,95	58,41	66,5	241,2	0,35

3. Hiérarchie des espaces selon les mesures de pollution : analyses et résultats

L'hypothèse développée ici porte sur l'existence d'une hiérarchie des espaces selon les niveaux de concentration mesurés entre 8h et 9h et ainsi identifier des espaces plus préservés de la pollution de proximité aux PM10. Il s'agit donc d'identifier des récurrences spatiales dans la répartition des niveaux relatifs de pollution. L'objectif de l'analyse présentée est de regrouper des segments du circuit qui se ressemblent du point de vue des niveaux de pollution mesurés. Pour cela, on s'appuie sur un découpage du circuit de mesure en segments de 40m (82 objets). Pour chaque série de mesures, on identifie le décile auquel chaque mesure appartient. On peut ainsi ordonner les valeurs mesurées et les comparer, toutes choses égales par ailleurs quant aux concentrations enregistrées chaque jour. Celles-ci peuvent en effet être très variées (voir Table 1 : les valeurs minimales de chaque série mesurées vont de 4.2 $\mu\text{g}/\text{m}^3$ à 83.4 $\mu\text{g}/\text{m}^3$).

On effectue une Analyse Factorielle des Correspondances (AFC - sur le χ^2), et une Classification Ascendante Hiérarchique de Ward (CAH - sur la distance euclidienne), à partir du tableau de contingence des segments et des déciles dans lesquels les mesures se positionnent. On tente comme cela à regrouper statistiquement les segments qui se ressemblent selon le positionnement statistique des mesures qu'ils rassemblent. Il en ressort cinq classes significatives, regroupant chacune les segments sur lesquels les mesures se situent principalement dans certains déciles : (1) Segments les moins pollués ; (2) Segments moyennement pollués ; (3) Segments avec des niveaux de pollution très variables, avec une prédominance statistique haute ; (4) Segments avec des valeurs moyennes-hautes et très élevées ; (5) Segments les plus pollués.

Une hiérarchie en termes de niveau relatif de pollution ressort ainsi de cette classification (Figure 2.a.). La cartographie des segments selon la classe à laquelle ils appartiennent (figure 2.b.) met en évidence des espaces plus souvent plus pollués que les autres (rouge foncé), et inversement (rouge clair). On note que les abords des écoles (en bleu sur la carte) ne sont pas forcément parmi les moins pollués et que certains grands axes ne figurent pas parmi les plus souvent les plus pollués (portion nord de l'Avenue du Général Leclerc et Avenue du Maine), comme on aurait pu l'attendre du fait du trafic routier important sur ces axes.

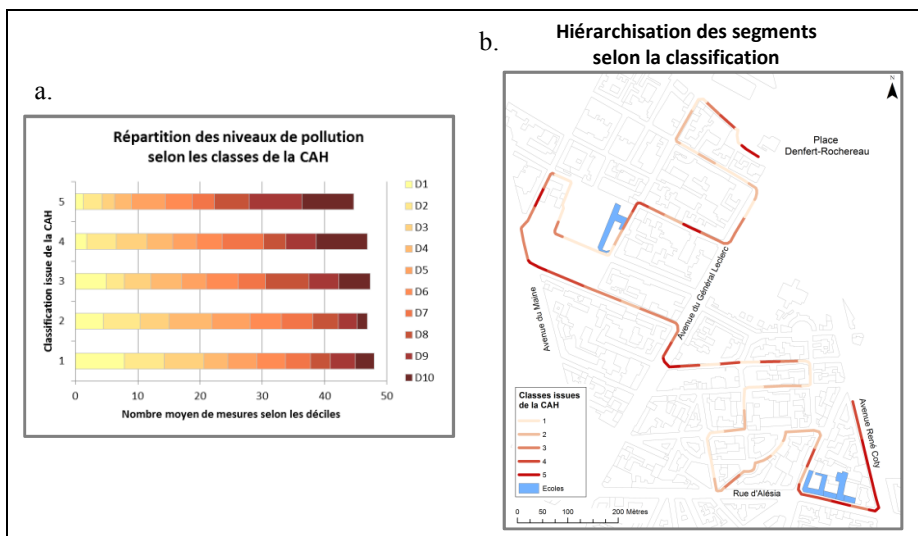


FIGURE 1. Présentation des résultats de la classification : (a) Répartition moyenne des niveaux de pollution selon les classes ; (b) Cartographie des segments selon la classification

Cette classification est le résultat d'une première mise en œuvre du protocole de mesure et d'analyse, sur un créneau horaire et un échantillon spatial restreint. Il pourra être reproduit avec une couverture plus importante des rues autour des écoles et sur d'autres créneaux horaires. Il pourra également être élargi à d'autres polluants, tel que le NO_2 , un autre polluant de proximité qui a des effets sanitaires importants.

SAGEO'2016 – Nice, 6-9 décembre 2016

Remerciements

Ce travail a bénéficié du soutien financier du LabExDynameTe (ANR-11-LABX-0046) dans le cadre du programme « Investissements d'Avenir ».

Références

- Anderson, H. R., Bremner, S. A., Atkinson, R. W., Harrison, R. M., & Walters, S. (2001). *Particulate matter and daily mortality and hospital admissions in the west midlands conurbation of the United Kingdom: associations with fine and coarse particles, black smoke and sulphate*. *Occupational and Environmental Medicine*, 58(8), 504-510.
- Annesi-Maesano, I., Ackermann, U., Céline Boudet, Filleul, L., Medina, S., Slama, R., Viegi, G. (2004). *Effets des particules atmosphériques sur la santé* Revue des études épidémiologiques. *Environnement, Risques & Santé*, 3(2), 97-110.
- Chaix, B., Méline, J., Duncan, S., Merrien, C., Karusisi, N., Perchoux, C., ... Kestens, Y. (2013). GPS tracking in neighborhood and health studies: A step forward for environmental exposure assessment, a step backward for causal inference? *Health & Place*, 21, 46-51. <https://doi.org/10.1016/j.healthplace.2013.01.003>
- Elbir, T., Mangir, N., Kara, M., Simsir, S., Eren, T., & Ozdemir, S. (2010). Development of a GIS-based decision support system for urban air quality management in the city of Istanbul. *Atmospheric Environment*, 44(4), 441-454. <https://doi.org/10.1016/j.atmosenv.2009.11.008>
- Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42(33), 7561-7578. <https://doi.org/10.1016/j.atmosenv.2008.05.057>
- Host, S. (2013). *Exposition à la pollution atmosphérique liée au trafic routier et risques sanitaires*. Vertigo - la revue électronique en sciences de l'environnement, (Hors-série 15).
- Jensen, S. S., Berkowicz, R., Sten Hansen, H., & Hertel, O. (2001). A Danish decision-support GIS tool for management of urban air quality and human exposures. *Transportation Research Part D: Transport and Environment*, 6(4), 229-241. [https://doi.org/10.1016/S1361-9209\(00\)00026-2](https://doi.org/10.1016/S1361-9209(00)00026-2)
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuvaroglu, T., ... Giovis, C. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Science and Environmental Epidemiology*, 15(2), 185-204. <https://doi.org/10.1038/sj.jea.7500388>
- Lin, M.-D., & Lin, Y.-C. (2002). The application of GIS to air quality analysis in Taichung City, Taiwan, ROC. *Environmental Modelling & Software*, 17(1), 11-19. [https://doi.org/10.1016/S1364-8152\(01\)00048-2](https://doi.org/10.1016/S1364-8152(01)00048-2)
- Merbitz, H., Buttstädt, M., Michael, S., Dott, W., & Schneider, C. (2012). GIS-based identification of spatial variables enhancing heat and poor air quality in urban areas. *Applied Geography*, 33, 94-106. <https://doi.org/10.1016/j.apgeog.2011.06.008>
- Puliafito, E., Guevara, M., & Puliafito, C. (2003). Characterization of urban air quality using GIS as a management system. *Environmental Pollution*, 122(1), 105-117. [https://doi.org/10.1016/S0269-7491\(02\)00278-6](https://doi.org/10.1016/S0269-7491(02)00278-6)

Mesures participatives de l'exposition individuelle à la pollution de l'air

Madelin M.¹, Duché S.², Dupuis V.³

1. CNRS UMR 8586 PRODIG, Sorbonne Paris cité, Université Paris Diderot UFR GHSS (cc 7001), 5 rue Thomas Mann, 75205 Paris Cedex 13, France
malika.madelin@univ-paris-diderot.fr
2. CNRS UMR 5194 PACTE, Université Grenoble Alpes, IGA
14bis, avenue Marie Reynoard, 38100 Grenoble, France
sarah.duche@univ-grenoble-alpes.fr
3. CNRS UMR 8234 PHENIX, Sorbonne Universités, UPMC Université Paris 6
4 place Jussieu, 75252 Paris Cedex 05, France
vincent.dupuis@upmc.fr

RESUME. Cette communication présente un projet de recherche participative « Quel est ton air ? », qui sensibilise les citoyens à la pollution de l'air, à leur exposition individuelle, depuis la fabrication d'un capteur jusqu'à la cartographie des concentrations. Au-delà de l'intérêt sociétal et de l'approche participative, ce projet permet également d'observer la variabilité des concentrations à une échelle spatio-temporelle fine, de soulever les enjeux de la représentation des données (mesures et indices, cartographie) et de questionner l'interopérabilité des données individuelles à des mesures de pollution, plus classiques, officielles, à une échelle régionale.

ABSTRACT. This presentation reports on a participatory research project named 'Quel est ton air' that aims at empowering citizens regarding air pollution and individual exposure from the assembly of a sensor to the mapping of observed pollutant concentrations. Beyond the social interest and the participatory approach, this project also allows observing the variation in pollutant concentrations at a fine spatiotemporal scale and raises the stake of data mapping (measurements, indexes, mapping) and questions the interoperability of individual data towards more standard pollution measurements at a regional scale.

MOTS-CLES : pollution de l'air, capteurs à faible coût, sciences participatives

KEYWORDS: air pollution, low cost sensors, participatory sciences

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

La pollution atmosphérique est l'une des problématiques environnementales majeures en milieu urbain, à la source de problèmes sanitaires et de préoccupations citoyennes (Ifop, 2014). L'essor des sciences participatives montre un véritable intérêt des citoyens pour les questions environnementales (Dickinson *et al.*, 2012 ; Goodchild et Li, 2012). En parallèle, le développement de l'électronique pour le grand public et des capteurs à bas prix, miniaturisés et portatifs, rend la mesure de pollution plus accessible à tout un chacun (Budde *et al.*, 2014 ; Duché et Madelin, 2015). Il ouvre de nouvelles perspectives de recherche sur la variabilité des polluants à une échelle spatio-temporelle fine et sa représentation, mais aussi sur la sensibilisation des citoyens à leur exposition individuelle à la pollution de l'air.

Dans cette communication, nous présentons un projet de recherche participative « Quel est ton air ? », retenu sur l'appel à projet *Culture Citoyenne et Scientifique* de la région Île-de-France 2015 (avec cofinancement UMR 8586 PRODIG et CIST Collège International des Sciences du Territoire) et qui repose sur la collaboration entre des géographes-climatologues et le FabLab de l'UPMC (Paris 6). Ce projet vise à sensibiliser le public à la thématique de la qualité de l'air et aux techniques de mesure et de cartographie, *via* des ateliers de fabrication de capteurs, des mesures *in situ* et une cartographie en ligne des données participatives. Dans cette communication, nous présenterons les choix pour la construction de la station portative, pour l'alimentation de la base de données géographiques et les questions autour de la représentation cartographique sur support web.

2. Conception et élaboration d'une station portative, à bas coût, de la qualité de l'air

Nous avons choisi, dans un premier temps, d'étudier un seul polluant, les particules en suspension. Après une étude des travaux et du marché existant (takingspace.org/ ; Budde *et al.*, 2014...), trois capteurs ont été testés (Duché et Madelin, 2015). Ils mesurent des particules fines d'une taille aérodynamique comprise entre 1 et 10 µm. Les tests en conditions forcées et calmes et en conditions réelles, à l'extérieur, ont montré des résultats parfois contrastés (bruit, différences entre capteurs, etc.), ce qui nous amène à garder des réserves quant à la pertinence et la fiabilité des mesures effectuées. Ce constat rejoint les résultats d'autres études, notamment de Budde *et al.* (2014), qui montrent cependant le potentiel de ces capteurs à faible coût pour des applications de sciences participatives.

Pour le montage de la station portative, plusieurs choix étaient possibles, du mode d'interrogation des données jusqu'au boîtier. Pour la première version (Fête de la Science 2015), nous avons retenu le capteur *Sharp GP2Y10*, le microcontrôleur *RFduino* pour sa compacité et son faible coût, qui communique *via Bluetooth* principalement avec des téléphones mobiles évolués (Android et iPhone ; Figure 1). L'application dédiée sur le téléphone portable interroge toutes les 30 secondes la

station, localise la mesure et alors envoie cette information, datée, sur un serveur pour constituer une base de données géographiques.

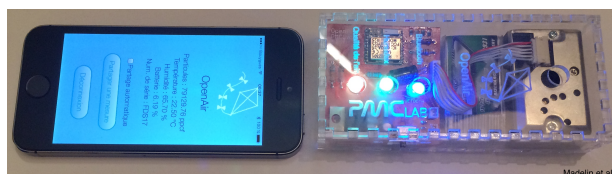


FIGURE 1. Station environnementale OpenAir et application sur téléphone

3. Montage de la station lors de la Fête de la science 2015 et 2016

Dans ce projet, nous tenons à permettre aux citoyens d'évaluer de manière active la qualité de leur environnement proche et en particulier de l'air qu'ils respirent, avec une participation dès la construction de la station de mesures de la pollution de l'air, lors d'ateliers (en particulier pendant la Fête de la science). Trois ateliers ont eu lieu en octobre 2015 et 2016, avec une trentaine de personnes (d'autres sont prochainement programmés). Les participants ont été initiés à la thématique de la qualité de l'air, ont fabriqué une station de mesures environnementales à l'aide d'un kit de montage et d'un livret, installé l'application sur leur téléphone et effectué quelques mesures à l'intérieur du FabLab, notamment en conditions forcées. Pour le 1^{er} atelier, le questionnaire préalable a montré une relative méconnaissance de la pollution de l'air et de sa variation spatiale. Les participants étaient plus motivés par la construction d'une station environnementale et leur contribution à la base de données a rapidement diminué après l'événement de la Fête de la science. Même si l'apport quant aux sciences participatives est quantitativement faible, l'atelier a permis de sensibiliser les participants à la thématique de la pollution de l'air. Le public des deux autres ateliers était plus averti (étudiants et personnes intéressées par la pollution) et la contribution s'avère plus forte, même s'il reste difficile de prolonger la contribution au-delà de l'événement. Ce projet de recherche participative s'inscrit clairement sur la sensibilisation à la pollution de l'air, sur l'apport d'une connaissance sur le sujet, sur la possibilité de fabriquer dans des FabLab et plus modestement sur l'alimentation d'une base de données géographiques.

4. Représentation cartographique des mesures participatives

Début octobre 2016, la base représentait plus de 7 000 données de température, humidité et concentrations de particules, associées à une date et heure et à une localisation. Cette base de données (PostGIS) est ensuite interrogée et cartographiée, en autres *via* les librairies Leaflet et d3. La précision de la localisation des mesures est parfois grossière, principalement en raison de la densité du bâti en milieu urbain et des effets de masque. Les données de concentrations en particules semblent

SAGEO'2016 – Nice, 6-9 décembre 2016

parfois aberrantes. Au-delà d'une critique des données, la cartographie des mesures questionne surtout sur les choix d'implantation et de représentation spatiale à privilégier (et *ipso facto* aux solutions techniques) : quelle est la représentativité d'une mesure ponctuelle ? les mesures de concentrations sont-elles à représenter par des points, par carroyage ou encore par agrégation le long d'un linéaire pour un itinéraire ? compte tenu de la relative précision des mesures, est-ce préférable de discrétiser l'information sous une forme assez grossière, telle qu'un indice à 6 classes ? etc. Cette réflexion non figée encore sur les choix cartographiques soulève la question du message à communiquer, en particulier lorsqu'il s'agit, comme ici, d'une information sensible et de mesures individuelles, faites dans des contextes différents et avec des capteurs présentant des incertitudes. Or la cartographie ici en ligne (Figure 2) s'avère un vecteur efficace pour communiquer, valoriser et diffuser l'information dans une démarche de sciences participatives.



FIGURE 2. Extrait du site web proposant la cartographie des mesures avec la station portative OpenAir (sources : Leaflet - OpenStreetMap ; données OpenAir)

Ce projet vise une démocratisation de la science et une volonté de sensibilisation aux débats scientifiques et aux questions environnementales. A l'issue des différents ateliers, les participants sont repartis avec un capteur construit par leur soin et une connaissance sur l'îlot de chaleur urbain, sur les particules en suspension et leur variabilité spatiale. Leur participation à l'enrichissement de la base de données a montré les difficultés d'assurer une collecte riche et continue dans le temps. Peut-être en aurait été-il différent avec un capteur acheté par les participants ? Nous menons actuellement cette réflexion, parallèlement au développement de nouvelles stations (avec d'autres polluants, avec des capteurs enregistreurs pour éviter la décharge de la batterie des téléphones). Enfin, nous réfléchissons aussi aux usages de ces stations et à la modification des comportements liée à l'utilisation de ces stations. Les réponses à une enquête post-atelier permettront d'alimenter cette piste et de modifier, éventuellement, l'application.

Bibliographie

- Budde M., Zhang L., Beigl M., 2014. Distributed, Low-Cost particulate matter sensing: scenarios, challenges, approaches. *Proc. 1st Int. Conf. Atmospheric Dust 2014*, 230-236.
- Dickinson J.L., Shirk J., Bonter D., Bonney R., Crain R.L., Martin J, Phillips T, Purcell K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, vol. 10, n° 6, p. 291-97.
- Duché S., Madelin M. (2015). Les capteurs *low cost* de pollution : un nouvel eldorado pour l'évaluation de l'exposition individuelle aux particules ? *Actes du XXVIIIe Colloque de l'Association Internationale de Climatologie*, Liège, p. 140-145.
- Goodchild M.F., Li L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, vol. 1, p. 110-120.
- Ifop - Institut français d'opinion publique (2014). *Conférence de citoyens : « Comment lutter contre la pollution de l'air à Paris ? »*. Rapport de l'Ifop pour la Mairie de Paris, 12 p.

Modélisation 3D de l'expansion urbaine : application au modèle SLEUTH

Guillaume Da Silva¹, Omar DOUKARI¹, Rahim AGUEJDAD²

1. Institut de Recherche en Constructibilité de l'ESTP

28, avenue du Président Wilson - 94234, Cachan cedex

gds75016@gmail.com, odoukari@adm.estp.fr

2. Laboratoire Image, Ville, Environnement UMR-7362 CNRS/Unistra

3, rue de l'Argonne - 67000, Strasbourg

rahim.aguejedad@live-cnrs.unistra.fr

RESUME. Parmi les modèles spatiaux de simulation de l'expansion urbaine, le modèle SLEUTH basé sur un automate cellulaire et une approche inductive est largement utilisé dans plusieurs applications. Cependant, son fonctionnement est basé sur une représentation en 2D. De plus, le facteur réseau routier reste statique au cours de la simulation, ce qui limite la performance du modèle quant à l'allocation spatiale à proximité des routes. L'objectif de ce travail est de faire de SLEUTH un modèle en 3D en intégrant la hauteur des bâtiments, la mise à jour de la carte du réseau routier et enfin le découpage de la tache urbaine en tissu urbain. Ces améliorations permettront d'affiner la simulation de l'expansion urbaine tout en facilitant le couplage du modèle SLEUTH avec d'autres outils de simulation en vue de l'estimation de l'îlot de chaleur urbain, le trafic, la pollution et les besoins énergétiques des bâtiments.

MOTS-CLES : SLEUTH, expansion urbaine, modélisation 3D, parcellisation.

ABSTRACT. Among all the existing urban growth inductive-pattern-based models, SLEUTH cellular-automata based model is widely used in many applications. Nevertheless, SLEUTH is limited to a 2D modeling framework. Moreover, the road network input variable remains static over the simulation time, which limits the performance of the model on the spatial allocation using the roads-influenced algorithm. Therefore, the objective of this work is to include the 3D aspect by automatically updating the transportation map, and integrating the buildings height and the urban fabric with parcels and streets. These improvements are expected to significantly improve the simulation of urban expansion. They would subsequently facilitate the coupling of SLEUTH with other simulation tools to deal the estimation of urban heat island, traffic, pollution and energy buildings requirements.

KEYWORDS: SLEUTH, urban growth, 3D modelling, partitioning.

SAGEO'16 – Nice, 6-9 décembre 2016

1. Introduction

Parmi tous les modèles dynamiques et spatialement explicites (Aguejdad *et al.*, 2016), ceux basés sur les automates cellulaires sont probablement les plus répandus pour leurs applications en milieu urbain (Clarke, 2008 ; Santé *et al.*, 2010). Ceci s'explique notamment par le caractère spatialement explicite et interactif des automates cellulaires, et par l'intégration aisée des classifications issues des données de télédétection dans ces modèles. C'est le cas du modèle SLEUTH dédié à la simulation dynamique de l'expansion urbaine (Candau, 2002 ; Clarke, 2008).

SLEUTH est un modèle robuste largement utilisé et ayant fait l'objet de plusieurs applications et travaux de recherche (Clarke, 2008). Parmi ses points forts, on peut citer sa capacité à simuler quatre grandes formes d'expansion urbaine : en continuité de l'urbain existant, par création de nouveaux centres ou foyers urbains, par diffusion et le long des axes routiers. Cependant, son fonctionnement reste basé sur une représentation en 2D et une simple répartition binaire de l'espace en deux classes : tache urbaine et zones non urbanisées. La tache urbaine englobe au sens large l'ensemble des surfaces urbanisées. Elle comprend ici les espaces bâtis, les parkings, la voirie, les surfaces en eau, les espaces verts et les espaces boisés associés à l'urbanisation. De plus, le facteur réseau routier reste statique au cours de la simulation, ce qui limite la performance du modèle quant à l'allocation spatiale le long des routes. Par conséquent, l'objectif de ce travail est de faire de SLEUTH un modèle 3D en intégrant la hauteur des bâtiments, la mise à jour automatique du réseau routier et enfin la parcellisation de la tache urbaine.

L'article est organisé en trois sections. La **première section** expose l'algorithme de parcellisation automatique des zones urbaines. La **deuxième section** revient sur la création dynamique de nouvelles routes reliant les fragments urbains simulés au réseau routier existant. La dernière **section** présente la modélisation 3D des bâtiments.

2. Méthode, résultats et discussion

2.1. Subdivision de la tache urbaine en parcelles

Faire de SLEUTH un modèle 3D réaliste nécessite le passage d'une tache urbaine homogène à des lotissements et des rues en appliquant un algorithme de partitionnement. Nous nous sommes donc penchés sur un algorithme nommé « *Parcel Divider* » (Dahal and Chow, 2014) qui semblait être efficace sur un bon nombre de formes urbaines. Cet algorithme, développé dans l'environnement ArcGIS, a d'abord nécessité son adaptation à SLEUTH, mais aussi au langage JAVA qui ne possède que peu de méthodes géométriques.

Cet algorithme repose tout d'abord sur une simplification des zones urbaines en les englobant par des rectangles de taille minimum (Figure 1). Chacun de ces rectangles est ensuite traité en fonction du rapport entre son étendue et la taille d'un

SAGEO'2016 – Nice, 6-9 décembre 2016

lotissement. Enfin, lorsque les subdivisions des rectangles sont achevées, il ne reste plus qu'à tronquer les délimitations obtenues selon la véritable forme des zones urbanisées.

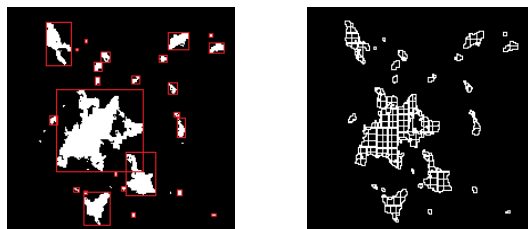


Figure 1 : Détection des zones à partitionner

2.2. Création des routes

Le modèle SLEUTH présente, quant à lui, déjà une carte du réseau routier mais cette carte reste inchangée au cours de l'évolution de la ville. Il paraissait donc pertinent de rendre le réseau routier dynamique et le mettre à jour automatiquement en s'inspirant de l'algorithme « *Parcel Divider* ». Cet algorithme consiste donc à raccorder les fragments individuels nouvellement simulés au réseau routier existant (Figure 2). Les nouveaux tronçons sont créés suivant le plus court chemin avec une prise en compte de la présence d'éventuelles zones protégées. Ceci est d'autant plus important que l'expansion urbaine simulée par SLEUTH est, entre autres, influencée par le réseau routier. Ainsi, la création de nouvelles routes n'était plus seulement nécessaire pour la modélisation 3D, mais aussi pour améliorer la performance prédictive du modèle.



Figure 2 : Raccordement des zones urbaines simulées au réseau routier existant

2.3. Modélisation 3D des bâtiments

Cette dernière étape consiste à créer une base pour l'environnement, appelée une scène. Cette dernière inclut tous les éléments du modèle, en l'occurrence le terrain, les bâtiments et les routes. Les bâtiments présentés sont intégrés à la scène à l'aide de leurs coordonnées 2D (Figures 3 et 4). Les informations concernant le nombre de bâtiments, leur hauteur ou bien leur type sont à déterminer au préalable par l'utilisateur.

L'algorithme de modélisation des bâtiments implique trois composantes élémentaires : sommet, rez-de-chaussée et étage. Ces trois composantes sont ensuite imbriquées afin de créer le modèle. Le sommet et le rez-de-chaussée du bâtiment restent uniques, tandis que l'étage est imbriqué n fois, n étant le nombre d'étages du bâtiment.



Figure 3 : Tissu résidentiel mixte



Figure 4 : Immeubles avec un nombre variable d'étages

3. Conclusion

Cette étude présente les premiers résultats du développement d'un modèle d'expansion urbaine 3D en utilisant l'approche BIM. Les principaux développements ont permis : (i) le découpage de la tache urbaine simulée par SLEUTH en lotissements ; (ii) le raccordement des fragments urbains individuels au réseau routier existant par création de nouveaux tronçons ; et (iii) la simulation de trois modèles de bâtiments : maison individuelle et immeuble. Par ailleurs, en plus des formes urbaines et des caractéristiques géométriques des bâtiments, des travaux en cours consistent à faire de SLEUTH 3D un modèle paramétrique capable d'intégrer des facteurs explicatifs du développement résidentiel et des paramètres additionnels associés aux bâtiments.

Bibliographie

- Aguejdad R., Doukari O., Houet T., Avner P., Viguié V., 2016. Modélisation prospective de l'étalement urbain : apports et limites des modèles de spatialisation. Application aux modèles SLEUTH, LCM et NEDUM-2D. *Cybergeo*, No. 782.
- Candau J., 2002. Temporal calibration sensitivity of the SLEUTH urban growth model. Master's thesis, Department of Geography, University of California, Santa Barbara, CA.
- Clarke K.C., 2008. A Decade of Cellular Urban Modeling with SLEUTH: Unresolved Issues and Problems. In Brail R.-K. (eds.), *Planning Support Systems for Cities and Region*, Lincoln Institute of Land Policy, Cambridge, MA, 47–60.
- Dahal K.J., Chow T.E., 2014. A GIS toolset for automated partitioning of urban lands. *Environmental Modelling and Software*, vol. 55, 222–234.
- Santé I., García A.M., Miranda D., Crecente R., 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, vol. 96, No. 2, 108–122.

Projet PLU++ : vers un Plan Local d'Urbanisme plus intelligible et plus participatif

Mickaël Brasebin¹, Sidonie Christophe¹,
Florence Jacquinod², Anouk Vinesse¹, Hortense Mahon²

1. ^b Université Paris-Est, COGIT, IGN, LASTIG, UPEMLV, 73 av. de Paris, 94160 Saint-Mandé, France

{mickael.brasebin@ign.fr}, {sidonie.christophe@ign.fr}, {anouk.vinesse@ign.fr}

2. ^b Université Paris-Est, SUN, EIVP, UPEMLV, 80 rue Rébeval 75019 Paris, France

{jacquinod@gmail.com}, {hortense.mahon@eleves.ec-nantes.fr}

RÉSUMÉ. Les Plans Locaux d'Urbanisme (PLU) sont des documents d'urbanisme qui régissent la constructibilité au niveau de la parcelle. Si la concertation portant sur ces documents est obligatoire, elle est difficile à mener. Une des raisons propre à cela est qu'il est difficile pour le non-initié de comprendre le contenu technique de ces documents et d'évaluer l'influence de ce type de règlement. L'idée du projet PLU++, présenté dans ce résumé, est de mener une réflexion sur la conception d'un système de géo-visualisation 3D visant à faciliter la concertation autour des PLU. Ce système s'articule autour d'un sous-système de simulation 3D qui permet de générer des configurations bâties respectant les règles du PLU et d'un sous-système de stylisation qui offre la possibilité de représenter les objets géographiques avec différents styles. Ainsi, à l'aide de ce système, différentes hypothèses sur les besoins des concertés en fonction de l'objet de la concertation pourront être testées, notamment en proposant différentes représentations et interactions avec une maquette 3D. L'objectif final derrière ce travail est de pouvoir faire émerger de bonnes pratiques de concertation 3D.

ABSTRACT. Local Urban Plan (LUP) is a urban document that regulates the Right to Build at the parcel scale. If public participation is mandatory for its elaboration or for its modification, it is difficult for non-expert people to evaluate the influence of this regulation on a territory and its urban morphology. The aim of the PLU++ project is to lead a reflexion about the design of a 3D geo-visualisation system that makes public participation about LUPs easier. This system is built over two sub-systems: a 3D simulation sub-system that enables the generation of built configurations that respect

SAGEO'2016 - Nice, 6-9 décembre 2016

LUPs rules and a 3D stylization sub-system that implements the possibility to represent geographic features with different styles. Thus, thanks to this system, different hypothesis about participants' needs according to the context and 3D representation effectiveness will be tested in order to foster good practices for public participation.

MOTS-CLÉS : PLU, simulation, geo-visualisation 3D, concertation, stylisation

KEYWORDS: urban planning, simulation, 3D geo-visualization, stylization

La géo-visualisation 3D est de plus en plus souvent utilisée comme support à la concertation sur des projets d'urbanisme. Cette concertation peut par exemple concerner des projets de construction ou de réaménagement d'un îlot. Elle est généralement très appréciée par l'ensemble des acteurs en tant que support d'explication et de communication plus accessible au plus grand nombre que les cartographies en deux dimensions car elle offre une vue relativement intuitive de l'évolution locale du territoire.

Parallèlement à cela, la concertation tend à se développer autour de l'élaboration ou de la révision de documents d'urbanisme qui régulent l'évolution du territoire. Cette phase de concertation est même obligatoire depuis la loi SRU votée en 2000 en ce qui concerne les Plans Locaux d'Urbanisme (PLU). Le règlement des PLU a notamment pour objectif de définir le droit à bâtir à l'échelle de la parcelle à travers un ensemble de contraintes portant sur la morphologie 3D et sur l'apparence de nouveaux bâtiments. Pourtant, il est assez difficile de motiver le public à participer aux concertations portant sur ce type de documents. Une des raisons à cela est qu'il est difficile à partir d'un document textuel et d'une carte de s'imaginer l'influence d'un règlement sur l'évolution du territoire pour les prochaines années.

Ainsi, l'objectif du projet PLU++ (<http://ignf.github.io/PLU2PLUS/>) présenté dans ce résumé est de mener une réflexion autour des outils de géo-visualisation 3D sur les bonnes pratiques pour concevoir un dispositif de concertation dédié aux discussions portant sur les PLU et permettent d'améliorer la compréhension des règles par le grand public.

Démarche du projet

La démarche globale du projet (présentée en figure 1 qui introduit les différentes parties de ce résumé) vise à un questionnement entre le dispositif technique et son adaptation à différents contextes de concertation. Un aspect préalable au projet est le fait d'avoir un certain nombre d'**hypothèses sur les besoins d'outils 3D en situation de concertation** (section 1) en fonction d'un cas d'usage (Jacquinod, 2014), que l'on peut définir par un public cible et un objectif (état de discussion du PLU, problématiques urbaines associées à la zone de discussion, etc.). Pour tester ces hypothèses, un prototype de géo-

SAGEO'2016 – Nice, 6-9 décembre 2016

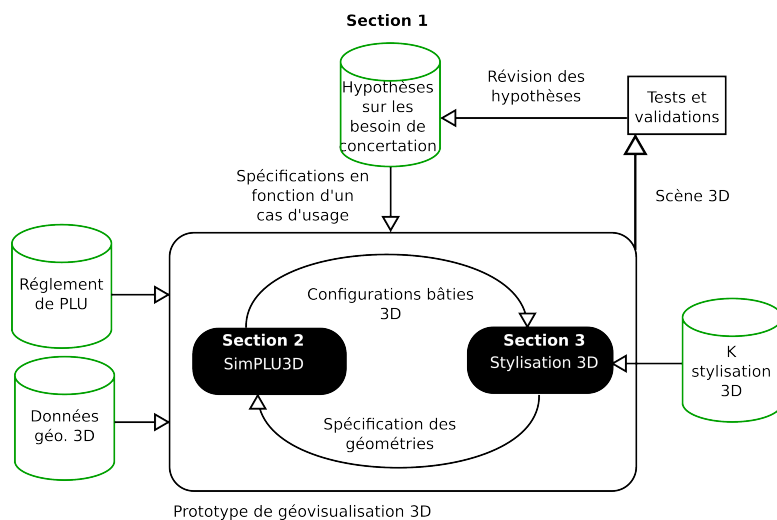


FIGURE 1. Illustration de la démarche globale du projet PLU++

visualisation 3D est conçu pour permettre la production de scènes 3D supposées faciliter la concertation. Afin d'explicitier les informations réglementaires dans des scènes 3D, il est nécessaire d'explicitier cette information par le biais d'une représentation géométrique. Pour faire cela, nous proposons d'utiliser des simulations de **configurations bâties 3D** (section 2) qui sont produites par la bibliothèque SimPLU3D¹ issu de travaux antérieurs (Brasebin, 2014). Ces configurations ainsi que les autres objets de la scène sont **stylisés en 3D** (section 3) à partir d'un ensemble de méthodes et de processus mis en œuvre par ailleurs et notamment dans le cadre du projet ANR MapStyle². De par ses sous-systèmes, le système de géo-visualisation offre de nombreuses possibilités de paramétrisation, ainsi, nous pouvons adapter les scènes produites et leurs interactions en fonction des hypothèses testées. Nous espérons pouvoir valider par expérimentation avec des praticiens certaines d'entre elles afin d'améliorer les bonnes pratiques de concertation voire en faire émerger de nouvelles.

1. Hypothèses sur les besoins d'outils 3D en situation de concertation

Pour formuler ces hypothèses, nous nous appuyons sur un ensemble d'études préliminaires, des entretiens avec des praticiens et des observations directes de

1. Site Internet du simulateur SimPLU3D : <https://github.com/IGNF/simplu3D>

2. Site Internet du projet MapStyle : <http://mapstyle.ign.fr/>

dispositifs existants de concertations. Ces hypothèses portent à la fois sur la manière de présenter le résultat des simulations (le focus) et la représentation des autres objets géographiques nécessaires de la zone (le contexte). Elles ont été formulées ainsi suite aux échanges avec les praticiens :

- La représentation des configurations bâties simulées doit atteindre un certain degré d’abstraction pour faire comprendre que l’objet de la concertation est la morphologie urbaine à l’échelle d’un quartier ou d’une ville et que les résultats des simulations représentent des possibilités d’évolution du territoire et non des projets qui seront amenés à être réalisés ;
- Les objets du contexte doivent également être rendus avec un certain niveau d’abstraction pour que leur représentation soit centrée sur la forme urbaine existante et non sur le détail des constructions individuelles ;
- La représentation globale de la zone étudiée doit mettre l’accent visuellement sur le focus, tout en permettant une comparaison avec le contexte en termes de forme urbaine ;
- La scène doit être la plus complète possible en termes d’objets par rapport au sujet de la concertation afin que les participants aient les éléments pour mener une réflexion sur la problématique ciblée.

2. Simulation de configurations bâties 3D

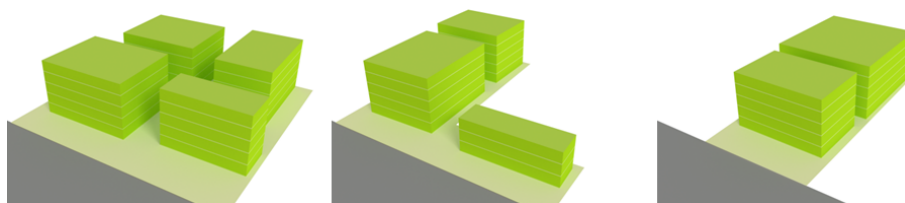


FIGURE 2. Quelques résultats de simulation issues de *SimPLU3D* sur différentes parcelles (source : <https://github.com/IGNF/simplu3D>)

Pour fournir une représentation géométrique du règlement, la bibliothèque *SimPLU3D* permet de générer des configurations bâties (voir par exemple la figure 2). Pour effectuer une simulation sur une parcelle, *SimPLU3D* prend en entrée des données géographiques 3D (parcelles, bâtiments existants, routes, etc.), les intègre dans un modèle permettant la vérification des règles d’urbanisme et simule la configuration qui respecte ces règles et optimise une fonction d’optimisation. Ces configurations sont formées d’un ensemble homogène d’objets (par exemple des boîtes dans les illustrations suivantes). La configuration optimisée proposée par le système est celle qui maximise une fonction donnée qui modélise le comportement d’un agent constructeur (par exemple, on pour-

SAGEO'2016 – Nice, 6-9 décembre 2016

rait choisir une optimisation du volume pour simuler la stratégie d'un agent qui souhaiterait maximiser le bénéfice du droit à bâtir).

Ainsi, pour adapter les simulations aux hypothèses à tester, il est possible de jouer sur plusieurs leviers à plusieurs niveaux. Pour s'adapter aux objectifs de concertation, il est capable de produire des simulations avec différents règlements pour discuter différents scénarios de PLU ou de générer des configurations bâties optimisant une fonction en rapport avec les enjeux d'une zone à urbaniser (par exemple, limiter la projection d'ombrage chez le voisin pour l'aménageur d'un lotissement).

Dans la perspective de prendre en compte la manière avec laquelle ces configurations sont représentées, il est possible d'ajuster :

- le niveau de détail des sorties, cela peut être utile afin de produire un niveau de détail homogène entre le focus et le contexte ou de rendre la forme plus ou moins définie en fonction de l'avancement des discussions;
- de les post-traiter pour permettre l'application de processus spécifiques de stylisation. Par exemple, le rendu "coup de crayon" sur des arêtes nécessite une sélection des arêtes pertinentes et une tessellation pour obtenir une surface sur laquelle on pourra appliquer la texture.

3. Stylisation 3D

Afin de tester les hypothèses formulées sur la compréhension des usagers en fonction des styles, un certain nombre ont été formalisés et étudiés. Ces styles sont sélectionnés car employés pour répondre à différents usages, les tests permettront ainsi de valider l'adéquation entre ces styles et leur usage de prédilection. De plus, dans le cadre de la concertation, le fait de pouvoir changer de style lors de la discussions permet de mettre en avant certaines problématiques de la concertation..

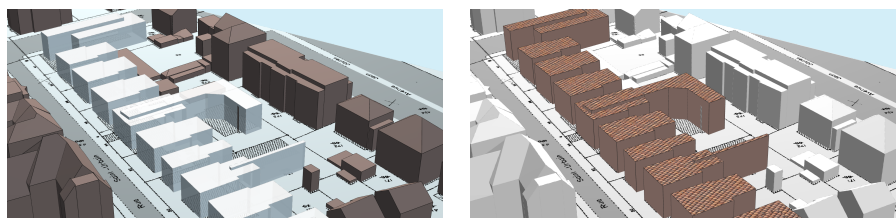
Pour débiter, trois catégories de styles sont considérés : photo-réaliste, typique³ et discret⁴.

Au sein de chaque catégorie, il est possible de produire une variété de styles en procédant à des variations des variables graphiques sous-jacentes (par exemple, en changeant la largeur des traits ou la couleurs des objets).

Ces styles peuvent être appliqués sur le focus ou le contexte de manière à les distinguer créant ainsi un contraste de styles entre ces différents thèmes d'objets.

3. Les bâtiments sont colorés avec des aplats de couleurs proches des rendus photo-réalistes avec des arêtes visibles.

4. Les objets sont rendus en blanc avec un certain niveau de transparence et une représentation fine des arêtes.



(a) Style discret pour les configurations simulées et typique pour le contexte.

(b) Style semi-réaliste pour les configurations simulées et discret pour le contexte.

Figure 3.

La figure 3 présente quelques premiers rendus qui mixent différentes catégories de style. Comme la catégorie "style abstrait" est un ensemble comportant de nombreuses familles (aquarelle, esquisse technique, etc.), nous envisageons d'explorer d'autres styles provenant de ces familles. Les hypothèses formulées ne seront pas uniquement testées au sein de styles statiques, mais également à travers les possibilités de changements de styles pendant les discussion afin d'en évaluer leur pertinence.

Conclusion

Le projet PLU++ s'appuie sur des travaux récents issus de différents domaines, comme la simulation, la géo-visualisation 3D et la formalisation de contenu adapté à l'utilisateur pour faire des propositions visant à simplifier la concertation autour des PLU. De premières expérimentations seront menées prochainement en relation avec une communauté de communes et permettront de valider les hypothèses sur leurs besoins en termes de concertation et donc d'avoir un retour sur l'intérêt des résultats de simulations et des propositions sur les méthodes de stylisation adaptées. De même, il se posera la question d'ajouter des informations thématiques relatives aux différentes simulations (par exemple, ensoleillement, équivalence en nombre de logements, etc.) afin de donner des éléments supplémentaires appréciatifs des différents scénarios étudiés.

L'ensemble des productions scientifiques du projet (Maquettes 3D, rapport d'enquête, etc.) sera mis à disposition sur le site Internet du projet. De futures évolutions sont envisageables comme la production de différentes catégories d'objets dans le simulateur comme les parkings et la végétations qui sont également des éléments régulés par les PLU et donc de potentiels sujets de discussion. À termes, nous espérons que les bonnes pratiques développées dans le cadre du projet permettront d'améliorer les systèmes de concertation territoriale dans leur ensemble.

SAGEO'2016 – Nice, 6-9 décembre 2016

Remerciements

Cette recherche a été financée par la COMUE Paris-Est dans le cadre de l'appel conjoint à Projets Exploratoires Premier Soutien (PEPS) Interdisciplinaires.

Bibliographie

- Brasebin M. (2014). Les données géographiques 3D pour simuler l'impact de la réglementation urbaine sur la morphologie du bâti (Thèse de doctorat, Université Paris-Est). *Mémoire de thèse*.
- Jacquino F. (2014). Production, pratique et usages des géovisualisations 3D dans l'aménagement du territoire (Thèse de doctorat, Université de Saint-Etienne). *Mémoire de thèse*.

Simulation par Automate Cellulaire et Système Multi-Agents des conséquences du changement climatique sur l'aire de distribution de *Pinus Cembra* dans le Mercantour.

Matthieu VIGNAL¹ Julien ANDRIEU²

1. UMR ESPACE, Université Nice Sophia Antipolis
98, bd E. Herriot, 06204 Nice Cedex
matthieu.VIGNAL@unice.fr

2. UMR ESPACE, Université Nice Sophia Antipolis
98, bd E. Herriot, 06204 Nice Cedex
julien.ANDRIEU@unice.fr

RESUME. Le dernier rapport du GIEC fait état de nombreuses conséquences possibles sur la biodiversité terrestre induites par l'élévation des températures. Arbre symbolique de la haute montagne, *Pinus cembra* est l'une des espèces qui constitue la grande richesse spécifique des Alpes du Sud franco-italiennes. L'objectif de cette étude est de définir l'aire de distribution future de cette espèce végétale, préalable nécessaire à toute réflexion sur la mise en place de stratégies adaptatives permettant d'assurer le maintien de cette espèce et l'intégrité de son milieu face au changement climatique. La méthode élaborée permet, à partir de relevés d'espèces issus de la base de données SILENE, de campagnes de terrain et de descripteurs écologiques de spatialiser l'aire de distribution d'une espèce afin d'en étudier les changements et l'intégrer dans un modèle dynamique, dont l'objectif est de formaliser les modifications de la distribution végétale à travers la simulation des processus biologiques.

ABSTRACT. The last IPCC report presents numerous consequences of temperature elevations on terrestrial biodiversity. Mountain's symbol, *Pinus cembra* is one of the species that constitute the richness of South Franco-Italian Alps. The aim of this study is define the distribution area of this species, necessary prerequisite to any discussion of adaptive strategies for maintenance of this species and the integrity of its environment face to climate change. The method developed here enable, with botanical observation form SILENE database, field campaign and ecological descriptors to give a spatial definition of the distribution areas in order to begin the changes study and to integrate it in a dynamic model which aims to formalize the changes in distribution areas through the simulation of complex biological processes.

MOTS-CLES : Biogéographie ; Modélisation ; Changement climatique ; Aires de distribution.

KEYWORDS: Biogeography; Modeling; Climate Change; distribution areas.

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction.

Le changement climatique est l'une des causes probables de la réduction de la biodiversité. Aujourd'hui, l'évolution des températures, plus rapide que les précédents changements climatiques, pourrait entraîner l'extinction de nombreuses espèces, végétales ou animales. Les Alpes du Sud franco-italiennes se caractérisent par une diversité climatique et géologique, expliquant une grande richesse spécifique (Thompson, 2005). Parmi ces espèces, *Pinus cembra* est un arbre dont les caractéristiques, notamment la chorologie, constituent un défi intéressant pour la modélisation spatialement explicite. Lors des phases où le climat change, les espèces risquent de s'éteindre si leurs aires de distributions ne connaissent pas l'opportunité de trouver, en un autre lieu, les conditions écologiques qui leur sont favorables, d'atteindre celui-ci et de s'y implanter durablement (Alexandre et Genin, 2011). Il est donc aujourd'hui essentiel de définir l'emprise spatiale à l'horizon 2100 de cette espèce végétale montagnarde afin d'identifier les effets probables du changement climatique sur cette dernière. C'est un préalable nécessaire à toute réflexion sur la mise en place de stratégies adaptatives permettant d'assurer le maintien des espèces et l'intégrité de son milieu face au changement climatique. Actuellement, la plupart des modèles de prévision de l'impact du changement climatique sur la biodiversité n'appliquent que des règles climatiques génériques sur une « aire de distribution potentielle actuelle » pour délimiter des « aires de distributions potentielles futures » (Casazza, 2010). Or, ces aires de distributions potentielles futures sont biaisées puisqu'elles ne prennent pas en compte la capacité de propagation de l'espèce, de l'aire actuelle vers de nouveaux milieux, ni même les modalités de la mortalité sans régénération *in situ*. Cette prise en compte est un enjeu majeur, d'une part en écologie et d'autre part en géomatique. Ainsi, l'objectif de cette étude est de coupler la démarche naturaliste avec la démarche de modélisation. Il s'est agi, à travers la littérature et les campagnes de terrain, de récolter une quantité de données suffisante afin de mettre en place un outil de modélisation capable de simuler les processus biologiques des espèces végétales. L'objectif est de mettre en avant les effets probables du changement climatique sur l'aire de distribution de *Pinus cembra* en tenant compte de la compétition interspécifique. Une attention particulière sera portée sur la manière dont ce modèle permet de simuler le mode de dissémination de cette espèce ornithochore.

2. Données.

Les données concernant les caractères biologiques de cette espèce proviennent d'un ensemble d'ouvrages. Les relevés botaniques proviennent de la base de données SILENE Flore. Ils fournissent la base de connaissance ponctuelle des présences de l'espèce. Des variables écologiques climatiques (Martin et al, 2013) et météorologiques (Météo France, 2014) ; topographiques (IGN ; GN) et géologiques (OneGeology) ont été utilisées afin de permettre l'approche auto-écologique. Certaines sont converties en variables de conditions environnementales en entrée du modèle, notamment à travers un indice de viabilité permettant d'évaluer la probabilité du pixel à être colonisé par une espèce.

3. Méthode.

Afin de compléter les données issues de la littérature, des campagnes de terrain sont réalisées. Ces campagnes ont quatre objectifs. D'abord, constituer une base de relevés floristiques des sites étudiés afin, premièrement, de compléter et vérifier les données SILENE Flore, d'obtenir un site de « vérité terrain » où l'absence signifie bien l'absence d'individu et, de spatialiser une population au sein de l'aire de distribution. Ensuite, une pyramide des âges est réalisée afin de permettre prudemment la validation du modèle. Enfin, les observations de *Nucifraga caryocatactes* permettent de caractériser le comportement de cette espèce animale notamment de ces déplacements et des cachettes qu'ils constituent.

L'analyse fréquentielle permet de caractériser la distribution des espèces à travers la détermination des optimums et limites qui conditionnent la mortalité et la viabilité future d'individus en cas de dissémination d'une graine en un lieu. Dès lors, cela permet de spatialiser les aires de distribution des espèces par le seuillage d'une carte de probabilité d'occurrence des espèces. Cette chaîne de traitements produit l'état actuel des aires pour l'élaboration du modèle dynamique. Afin d'ajouter une dimension « population » à l'objet « aire de distribution actuelle », la densité d'individus est spatialisée par l'intermédiaire d'une méthode d'interpolation spatiale de type krigeage.

Afin de modéliser les aires de distribution futures, le modèle va se baser sur les processus qui régissent les dynamiques de population (cf. figure 1). Le modèle est spatialement explicite, c'est-à-dire que les données sont spatialisées (températures, indice de viabilité, etc.) ainsi que l'ensemble des processus biologiques. C'est un système multi-agents (SMA) qui comporte un automate cellulaire (AC). Il est composé de deux types d'entités : les pixels (200 mètres) qui représentent l'environnement, et les individus que sont les individus de *Pinus cembra*, les graines et les oiseaux. Les aires de distribution actuelles sont modifiées, au fil du changement climatique, à condition que les processus biotiques le permettent. Les processus de ce modèle s'enchaînent au fil des mois. Les individus aptes à produire des graines entament le processus de dissémination. Par la suite, les graines peuvent entamer ou peuvent ne pas entamer le processus de germination qui dépend de l'indice de viabilité *in situ*, de la viabilité des graines et de l'occupation du sol. Les dix premières années, la présence de l'espèce est soumise à la mortalité infantile en appliquant un seuil thermique.

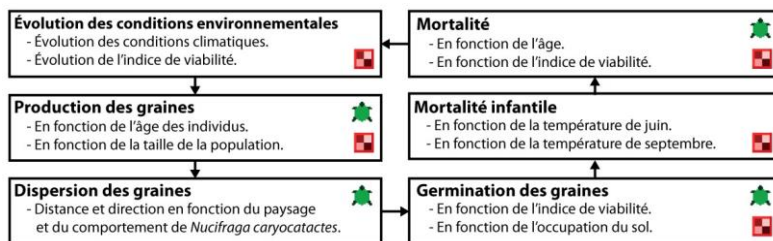


FIGURE 1. Diagramme d'activités du modèle.

SAGEO'2016 – Nice, 6-9 décembre 2016

4. Résultats.

L'analyse de l'efficacité des descripteurs met en évidence que les descripteurs climatiques possèdent des efficacités bien plus importantes que les descripteurs topographiques ou géologiques. L'aire de distribution potentielle de *Pinus cembra* couvre, aujourd'hui, l'ensemble des versants de montagnes de 900 à 2800 mètres. Néanmoins, au sein de cette aire la densité d'individus n'est pas répartie uniformément. En paramétrant selon les connaissances actuelles de la biologie de cette espèce, le déplacement est possible et les pixels viables dans le futur semblent être colonisés. Le changement climatique permettrait à cette espèce de subsister dans de nombreux espaces. Dans un premier temps seront discutés les résultats des simulations sans la prise en compte de la compétition interspécifique et dans un deuxième temps, les résultats des simulations avec la prise en compte de la compétition interspécifique.

5. Discussion et conclusion.

D'abord, les données en entrée sont à prendre avec un certain nombre de précautions. La base de données Silène présente des problèmes de représentativité. Les données climatiques présentent des marges d'erreur en particulier pour les pluies (Martin et al. 2013). Ensuite, les aires de distribution actuelles sont issues d'un traitement des données générant également une certaine marge d'erreur. Enfin, l'individu graine dans le modèle, bien que considéré comme telle, n'est pas un individu à part entière. Il serait particulièrement intéressant d'intégrer à chaque graine un comportement propre. En conclusion, les campagnes de terrain ont permis de confronter la littérature au terrain. Les affinités écologiques de l'espèce ont été déterminées par l'analyse statistique à travers la définition de l'optimum écologique et des limites de tolérances. La détermination de l'aire de distribution a permis de définir l'emprise spatiale de l'espèce. Afin de modéliser les aires de distribution futures, la production, la dispersion et la germination des graines, mais également la prise en compte de la mortalité ont été implémentés dans un système multi-agents. Le modèle fonctionne et permet d'engendrer des modifications qui s'approchent de celles exposées dans la littérature.

Bibliographie.

- Alexandre F. Génin A. (2011). *Géographie de la végétation terrestre ; Modèles hérités, perspectives, concepts et méthodes*, Edition Armand Collin.
- Casazza G. Zappa E. Mariotti M. G. Médail F. Minuto L. (2010). Distribution pattern and richness of endemic plant species in Maritime and Ligurian alps. *Boll. Mus. Ist. Biol. Unic*, n° 72, p. 130-136.
- Martin N. Carrega P. Adnes C. (2013). Downscaling statistique des températures et des précipitations 2071-2100 du modèle ALADIN-Climat pour les scénarios A1B, A2 et B1 dans les Alpes-Maritimes, *Colloque de l'Association Internationale de Climatologie 2013*, Benin.
- Thompson J. (2005). *Plant evolution in the Mediterranean*, Oxford university press.

Stratégies ouvertes d'intégration de règles métier pour la géomatique dans un système d'information de gestion

Mounir Touzani¹, Christophe Ponsard², Anne Laurent¹,
Thérèse Libourel^{1,3}, Joël Quinqueton⁴

1. Université de Montpellier (UM), LIRMM, France

touzani@lirmm.fr, laurent@lirmm.fr

2. CETIC - Centre de recherche, Gosselies, Belgique

christophe.ponsard@cetic.be

3. Espace-Dev (UM, UAG, UR, IRD), France

therese.libourel@umontpellier.fr

4. Université Paul-Valéry Montpellier 3 (UPVM), LIRMM, France

jq@lirmm.fr

RÉSUMÉ. L'information géographique est si présente dans notre quotidien que sa prise en compte dans les systèmes d'information (SI) devient incontournable. Ces SI doivent pouvoir intégrer des aspects spatiaux et temporels au sein de leurs règles métier. Notre objectif est donc de proposer un processus d'évolution d'un SI permettant d'identifier ce type de règles métier à travers un processus d'ingénierie des exigences (IE). Nous avons retenu la méthode KAOS pour mettre en œuvre notre proposition à travers une étude de cas. Nous montrons également comment déployer de telles règles dans les composants les plus appropriés en faveur d'une architecture ouverte.

ABSTRACT. Geographic information is so present in our daily lives that its inclusion in information systems (IS) becomes unavoidable. These IS must be able to integrate spatial and temporal aspects into their business rules. Our goal is therefore to provide an IS evolutionary process to identify this type of business rules through a requirements engineering process (IE). We chose the KAOS method to implement our proposal through a case study. We also show how to deploy such rules in the most appropriate components for an open architecture.

MOTS-CLÉS : Information Géographique, Règles Métier, Ingénierie des Exigences, Architecture ouverte, Méthode KAOS.

KEYWORDS: Geographic Information, Business Rules, Requirements Engineering, Open Architecture, KAOS method.

SAGEO'2016 - Nice, 6-9 décembre 2016

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Introduction

Dans un monde où les personnes et les objets sont de plus en plus connectés et localisés, l'information géographique (IG) devient omniprésente et indispensable pour l'émergence de nouvelles applications telles que les villes intelligentes et les usines du futur. Dès lors, de nombreux systèmes logiciels sont amenés à entretenir une relation de plus en plus étroite et précise avec le monde réel. L'ampleur de cette évolution est majeure puisqu'en termes de données, au moins 80% sont concernées par la dimension géographique¹.

L'intensité de la relation avec le monde physique couvre un spectre très large : les systèmes cyber-physiques sont un cas de figure extrême car ils sont connectés en permanence et nécessitent des solutions spécifiques. À l'opposé, les Systèmes d'Information (SI) de gestion doivent régulièrement être augmentés afin d'inclure des fonctions géomatiques spécifiques (Pornon, 2011). Cet article est ancré dans cette seconde optique d'évolution.

Dans cet article, nous examinons plus spécifiquement des stratégies ouvertes d'intégration du SI existant, c'est-à-dire exploitant des briques ouvertes en matière de données et/ou de services pour répondre à des besoins géomatiques. Le terme ouvert s'applique ici à la fois au type d'architecture, aux standards et à la disponibilité de composants libres. Ceci offre plusieurs avantages : une architecture modulaire, la possibilité de déployer des services spécifiques en interne ou en externe, ainsi qu'un risque moindre de verrouillage de la solution.

La démarche d'évolution proposée est centrée sur la notion de règle métier (*business rule en anglais*), définie comme une déclaration structurée de haut niveau, permettant de contraindre, contrôler et influencer un aspect du métier (Hay, 2000). En termes d'ingénierie des exigences (IE), une règle métier est une exigence qu'on désire pouvoir faire évoluer aisément et que l'architecture du SI doit clairement expliciter et isoler de la logique applicative (Halle, 2001). Nous considérons donc, par la suite, l'utilisation d'un composant BRMS (système de gestion de règle métier, en anglais : Business Rule Management System) permettant de spécifier, gérer et exécuter les règles métier (Diouf, 2007).

Nous proposons un processus d'évolution où les règles métier sont replacées dans une démarche de modélisation des exigences, afin de déterminer l'impact de l'introduction de nouvelles exigences spatiales et temporelles. Sur cette base, nous guidons notre analyse vers le choix de moteurs de règles adéquats permettant de mettre en place des règles métier modifiées. Nous illustrons notre

1. <http://www.esrifrance.fr/sig1.aspx>

démarche par plusieurs règles métier à caractère spatial et/ou temporel à identifier dans l'évolution d'un SI dans le contexte de la fusion de deux universités.

Cet article est structuré comme suit : la section 2 présente un état de l'art qui fait l'inventaire des principaux composants capables de soutenir des règles métier de type spatial et temporel. La section 3 décrit la méthodologie proposée afin d'identifier les règles et de concevoir une architecture adaptée pour leur mise en œuvre. La section 4 illustre son application sur plusieurs types de règles métier de notre étude de cas. Enfin, la section 5 conclut et dresse quelques perspectives pour la suite de nos travaux en la matière.

2. État de l'art

Cette section décrit des travaux sur lesquels notre approche est bâtie. Elle donne une synthèse des techniques de modélisation spatio-temporelle du point de vue des systèmes d'information géographiques (SIG) et de l'ingénierie des SI. Ensuite, la notion de règle métier sera approfondie en particulier au niveau des capacités des moteurs de règles à traiter les aspects spatiaux et temporels.

2.1. Approche spatio-temporelle (ST) dans les SIG

L'information géographique (IG) décrit un objet, un phénomène ou un processus du monde réel. Un objet peut être représenté par son nom, son type, sa forme, sa localisation, ses propriétés thématiques ou encore par les relations de proximité qu'il peut avoir avec d'autres objets (Becker, al, 1990). L'analyse spatio-temporelle intègre simultanément les notions d'espace et de temps. L'espace vise des informations de localisation et de relations spatiales entre les objets, de manière qualitative (topologique, de connectivité, de position relative) ou quantitative (distances) entre objets spatiaux (Egenhofer, Herring, 1990; Randell *et al.*, 1992; Ligozat, 1998; Clementini, 2009). La dimension temporelle, quant à elle, cible des événements se produisant soit à un instant ou à une période, soit à des changements intervenant sur plusieurs instants, ou encore sur plusieurs périodes (Allen, 1984; Frank, 1994). Des typologies complètes de concepts combinant à la fois le spatial et le temporel couvrent des phénomènes de déplacement et de transformation (Claramunt, Jiang, 2001; Mathian, Sanders, 2014).

La représentation d'un objet spatial ou temporel demande une modélisation spécifique. Des chercheurs ont proposé des formalismes de représentation par l'utilisation de pictogrammes spatiaux et temporels dans les modèles, tels que ceux affinés par (Bédard, Larrivée, 2008). Ils ont évolué depuis Modul-R puis PVL (Plugin for Visual Language) sur PictograF².

2. PictograF <http://pictograf.scg.ulaval.ca>

SAGEO'2016 – Nice, 6-9 décembre 2016

2.2. Ingénierie des exigences (IE) et contraintes spatio-temporelles

L'Ingénierie des Exigences (IE) a pour objet d'établir et de maintenir un accord avec les parties prenantes sur les exigences du système à construire (Lamsweerde, 2009). Durant ce processus, les experts métier capturent, structurent, documentent et valident les besoins fonctionnels et non-fonctionnels.

Parmi les méthodes d'IE existantes, les méthodes orientées buts se démarquent par des garanties de complétude, de précision et leur capacité de raisonnement (Lamsweerde, 2001). Notre travail s'appuie principalement sur les modèles de buts et d'objets de la méthode KAOS (Keep All Objectives Satisfied) (Lamsweerde, 2009).

Un **but** est un objectif à atteindre par le système à concevoir. En KAOS, un **modèle de buts** est composé de buts système progressivement décomposés en buts fils, opérationnels selon une logique "ET" (coopération de tous les fils) ou "OU" (alternative entre les fils), traduisant respectivement les conditions "tous les fils sont nécessaires" ou "un des fils est suffisant". La question du « POURQUOI » permet de remonter jusqu'au but père, alors que la question du « COMMENT » permet de raffiner le but père en une multitude de buts fils. Ce raffinement s'arrête jusqu'au dernier but fils contrôlable par un **agent**, qui a cette responsabilité et qui se traduit par une réponse à la question du « QUI ». Ces derniers buts correspondent finalement, soit à des **attentes** sur le comportement d'agents de l'environnement, soit à des **exigences** sur le logiciel ou agent système. Parmi ces exigences figurent les règles métier.

Les buts, les exigences et donc les règles métier sont explicités dans les termes du domaine et capturés dans un **modèle des objets** en ayant recours à un formalisme de diagramme de classe UML (Unified Modelling Language). La plupart des SI standards intègre déjà des propriétés de type temporel, notamment des propriétés invariantes dans le temps ("TOUJOURS") ou de progression ("SI <événement> ALORS <état cible atteint> DANS <délai>"³).

Les propriétés spatiales peuvent être exprimées selon des principes similaires, via des patrons de formulation et des logiques spécifiques, permettant d'exprimer des relations structurelles spatiales (inclusion, intersection, adjacence,...) telle que DE-9IM (Clementini *et al.*, 1994), ou ayant une dynamique temporelle (transformation, mouvement...). A titre d'exemple, la règle métier suivante intervient dans notre cas d'étude : AUCUN étudiant ne devra effectuer plus d'un déplacement ENTRE sites DURANT une journée afin d'assister aux cours auxquels il est inscrit. Les notations SIG notamment "PictograF" ont aussi été intégrées dans le cadre de notation graphique d'IE, afin de décorer des buts et des entités ayant un caractère spatial ou temporel. Une description plus détaillée est disponible dans (Touzani *et al.*, 2016).

3. Par convention, nous identifions des mots clefs spécifiques en majuscules.

2.3. Approche par règles métier

La mise en œuvre des règles métier s'appuie généralement sur un système de gestion de règles métier (BRMS, en anglais : Business Rules Management System). La figure 1 adaptée de l'originale ⁴, montre l'architecture d'OpenRules, un BRMS très représentatif (Jacob Feldman, 2015).

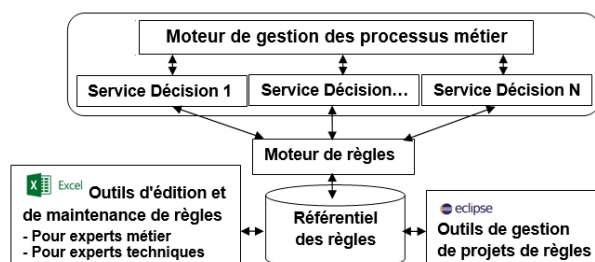


FIGURE 1. Structure d'un BRMS

- Les règles sont stockées indépendamment du code dans une *base de données de règles métier* qui peut être alimentée par des *outils d'édition*, facile d'accès comme des feuilles de calcul (table de décision) mais aussi des *outils de gestion* plus pointus (analyse des règles).

- Le *moteur d'exécution* des règles peut avoir des capacités d'inférence de nouveaux faits sur base de chaînage avant (déduction de faits ajoutés à la base de règles), le chaînage arrière (recherche de faits pouvant engendrer le fait à démontrer) ou mixte (combinaison des deux techniques précédentes).

- Une *couche de services* peut aussi être présente pour faciliter l'intégration du BRMS dans un environnement ouvert, notamment via des services web.

Les règles métier peuvent être définies de différentes manières :

- des règles simples de type « Si <condition> Alors <action> ». Ce type de règle a été décrit dans la section 2.2. Tous les BRMS supportent ce niveau dans divers formats textuels ou XML, notamment XACML qui représente un langage de règle de contrôle d'accès (OASIS, 2013).

- des tables de décision comprenant plusieurs colonnes de conditions et des actions. Celles-ci permettent de décrire et de comprendre des règles plus complexes. Beaucoup d'outils dont OpenRules ⁵ et Drools ⁶ proposent ce type de tables.

4. http://openrules.com/docs/man_4.html

5. OpenRules <http://openrules.com>

6. Drools <http://www.drools.org>

SAGEO'2016 – Nice, 6-9 décembre 2016

- des arbres de décision composés de nœuds de conditions, de branches et de feuilles comme actions. Ils peuvent être spécifiés graphiquement. Ce niveau de formalisme est supporté par JRules⁷.

La plupart des BRMS disposent de capacités de raisonnement temporel sur des événements et des durées. Certains BRMS ont développé des extensions pour traiter des caractéristiques spatiales, notamment :

- OpenRules a intégré la suite topologique Java (JTS) (Alex Karman, 2014) qui permet d'exprimer des règles intégrant des relations spatiales.
- Drools dispose d'un mécanisme de raisonnement spatial aisé à adapter et à mettre en œuvre (Merilina, 2014).
- GeoXACML est une extension d'XACML permettant d'exprimer des restrictions géographiques pour le contrôle d'accès IT (OGC, 2007).

2.4. Composants ou services SIG ouverts

La plupart des BRMS cités précédemment sont des outils libres, aisés à intégrer. D'autres composants peuvent leur venir en appoint dans un contexte spatial et/ou temporel pour offrir des fonctions d'accès aux données, de rendu cartographique, de calcul spatial, de géocodage ou de routage de véhicules. Ces fonctionnalités sont intégrées dans des "SIG d'entreprises" monolithiques mais sont également proposées par les solutions ouvertes suivantes de :

- données, rendu cartographique et géocodage. Google et Bing offrent des web-services commerciaux mais des services libres peuvent aussi être déployés sur base d'OpenStreetMap⁸. L'outil libre QGIS peut être utilisé comme serveur de données⁹. L'IGN propose aussi des données ouvertes¹⁰,
- calculs spatiaux. GEOS est la transposition en C++ de la librairie Java JTS déjà évoquée, elle-même utilisée par la librairie GDAL¹¹. Ces solutions matures aux licences permissives (LGPL, MIT) sont largement utilisées et peuvent également être intégrées directement par exemple pour rendre possible l'implémentation de règles simples dans un système développé en interne,
- routage de véhicules. Des services de calcul de chemins ou de matrices de distances sont disponibles sous un format libre. Certains sont très efficaces tels que OSRM¹² et Graphhopper¹³. OpenTripPlanner est spécialisé dans le transport multimodal¹⁴.

7. JRules http://www.ibm.com/support/knowledgecenter/SS6MTS__7.1.1

8. OpenStreetMap <https://www.openstreetmap.org>

9. QGIS <http://www.qgis.org>

10. IGN <http://www.ign.fr/institut/activites/lign-lopen-data>

11. GDAL <http://gdal.org>

12. OSRM - Open Source Routing Machine <http://project-osrm.org>

13. Graphhopper - Road Routing in Java with OpenStreetMaps <http://graphhopper.org>

14. OpenTripPlanner <http://www.opentripplanner.org>

3. Processus d'évolution

Dans cette contribution, nous proposons un processus d'évolution qui s'appuie sur la méthode d'IE *KAOS* décrite précédemment et qui se base sur une structure en trois étapes. Nous l'illustrons par la figure 2 contextualisée schématiquement à l'étude de cas pour aider à la compréhension du processus qui sera élaboré dans la section 4.

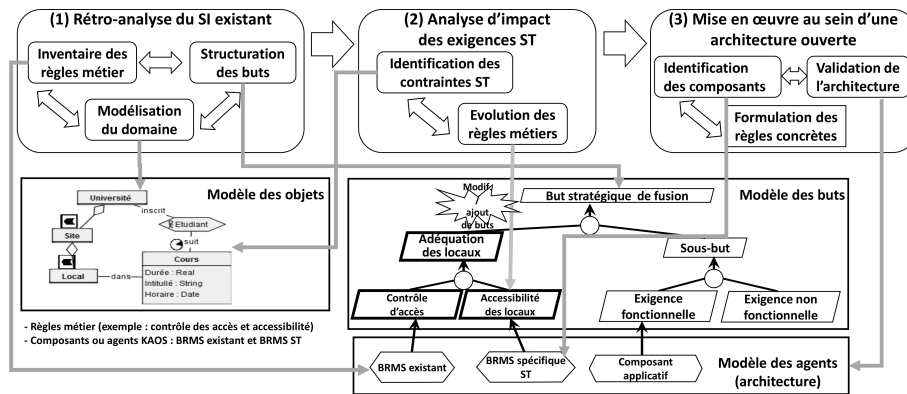


FIGURE 2. Processus d'évolution dirigé par les buts

Étape 1 - Rétro-analyse orientée buts du SI existant

Cette étape initiale vise à produire un modèle des buts et règles métier en lien avec un modèle du domaine. Les activités suivantes sont réalisées pour aboutir à un modèle cohérent et détaillé. Il s'appuie sur la documentation existante et de la rétro-ingénierie :

- structuration des buts selon une démarche « top-down ». Il s'agit de produire un modèle minimal pour comprendre le système à faire évoluer ;
- identification des règles métier existantes. Nous considérons ici une démarche « bottom-up » en partant de l'architecture et en remontant vers les buts ;
- modélisation du domaine. Un modèle de données abstrait est reconstruit sur la base d'objets métier, intervenant notamment dans la formulation des règles métier et des buts. Des objets ayant certaines caractéristiques spatiales ou temporelles peuvent déjà être annotés à ce stade.

Étape 2 - Analyse ST de l'évolution et adaptation des règles métier

Les nouvelles exigences sont identifiées et analysées pour leurs aspects spatio-temporels. Une analyse d'impact est ensuite réalisée sur :

- le modèle du domaine pour identifier les objets impactés et leurs nouvelles caractéristiques ;

SAGEO'2016 – Nice, 6-9 décembre 2016

– le modèle des buts pour identifier l'impact sur la structure de raffinement et *in fine* sur les règles métier en termes d'ajout et de modification.

Étape 3 - Mise en œuvre au sein d'une architecture ouverte

Afin de mettre en œuvre notre proposition d'évolution du système, voici les points considérés :

- identifier les composants de gestion des nouvelles règles métier, notamment sur base des briques potentielles décrites dans l'état de l'art ;
- valider la cohérence de l'architecture modifiée concernant les aspects non-fonctionnels de maintenabilité (limiter la diversité des composants) et de sécurité (en particulier par rapport à des services déployés en externe) ;
- traduire des règles métier dans le langage spécifique du moteur de règles et en particulier des nouvelles contraintes spatiales et temporelles.

4. Étude de cas : évolution d'un SI lors d'une fusion d'universités

Dans cette section, nous présentons l'étude de cas avant d'appliquer le processus d'évolution décrit à la section précédente.

Notre étude de cas s'inspire de la fusion des universités de Montpellier 1 et 2 (UM1 et UM2). Il s'agit d'un processus initié en 2012¹⁵ et qui a abouti à la naissance de la nouvelle Université de Montpellier (UM) en 2015. Les objectifs de la fusion sont de renforcer l'offre de formation, d'augmenter le potentiel de recherche et *in fine* d'accroître le rayonnement international de l'université.

Pour une raison d'abstraction et de simplification, nous considérons uniquement le site principal des deux organisations universitaires initialement indépendantes. Ces sites sont situés à une distance de l'ordre d'une dizaine de kilomètres qui est suffisamment faible pour permettre une mobilité mais suffisamment élevée pour nécessiter de repenser les affectations géographiques et la mobilité des étudiants et enseignants-chercheurs dans le contexte de la fusion.

Afin d'illustrer l'approche en détail, nous ne détaillerons que le but relatif à la mise en adéquation des locaux en termes de qualité, sécurité et accessibilité. Une analyse plus stratégique est décrite dans (Touzani *et al.*, 2016).

4.1. Etape 1 - rétro-analyse du but relatif à l'adéquation des locaux

Ce but présente des caractéristiques spatiales et temporelles importantes. La figure 3 décrit son raffinement jusqu'à l'identification de six exigences qui seront formulées en termes de règles métier. L'idée étant de pouvoir introduire un tel composant dans le cadre du processus d'évolution.

15. <http://www.lenouveaumontpellier.fr/nouvelle-universite-montpellier>

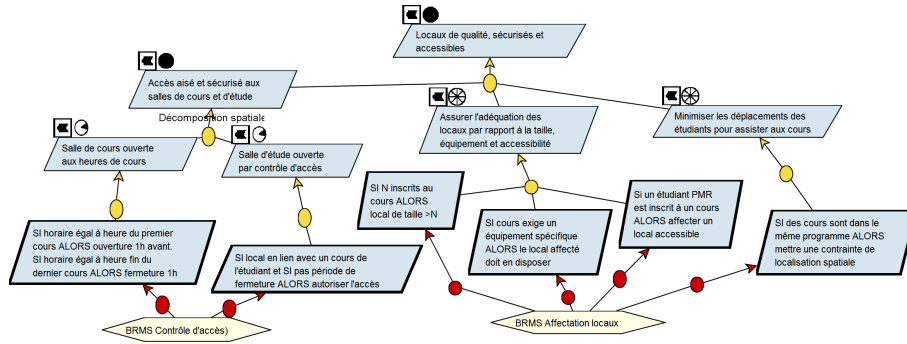


FIGURE 3. Raffinement des buts relatifs à la gestion des locaux

Les buts sont représentés par des parallélogrammes avec des bords fins et les règles métier par des parallélogrammes avec des bords épais. La formulation de ces dernières utilise une syntaxe classique « Si ... Alors ... ». Des décorations spatiales et temporelles étendant les notations PictograF. Elles sont décrites et utilisées dans notre article (Touzani *et al.*, 2016). Ainsi le but racine a une dimension spatiale bi-dimensionnelle (symbole \blacksquare) de gestion des locaux qui doit être assurée en permanence (symbole \bullet). Le but plus détaillé, relatif au contrôle d'accès est aussi de nature spatiale et comprend deux aspects complémentaires relatifs à des durées (durant/hors des périodes de cours). Quant aux buts de mise en adéquation relatifs à la taille et aux contraintes de déplacements, ils sont réexaminés de manière plus périodique (symbole \otimes).

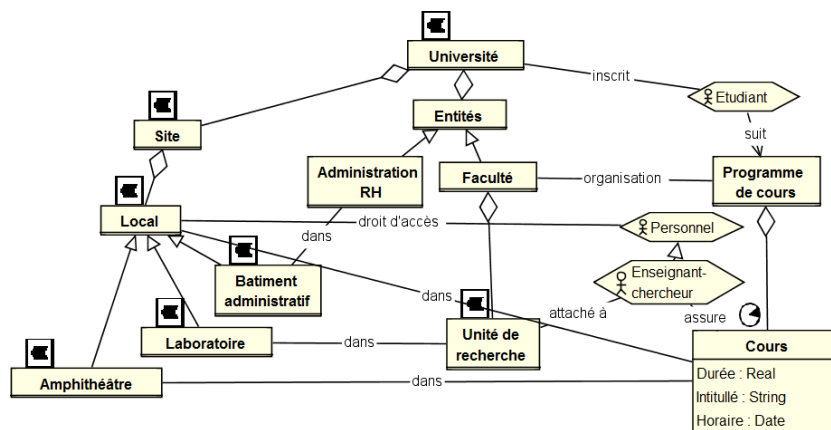


FIGURE 4. Modélisation du domaine (modèle des objets)

Le vocabulaire utilisé dans les descriptions des buts et des règles est formalisé dans la figure 4. Celle-ci décrit les aspects du domaine qui sont nécessaires

SAGEO'2016 – Nice, 6-9 décembre 2016

et suffisants à la description des buts et règles métier traités ici. Les notations étendues PictograF sont utilisées pour identifier les éléments ayant des dimensions spatiales (sites, bâtiments,...) et temporelles (cours).

Bien que les SI existants ne soient pas basés sur des BRMS, nous pouvons les illustrer dans un langage existant tel que le langage dédiée (DSL) de Drools afin de les exprimer plus précisément mais aussi de préparer les étapes suivantes de l'évolution. Nous détaillons deux des six règles identifiées :

– La règle « **SI** le local est en lien avec un cours d'étudiants et **SI** cela se déroule en période de travail **ALORS** autoriser l'accès » peut s'exprimer comme suit en utilisant une variable d'horloge et un test de compatibilité de la salle avec l'étudiant :

```
rule "Acces Conditionnel"
when
  $salle : Salle( $suj_salle=sujets , SessionClock() during periode_travail )
  $etudiant : Etudiant( actif == true , sujets intersect $suj_salle )
then
  $salle . autorisation( $etudiant );
```

– La règle : « **SI** un étudiant en situation de handicap (PMR) est inscrit à un cours **ALORS** affecter un local accessible » nécessite un composant de planification. La règle suivante peut être utilisée pour générer la contrainte qui exclura les salles inadaptées. La salle sera sélectionnée par le planificateur sur base de critères supplémentaires de taille et d'équipement.

```
rule "Accessibilite"
when
  $salle : Salle( accessibilite == false )
  $etudiant : Etudiant( PMR == true )
  $cours : Cours( inscrits contains $etudiant )
then
  $planificateur . exclusion( $cours , $salle );
```

4.2. Etape 2 - Analyse ST de l'évolution et adaptation des règles

La structure de fonctionnement des universités reste globalement similaire et permet d'appréhender la fusion au niveau des relations d'appartenance représentée par la figure 4. Les entités UM1 et UM2 disparaissent au profit de l'entité UM et les sites UM1 et UM2 deviennent des sites de l'UM. Les programmes de cours sont également unifiés même si l'offre de formation peut comporter une certaine redondance durant une période de transition.

La fusion a plusieurs conséquences sur les buts considérés ici : au niveau des cours, la refonte des programmes et les modifications de volumétrie impliquent de reconsidérer les affectations des locaux, avec la contrainte de limiter les déplacements des étudiants. Une règle serait de garder les étudiants sur le même site toute la journée. En cas d'impossibilité, une période de battement doit être prévue pour permettre le déplacement. D'autre part, une unification des contrôles d'accès est fortement recommandée.

4.2.1. Les besoins supplémentaires de déplacement

Ces besoins entre les différents sites peuvent être capturés par de nouvelles règles métier qui seront communiquées à l'outil semestriel de planification. Ce type de problème étant sur-contraint, il est également important de disposer d'informations sur l'importance relative d'une règle par rapport à une autre, ce qui est possible de faire via la table de décision spatiale décrite à la figure 5.

DecisionTableSpatial ContraintesAffectationsLocaux							
#	Main Type	Relationship	Related Type	Oper	Value	Conclusion	
						Spatial Significance	Score
0						=	0
1	CurrentLoc	!sameSite	NextLoc	Is	True	+=	10
2	CurrentLoc	distance	NextLoc	> 10	True	+=	2
3	CurrentLoc	distance	NextLoc	> 20	True	+=	5

FIGURE 5. Table de décision pour une localisation spatiale

Cette table est basée sur les conventions OpenRules et considère tous les couples d'emplacements successifs pour tous les programmes de cours afin de calculer une pénalité liée au non respect de la contrainte de localisation que l'outil de planification tente de minimiser. D'autres éléments calculés par d'autres règles déjà existantes peuvent entrer également en jeu, par exemple le fait d'allouer un local n'ayant pas les équipements requis, une salle en surcapacité, ou un local qui n'est pas facilement accessible aux étudiants en situation de handicap (règle détaillée à la section précédente).

Pour une meilleure optimisation, il est important de prendre en compte le temps de parcours au lieu de la distance, celui-ci pouvant varier en fonction de l'heure de la journée. Un service web peut être utilisé à cette fin, la table peut alors être adaptée pour spécifier des pénalités en fonction du temps de transport, puisque c'est la dimension temporelle qui compte.

Enfin, des services externes peuvent aussi alimenter le système avec des faits spécifiques, relatifs par exemple à l'accessibilité : il existe des initiatives collaboratives (crowdsourcing) telles que *j'accède*¹⁶ ou même plus spécifiques aux universités¹⁷.

4.2.2. L'unification du contrôle d'accès

Cette unification a pour but de mettre en place un contrôle d'accès avec un moyen d'authentification unique (nouvelles cartes d'accès ou reconnaissance de celles qui existent déjà) mais surtout qui unifie les règles d'accès tout en permettant des spécificités à un site géographique, soit pour des raisons transitoires, soit liées à des particularités telle qu'une sécurisation plus élevée de locaux contenant du matériel coûteux ou dangereux. Une fonction permettant

16. <https://www.jaccede.com>

17. <http://www.handi-u.fr>

SAGEO'2016 – Nice, 6-9 décembre 2016

d'identifier le site est nécessaire à cet effet. Par exemple, si le nettoyage à l'UM1 demande une ouverture des salles 1h avant le début des cours sur ce site, nous pourrions ajouter la condition suivante à la règle décrite précédemment :
`Salle(amphi == true, SessionClock() before[1h] ouverture, site == "UM1").`

4.3. Etape 3 - Mise en œuvre au sein d'une architecture ouverte

La figure 6 représente une architecture permettant de réaliser l'évolution des SI. Elle est centrée autour d'un composant BRMS unique qui interagit avec les systèmes existants de contrôle d'accès et les bases de données d'informations. Il a un lien avec deux composants ou services externes pour accéder à des données ouvertes sur l'accessibilité et un service en ligne de calcul de distance. Enfin, un outil de planification unifié est également redéployé.

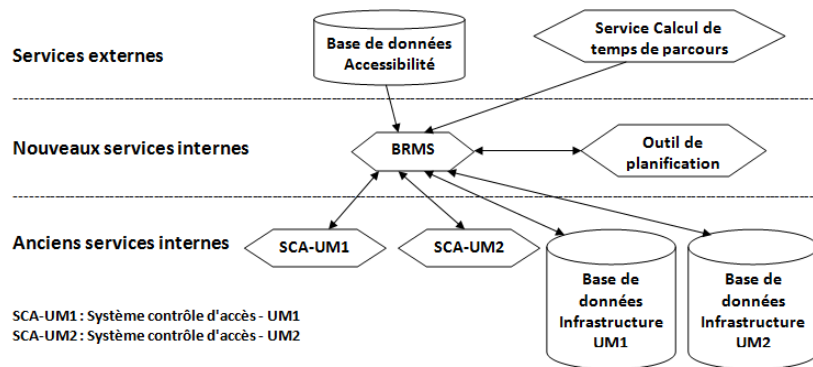


FIGURE 6. Nouvelle architecture ouverte

Concernant l'unification du contrôle d'accès, chaque université dispose déjà d'un système existant et pas forcément basé sur un BRMS. Une manière progressive consiste à définir une interface commune pour accéder aux données d'autorisation et au mécanisme de contrôle. Le BRMS pourra les exploiter pour un contrôle central, cependant à ce stade, la mise à jour restera locale. Par la suite, cette interface pourra devenir celle du système unifié, afin de permettre une plus grande unification des règles de contrôle d'accès et de faciliter une maintenance centralisée.

5. Conclusion et perspectives

Cet article présente une approche d'évolution de SI pour la prise en compte d'exigences émergentes à caractère spatial et en lien avec la dimension temporelle classiquement présente. Nous nous sommes appuyés sur une approche par règles métier. Sur base d'un état de l'art, nous avons identifié des composants capables de gérer et raisonner sur de telles règles dans un contexte

spatio-temporel mais aussi d’y adjoindre des services ouverts connexes et utiles pour élargir les missions d’un SI existant à certaines fonctions géomatiques. Afin de guider cette évolution, nous avons proposé une méthode orientée buts basée sur nos travaux antérieurs. Nous l’avons illustrée sur un cas d’étude ciblé et avons décrit des règles concrètes dans plusieurs BRMS populaires. Les principaux enseignements tirés à ce stade sont que :

- l’approche orientée buts permet de systématiser la collecte et la classification des règles métier ainsi que l’identification des règles implicites ;
- l’utilisation d’un BRMS est une approche intéressante dans le contexte de fusion, notamment pour gérer des variantes qui doivent coexister en phase de transition. Ceci serait beaucoup plus fastidieux avec des systèmes dits « codés » ;
- un BRMS permet de réaliser des raisonnements spatiaux et/ou temporels ;
- bien qu’il existe des BRMS dédiés à certains raisonnements, se limiter à un seul permet de centraliser la connaissance et faciliter sa maintenance.

De nombreuses perspectives sont ouvertes par ce travail, notamment dans la perspective de l’évolution vers des villes dites « intelligentes » (Smart Cities en anglais), tant du point de vue économique que social et environnemental. Dans nos actions futures, nous envisageons notamment d’affiner notre raisonnement spatio-temporel en proposant des ontologies spécifiques à ce domaine. Nous envisageons également d’adapter nos travaux aux phénomènes de fusion d’autres organisations, dans le domaine agricole par exemple. Au niveau technique, en fonction des besoins, nous envisageons également d’investiguer des composants plus spécialisés tels que le *Complex Event Processing* permettant le traitement réactif de flux de données et utilisé pour le traitement d’alertes, par exemple de vitesse ou de proximité dans les flottes de véhicules, ou encore des *systèmes experts* qui sont capables d’apporter une aide à la décision en reproduisant des mécanismes cognitifs plus complexes que les BRMS et en s’appuyant sur des bases de connaissances (Laurini, Thompson, 1992).

Bibliographie

- Alex Karman. (2014, Oct). *The Decision Table Template For Geospatial Business Rules*. Revolutionary Machines, DecisionCAMP.
- Allen J. F. (1984). Towards a general theory of action and time. *Artif. Intell.*, vol. 23, n° 2, p. 123–154.
- Becker R., al. (1990, Juillet). Network visualization. *4th International Symposium on Spatial Data Handling, Zurich, Switzerland*.
- Bédard Y., Larrivée S. (2008). Modeling with pictogrammic languages. *Shekar S, Xiong, H. (ed(s)), Encyclopedia of Geographic Information Sciences*, p. 716–725.
- Claramunt C., Jiang B. (2001). An integrated representation of spatial and temporal relationships between evolving regions. *Journal of Geogr. Syst.*, vol. 3, n° 4.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Clementini E. (2009). *A conceptual framework for modelling spatial relations*. Thèse de doctorat non publiée, EPFL, Lausanne.
- Clementini E., Sharma J., Egenhofer M. J. (1994). Modelling topological spatial relations: Strategies for query processing. *Computers & Graphics*, vol. 18.
- Diouf M. (2007). *Spécification et mise en oeuvre d'un formalisme de règles métier*. Thèse de doctorat non publiée.
- Egenhofer M. J., Herring J. R. (1990). A mathematical framework for the definition of topological relationships.
- Frank A. U. (1994). Qualitative temporal reasoning in gis - ordered time scales. *Proc. of the 6th Int. Symposium on Spatial Data Handling..*
- Halle B. von. (2001). *Business rules applied: Building better systems using the business rules approach* (1st éd.). Wiley Publishing.
- Hay D. (2000). *Defining business rules – what are they really. final report*. Final Report. Consulté sur http://www.businessrulesgroup.org/first_paper/br01c1.htm
- Jacob Feldman. (2015, Dec). *OpenRules - Business Rules and Decision Management System V6.3.4*. <http://openrules.com>.
- Lamsweerde A. van. (2001). Goal-Oriented Requirements Engineering: a Guided Tour. In *Proc. fifth ieee int. symposium on requirements engineerin*, p. 249-262.
- Lamsweerde A. van. (2009). *Requirements Engineering - From System Goals to UML Models to Software Specifications*. Wiley.
- Laurini R., Thompson D. (1992). *Fundamentals of geographic information systems*. Academic Press Limited. 0-12-438380-7.
- Ligozat G. (1998). Reasoning about cardinal directions. *Journal of Visual Languages and Computing*, vol. 9, n° 1, p. 23 - 44.
- Mathian H., Sanders L. (2014). *Objets géographiques et processus de changement*. London, ISTE.
- Merilina J. (2014). A mechanism to enable spatial reasoning in jboss drools. In *Int. Conf. on Industrial Automation, Information and Communication Technology*.
- OASIS. (2013). *eXtensible Access Control Markup Language (XACML) Version 3.0*. <http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html>.
- OGC. (2007). *GeoXACML Implementation Specification*. <http://www.opengeospatial.org/standards/geoxacml>.
- Pornon H. (2011). *SIG - La dimension géographique du système d'information*. Dunod. Consulté sur https://books.google.be/books?id=eM_OVMeOV1MC
- Randell D., Cui Z., Cohn A. (1992). A Spatial Logic Based on Regions and Connection. In *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*. San Mateo, Morgan Kaufmann.
- Touzani M., Ponsard C., Laurent A., Libourel T., Quinqueton J. (2016). Vers une modélisation et une analyse des exigences spatio-temporelles. In *Actes du 34e Congrès INFORSID, Grenoble, France, May 31 - June 3*, p. 51–66.

Towards Automated Image Geolocalization Refining

Bernard SEMAAN¹, Guillaume MOREAU¹,
Myriam SERVIERES¹, Bilal CHEBARO²

1. Ecole Centrale Nantes, UMR CNRS 1563 AAU

bernard.semaan,guillaume.moreau,myriam.servieres@ec-nantes.fr

2. Lebanese University

bchebaro@ul.edu.lb

RÉSUMÉ. Cet article décrit un système destiné à améliorer la géolocalisation des photos en utilisant plusieurs sources de données. Ce système regroupe plusieurs méthodes d'extraction et de traitement des données. La première étape est l'extraction d'une image, de ses méta-données et d'autres informations associées si elles sont disponibles. Ensuite, des informations géométriques et sémantiques sont extraites de l'image et mises en correspondance pour proposer une ou plusieurs poses de l'image. L'implémentation de la première partie de ce système est présentée.

ABSTRACT. In this paper we describe a system for refining image geolocalization using multiple sources of data. It combines multiple type of data retrieval and processing methods. Our system starts by retrieving the image, its meta-data and description when available. Then it proceeds with geometric and semantic information extraction. Finally it combines all these outputs and suggests one or many possible poses for the image. The implementation of the first part of this system is presented.

MOTS-CLÉS : SIG, Géolocalisation des images, Détection d'immeubles, Méta-données

KEYWORDS: GIS, Image geolocalization, Buildings detection, Meta-data

1. Introduction

Nowadays, city managers use more and more often a numeric duplicate of the city to manage it. As the city changes much faster than GIS updates can be made, GIS are always slightly outdated. At the same time, technology is rapidly evolving. People own new equipment capable of recording videos, pictures, geolocalize them, record time and other information. These information are

SAGEO'2016 - Nice, 6-9 décembre 2016

SAGEO'2016 – Nice, 6-9 décembre 2016

saved directly in the photo or video header. We assume using active sources of data such as (Picasa, Flickr, etc.) websites or specific institutional applications can help keeping a GIS up to date. Finally we can also use collaborative cartography as Open Street Map¹ in order to get buildings shape, roads and other cartographic data.

2. Geolocalization Refining Approaches

Several works already used feature extraction from images to retrieve their localization by matching them with a known GIS information. Other approaches use directly image comparisons or semantic information to retrieve the initial image localization. In (Bioret *et al.*, 2009), Bioret manually extracts the building facades limits and computes the angle and relation between them. The algorithm reconstructs the 2D geometric shape of the building's visible side. He finally compares it to the GIS map to estimate the picture pose (position and orientation of the camera). Other research teams work on texture similarity (Ventura *et al.*, 2014) in order to identify the represented building in the photo. First part is training the system to find all the SIFT descriptors in the existing database. Then they find descriptors in a query image. Finally the system compares all the descriptors in order to find the location of the photo. Images and video streams contain a lot of semantic information that we cannot directly compare to a GIS map or to existing geolocated textures. Text detection (Neumann, Matas, 2012) in photos may reveal a road name, place name we are trying to find. Other semantic information that may be extracted from a photo in an urban environment are road signs (Iii, 2013), shops logos (e.g. McDonalds), prominent building (e.g. Eiffel tower), etc.

3. Image Geolocalization System Proposal

Automating GIS update is a complex process. To facilitate this process we present our proposal of retrieving images from online social media databases and processing them to find their localization which will allow to use them for GIS updates in urban zone. This system can be divided to three layers: Data retrieval and preprocessing, Features extraction, Decision making. In Figure 1, we present the main modules of our system.

Data Retrieval and Preprocessing We start by retrieving images from a local or an online database regarding users needs (Images shooting zone, images from social media or taken by the users, etc.). The aim is to return a list of meta-data and Geo-located images that may contain buildings. Then we proceed in extracting and processing image meta-data. Some additional data may exist in social media websites like comments, description, tags that

1. <http://www.openstreetmap.org>

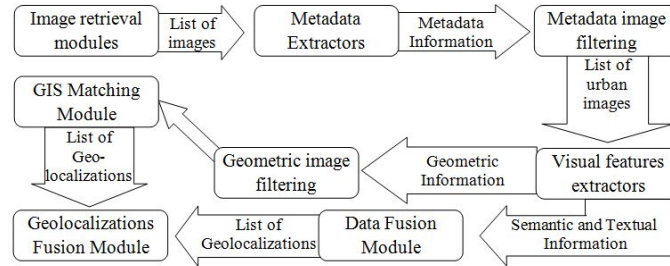


Figure 1. Our Image Geolocalization System Proposal

may include geographic and temporal information. Then we dismiss images that do not belong to an urban zone, are outdated or simply are missing their meta-data information.

Features Extraction This layer extracts basic image features by using image processing algorithms. Some possible features to extract are semantic (text detection, road signs detection, known logos detection, etc.) while others are geometric (vanishing points, building shape, image segmentation, etc.). Simple urban man made scenes are made of standing boxes like buildings, they can be represented with a group of segments and polygons. For this layer, we have created so far a method that uses only LSD (Grompone Von Gioi *et al.*, 2010) segments and vanishing points (Kalantari *et al.*, 2009) to find buildings edges. We start by assembling segments with close slopes and small separation distance. Assuming that one and only one big building exists in the image, its edges should be defined by the horizontal segments on the top of the image thus we compare the segments two by two and keep the higher ones. The remaining edge segments are then joined to find the contour polygon of a building. Finally we add missing vertical segments. Our building edges detection method is robust against textures change, weather conditions and some occlusion problems. On the other side our method would not work with complex architectures containing arcs and tilted walls. Figure 2 shows some results.

Decision Making We use the result of preceding layers, and the imprecise location we already have as an input for the implementation of (Bioret *et al.*, 2009) that will return a list of possible camera poses. On the other hand we use semantic extracted features to find another list of geolocalizations based on street names, shops and points of interest that may appear in the image. Finally, we will compare the various geolocalizations that we have collected from different sources and find the ones that occur the most, remove the incoherent and duplicate ones.

4. Limitations and inaccuracy management

As mentioned before, our proposal is a sequence of modules with different outputs. Some modules in this sequence, like the building detection one, may

SAGEO'2016 – Nice, 6-9 décembre 2016



Figure 2. Automatic buildings detection results.

affect the whole results. In fact a wrong building shape detection wrecks the GIS matching process. Other modules, like the text detection one, will only provide additional and optional information. Text, for example, may not be available in every image, when available it may not always be helpful. Thus the decision making layer will give more weight to the GIS matching module localisation than to locations based on detecting text in images. This will be managed using importance factors in the final decision equation.

5. Conclusion

In this article we have introduced a system that helps finding the most accurate geolocalization and camera pose of an image. We have also briefly presented image segmentation method to detect a building in our images and we have presented some results of its detection. The advantage of such system is the possibility of combining information from several sources and from different forms. This system uses existing techniques and provides the possibility of introducing new ones.

Bibliographie

- Bioret N., Moreau G., Servières M. (2009). Pose computation based on correspondence between 2d gis and street images. In *Intelligent systems and automation: 2nd mediterranean conference on intelligent systems and automation (cisa'09)*, vol. 1107, p. 169–174.
- Grompone Von Gioi R., Jakubowicz J., Morel J. M., Randall G. (2010). LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, p. 722–732.
- W. G. (2013). Combinatorial clustering and its application to 3d polygonal traffic sign reconstruction from multiple images. , p. 1–8.
- Kalantari M., Jung F., Guedon J. (2009). Precise, automatic and fast method for vanishing point detection. *The Photogrammetric Record*, vol. 24, p. 246–263.
- Neumann L., Matas J. (2012). Real-time scene text localization and recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 3538–3545.
- Ventura J., Arth C., Reitmayr G., Schmalstieg D. (2014). Global Localization from Monocular SLAM on a Mobile Phone. *Visualization and Computer Graphics, IEEE Transactions*, vol. 20.4, p. 531–539.

Une application SIG pour l'optimisation de la collecte des déchets post-inondation

Serge Lhomme¹, Mickael Brasebin², Arnaud Le Guilcher², Oumayma Kaabi^{1,2}, Sofiane Martel², Bruno Barroca³

1. Université Paris-Est Créteil (UPEC), Lab'Urba (EA 3482)

61 avenue du Général de Gaulle - 94010 Créteil - France

serge.lhomme@u-pec.fr

2. Université Paris-Est (UPEMLV), COGIT, LASTIG, IGN

73 avenue de Paris - 94160 Saint-Mandé - France

mickael.brasedin@ign.fr ; arnaud.le-guilcher@ign.fr

3. Université Paris-Est Marne-la-Vallée (UPEM), Lab'Urba (EA 3482)

Cité Descartes - 5 boulevard Descartes - 77454 Marne la Vallée - France

bruno.barroca@u-pem.fr

RESUME. Pour faciliter la reprise d'un fonctionnement normal à la suite d'une inondation, les collectivités territoriales doivent notamment faire face aux déchets produits par celle-ci. Dans ce contexte, le principal objectif du projet SIGOPT est de produire une extension à un Système d'Information Géographique permettant d'optimiser la collecte des déchets générés par une crue. Pour disposer d'une évaluation de la quantité de ces déchets, une extension SIG de la méthode Mécadépi a été développée. Pour modéliser la problématique d'optimisation, les modèles de localisation-allocation et de tournées de véhicules ont été articulés selon une logique à deux niveaux. Pour faire face aux nombreuses incertitudes, il est proposé d'adapter les algorithmes traditionnels d'optimisation à un cadre stochastique. Cette recherche a été appliquée au territoire francilien.

ABSTRACT. In order to recover quickly after a flood, local authorities have to deal with flood waste. In this context, the main objective of the SIGOPT project is to develop a Geographic Information System plugin for optimizing flood waste collection. A plugin, based on Mécadépi method, has been developed to quantify flood waste. Then, to model the optimization problem, location-allocation models and vehicle routing models have been articulated following a two-level logic. To deal with uncertainties, it is proposed to adapt traditional optimization algorithms to a stochastic framework. This research is applied to the Paris region.

MOTS-CLES : SIG ; Inondation ; Résilience ; Optimisation spatiale ; Modèles localisation-allocation ; Modèles de tournées de véhicules.

KEYWORDS: GIS; Flood risk; Resilience; Spatial optimization; location-allocation models; Vehicle routing problem.

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Contexte de la recherche

Pour faciliter la reprise d'un fonctionnement normal à la suite d'une inondation et ainsi contribuer à la résilience d'un territoire, les collectivités territoriales doivent notamment faire face aux déchets produits par celle-ci. En effet, les volumes de déchets produits par une inondation peuvent être considérables. A titre d'exemple, les déchets générés par l'inondation consécutive à la tempête Xynthia ont représenté l'équivalent de 12 années de collecte « normale » pour le territoire touché (Beraud, 2013). Dans les faits, lors d'une inondation, l'eau dégrade tout ce qu'elle touche, générant par là même des déchets en quantité très importante. De surcroît, ces déchets mouillés ne sont pas des déchets « habituels », ce qui accroît la difficulté de la gestion et peut retarder la collecte.

Ces déchets soulèvent alors deux problèmes majeurs. Premièrement, s'ils ne sont pas collectés rapidement, ces déchets engendrent des nuisances importantes : dégradation de l'environnement, impacts sanitaires... Deuxièmement, les déchets non collectés, qui encombreront les axes routiers, peuvent gêner les secours et par conséquent compliquer la gestion de crise et le retour à la normale. Le maintien d'un service de gestion des déchets est donc stratégique pour la résilience des territoires face aux catastrophes. Dans ce contexte, peu de temps après une inondation, des questions d'optimisation se posent : Quels sites potentiels choisir pour entreposer les déchets afin de minimiser les nuisances environnementales ? Comment organiser la collecte des déchets ? Comment optimiser les tournées des véhicules de collecte ? Ces problèmes d'optimisation ne sont pas triviaux à formuler et à résoudre.

Plus spécifiquement, à Paris et en Ile-de-France, le coût d'une crue de la Seine équivalente à celle de 1910 apparaît difficilement estimable. Ainsi, les estimations existantes varient du simple au sextuple : de 5 à 30 milliards d'euros (Reghezza, 2015). Quoiqu'il en soit, ce coût est sans commune mesure avec ceux estimés pour d'autres territoires français. Si en France les inondations peuvent être considérées comme le risque naturel majeur, les coûts induits par les inondations étant les plus importants, une inondation de type 1910 en région parisienne peut dès lors apparaître comme l'aléa naturel le plus redoutable et la collecte des déchets générés par celui-ci peut être considérée comme la plus complexe.

Dans ce contexte, le principal objectif du projet SIGOPT (<https://github.com/IGNF/SIGOPT>) est de produire une extension à un Système d'Information Géographique (SIG) permettant d'optimiser la collecte des déchets générés par une crue et de l'appliquer au territoire francilien (plus précisément Paris et les départements de la petite couronne). Pour cela, il s'est tout d'abord révélé nécessaire de caractériser et de quantifier les déchets pouvant être produits par une inondation. Il faut préciser ici que l'application de cette recherche au territoire francilien s'est notamment appuyée sur des échanges avec les gestionnaires franciliens qui ont ainsi contribué à l'orienter.

2. Implémentation SIG d'une méthode permettant de caractériser et de quantifier les déchets générés par une inondation

Les recherches sur les impacts induits par des inondations se sont longtemps focalisées sur l'évaluation de leur coût économique. Ces recherches prennent peu en compte l'organisation de l'après crise, et le redémarrage des territoires (Béraud, 2013). Par exemple, concernant la gestion des déchets générés par l'inondation, le manque de méthodes d'évaluation (tenant compte de l'amplitude de l'inondation et de la vulnérabilité des enjeux) empêche les acteurs publics d'anticiper le problème et de mettre en place des plans de gestion efficaces. C'est pourquoi, une méthode facile d'utilisation permettant de quantifier et de qualifier les déchets des ménages qui pourraient être produits par une inondation sur un territoire a été développée très récemment pour le territoire français : Mécadépi (Béraud et al., 2012). La méthode Mécadépi permet ainsi d'estimer et de caractériser les déchets des ménages, mais ne s'attache pas aux autres déchets du territoire.

L'application de cette méthode sur un territoire requiert l'utilisation d'un SIG, car cette méthode repose sur des données géographiques facilement mobilisables qu'il convient de croiser : notamment la BDTOPO de l'IGN (plus précisément la couche des bâtiments indifférenciés) et l'emprise de la zone inondable (issue par exemple de la base de données Géorisque). La principale difficulté, pour appliquer cette méthode, réside alors dans le calcul du nombre de logements situés en zone inondable. En effet, Mécadépi fournit les coefficients qui permettent à partir de ce nombre de logements d'évaluer de manière sommaire et assez simple les quantités de différents types de déchets ménagers (ameublement, équipements électroniques et électriques, véhicules...), mais la méthode proposée pour évaluer le nombre de logements inondés peut apparaître assez contraignante.

Afin de faciliter l'implémentation de cette méthode par les collectivités territoriales et de disposer d'une évaluation de la quantité des déchets générés par une crue de la Seine sur le territoire étudié, un des objectifs du projet SIGOPT a été de créer une extension de cette méthode à un SIG libre. Les données nécessaires en entrée sont : une couche vectorielle des bâtiments, une couche vectorielle des IRIS, une couche vectorielle de la zone inondable et une table attributaire issue d'une enquête logement de l'INSEE. En sortie, les résultats sont stockés au sein de la couche des bâtiments et de la couche des IRIS sous la forme de champs qui correspondent aux différents types de déchets quantifiés. Ce stockage des résultats à une échelle très fine comme celle des bâtiments, s'il est très critiquable compte-tenu des nombreuses incertitudes présentes à cette échelle, permet d'obtenir facilement, par agrégation, des résultats à différentes échelles : îlots (Fig. 1), communes, intercommunalités, départements... Les résultats obtenus au niveau des bâtiments peuvent même être projetés sur les axes routiers, ce qui se révèle très pertinent dans le cadre de cette recherche.

SAGEO'2016 – Nice, 6-9 décembre 2016

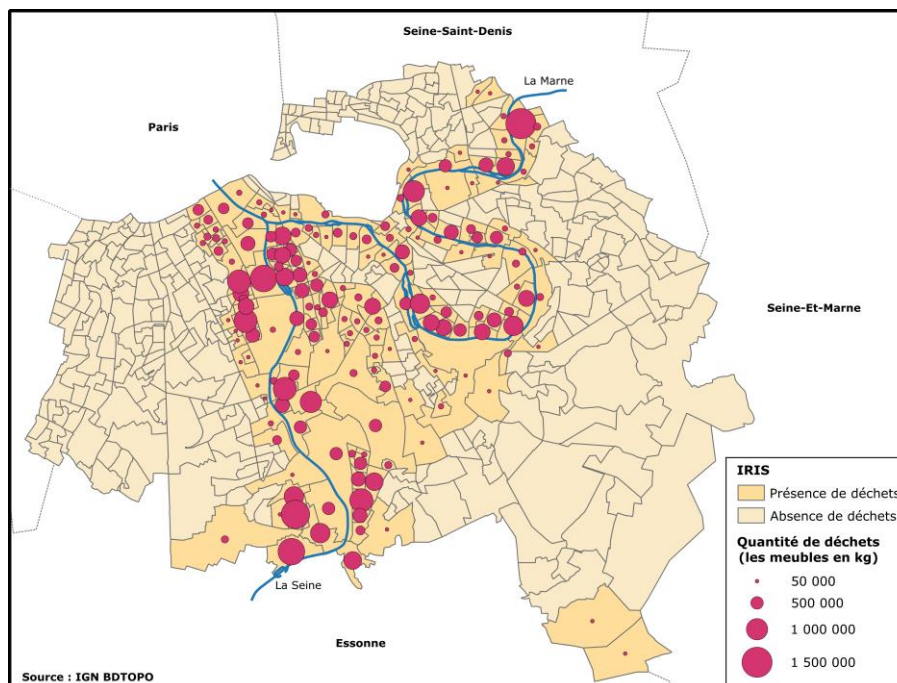


FIGURE 1. Cartographie des quantités de déchets (les meubles en kg) par IRIS dans le département du Val-de-Marne.

3. Optimisation de la collecte des déchets post-cruce

Une fois les quantités de déchets quantifiées, l'objectif de cette recherche est de formaliser de manière mathématique le problème de la collecte des déchets post-cruce afin de minimiser le temps de collecte tout en respectant des contraintes économiques, matérielles et humaines. Dans les faits, ce problème peut être rattaché à deux familles de problèmes classiques en recherche opérationnelle : les problèmes de localisation-allocation (Revelle et al., 1977) et les problèmes de tournées de véhicules (Dantzig et Ramser, 1959). Ces problèmes font depuis leur origine l'objet de travaux réguliers, ils sont donc nombreux, mais dans le cadre précis de la gestion des déchets post-cruce il existe peu de travaux, comme l'atteste un état de l'art d'une soixantaine d'articles sur l'optimisation spatiale et les SIG (Méneroux et al., 2015).

Les deux familles de problèmes d'optimisation abordées ici peuvent être modélisées sous la forme d'un graphe caractérisé par deux types de sommets : les points de collecte, caractérisés par une quantité de déchets à collecter, et les centres de stockage, caractérisés par une certaine capacité de stockage (Fig. 2). Les localisations envisagées des centres de stockage sont déterminées par les gestionnaires en charge de la collecte. Ce faisant, il est proposé dans cette recherche

d'articuler les modèles de localisation-allocation et de tournées de véhicules selon une logique à deux niveaux.

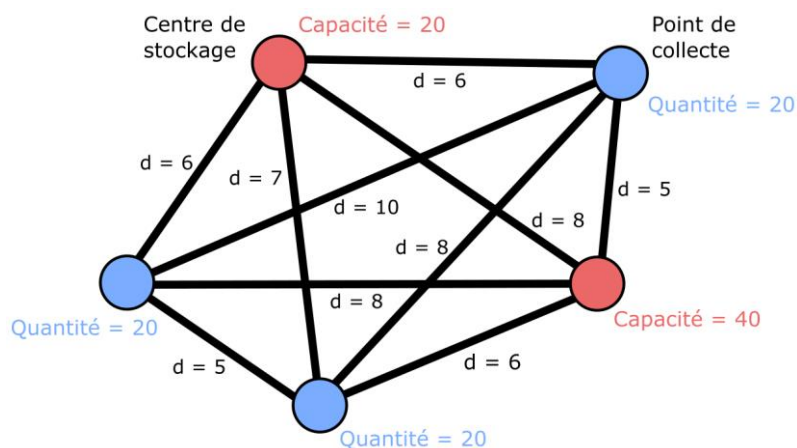


FIGURE 2. Représentation graphique simplifiée de la modélisation sous forme de graphe des problèmes d'optimisation abordés.

Ainsi, dans un premier temps, la résolution d'un modèle simple de localisation-allocation permet de minimiser le temps de collecte en affectant au mieux les points de collecte aux centres de stockage pour l'ensemble du territoire étudié. Parallèlement, comme les centres de stockage envisagés sont généralement des centres ouverts exceptionnellement pour faire face aux déchets générés par la crue, ce qui pose des problèmes politiques majeurs (les riverains de ces centres subissant des nuisances inhabituelles), il convient d'utiliser un deuxième modèle afin de déterminer s'il est possible de ne pas ouvrir tous les centres de stockage initialement prévus, sans pour autant excessivement dégrader le temps de collecte. Dans un deuxième temps, le modèle de tournées de véhicules est appliqué indépendamment pour chaque centre de stockage afin de minimiser le temps de collecte au sein des centres de stockage et d'organiser plus précisément la collecte sur un territoire restreint à un « bassin de collecte ».

Pour modéliser la collecte des déchets post-inondation, plusieurs modes de collecte ont été envisagés. Le mode de collecte le plus simple correspond à une collecte indifférenciée des déchets (tous les déchets sont mélangés et rien n'est trié) qui doivent être ensuite entreposés dans des points de transfert (où les déchets seront triés voire traités). Un deuxième mode de collecte est une collecte différenciée des déchets (ce mode de collecte permet d'effectuer un premier tri des déchets) avec des points de transfert eux aussi différenciés. Dans les faits, les gestionnaires semblent privilégier le premier mode de collecte. C'est donc sur ce mode de collecte que se fonde les modélisations effectuées, où les points de transfert sont considérés comme des centres de stockage.

SAGEO'2016 – Nice, 6-9 décembre 2016

Enfin, différents algorithmes de résolution ont été envisagés pour résoudre ces problèmes d'optimisation : des heuristiques relativement simples (algorithme glouton, algorithme de voisinage...) et des métaheuristiques plus complexes (recuit simulé, algorithme génétique, GRASP...). La métaheuristique du recuit simulé a finalement été retenue, car elle a la propriété d'être bien adaptée à la résolution des problèmes de localisation-allocation et des problèmes de tournées de véhicules (Kokubugata et al., 2008 ; Jiang et al., 2015 ; Wang et al., 2015).

La conception de telles méthodes d'optimisation se heurte cependant à une difficulté notable dans la réalité qui est l'incertitude sur les paramètres du modèle : les quantités de déchets évaluées par les méthodes de type Mécadépi sont soumises à une certaine variabilité et, dans un contexte d'inondations, la praticabilité de certains tronçons de route n'est pas acquise. Il convient alors d'adapter les algorithmes traditionnels d'optimisation à un cadre stochastique. Pour cela, à partir d'un modèle stochastique convenable, une approche de « quantification des incertitudes », à l'origine développée pour des modèles mécaniques (Sudret, 2007), est envisagée. Une telle approche permet de suivre les incertitudes à travers le processus d'optimisation et de rendre plus efficaces les prises de décision en environnement incertain.

4. Perspectives

Les collectivités territoriales sont confrontées à de nombreux problèmes d'optimisation qu'elles ne savent pas nécessairement résoudre. Le cas de la collecte des déchets post-inondation est un bon exemple. Un des défis sous-jacents à ce projet est donc d'identifier les modes de formulation géographique les plus pertinents pour « traduire » des problèmes mathématiques d'optimisation que seuls des mathématiciens ou des informaticiens peuvent résoudre à l'aide de logiciels de calculs numériques. A terme, relever ce défi permettrait à certains utilisateurs SIG de résoudre une grande diversité de problèmes d'optimisation spatiale.

Remerciements

Cette recherche a été financée dans le cadre d'un projet PEPS de l'université Paris-Est. Le projet SIGOPT s'inscrit ainsi dans le cadre d'une recherche exploratoire visant plus généralement à faciliter la formalisation et la résolution de problèmes d'optimisation spatiale au sein de Systèmes d'Information Géographique.

Références

Béraud H. (2013). *Initier la résilience du service de gestion des déchets aux catastrophes naturelles : le cas des territoires urbains et de l'inondation*. PhD thesis, Université Paris-Est Mane-La-Vallée.

- Beraud H., Jadot J., Barroca B., Hubert G., Bauduceau N., (2012). *Mécadépi. « Méthode d'Evaluation et CARactérisation des DEchets Post Inondation »*. Rapport final, Université Paris Est Marne-la-Vallée et CEPRI, 133 p.
- Dantzig G. B., Ramser J. H., (1959). The Truck Dispatching Problem. *Management Science*, Vol. 6, pp. 80-91.
- Jiang Q., Wang C., Mu D., Zhou L. (2015). A hybrid metaheuristic algorithm for the VRPSD problem. *Journal of Computational Information Systems*, Vol. 11, n°13, pp. 4845–4856.
- Kokubugata H., Kawashima H. (2008). Application of Simulated Annealing to Routing Problems in City Logistics. In *Simulated Annealing*, Cher Ming Tan (Ed.), ISBN: 978-953-7619-07-7, pp.131–154.
- Méneroux Y., Brasebin M., Le Guilcher A., (2015). *Revue de l'état de l'art : SIG et Optimisation spatiale*, document de travail du projet SIGOPT, Laboratoire COGIT, 25 p.
- Reghezza M. (2015). *De l'avènement du Monde à celui de la planète : le basculement de la société du risque à la société de l'incertitude*. HDR, Université Paris 1- Panthéon Sorbonne.
- ReVelle C. S., Bigman D., Schilling D., Cohon J., Churuch R., (1977). Facility Location: A Review of Context-free and EMS Models. *Health Services Research*, Vol. 12, n°2, pp. 129-146.
- Sudret B., (2007). *Uncertainty propagation and sensitivity analysis in mechanical models - Contributions to structural reliability and stochastic spectral methods*. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand, 252 p.
- Wang C., Mu D., Zhao F., Sutherland J. W., (2015). A parallel simulated annealing method for the vehicle routing problem with simultaneous pickup–delivery and time windows. *Computers & Industrial Engineering*, Vol. 83, pp. 111–122.

Une approche semi-automatique pour l'extraction d'informations liées aux itinéraires culturels à partir des réseaux sociaux : cas de la Via Francigena

Nathalie Valmond-Leblanc¹, Eric Kergosien², Natalia Grabar¹, Marta Severo²

1. STL, Université de Lille

Domaine Universitaire Point de bois, BP 60149, 59653, Villeneuve d'Ascq, France
prenom.nom@univ-lille3.fr

2. GERiCO, Université de Lille

Domaine Universitaire Point de bois, BP 60149, 59653, Villeneuve d'Ascq, France
prenom.nom@univ-lille3.fr

RESUME. Dans le cadre de l'analyse des itinéraires culturels, objet complexe du point de vue territorial, symbolique et social, le projet Itinéraire Numérique s'attache à analyser les contenus numériques disponibles en ligne pour la valorisation de l'itinéraire Via Francigena. Dans cet article, nous proposons une méthodologie semi-automatique pour l'extraction d'information spatiale liées à la Via Francigena à partir des données Instagram.

ABSTRACT. As part of the analysis of cultural routes, complex object of the symbolic, social and territorial point of view, the project Digital Route focuses on the analysis of digital content available online for the development of the Via Francigena route. In this article, we propose a semi-automatic methodology for the extraction of spatial information related to Via Francigena from Instagram data.

MOTS-CLES : Itinéraire culturel, Fouille de textes, Information spatiale, réseaux sociaux.

KEYWORDS: Cultural route, Text Mining, Spatial information, Social networks

1. Introduction

Le fait que le patrimoine culturel soit disponible au format numérique modifie la manière de le comprendre et de le représenter. En particulier, cette abondance des données numériques a produit une nouvelle effervescence vers les possibilités de représentation du patrimoine. Des projets comme la plateforme de l'Institut Culturel de Google, le projet ORBIS de Stanford (Stanford Geospatial Network Model of the Roman World), ou les différents projets Europeana constituent de bons exemples d'une telle effervescence. L'objectif du projet « Itinéraire Numérique »¹ est de fédérer une approche interdisciplinaire pour étudier les modalités de représentation numérique du patrimoine culturel lié à l'itinéraire Via Francigena. Dans cet article nous décrivons une méthode semi-automatique pour extraire les entités spatiales (ES) à partir de messages Instagram. L'objectif à terme est de pouvoir proposer au grand public une interface dynamique permettant de découvrir la Via Francigena à travers les informations diffusées par des utilisateurs d'instagram (commentaires, photos, etc.). Nous traitons dans ces travaux des ES dites simples, aussi nommées entités spatiales absolues (ESA), à la différence des entités spatiales plus complexes intégrant des relations spatiales appelées entités spatiales relatives (ESR) (Lesbegueries et al., 2006). La méthode combine une approche morpho-syntaxique à une approche lexicale s'appuyant sur la liste des étapes de l'itinéraire et sur Wikipedia pour identifier les ES dans les messages Instagram. En section 2, nous présentons un rapide état des travaux pour l'extraction des ES à partir de données de réseaux sociaux. En section 3, nous présentons notre méthodologie et les premiers résultats obtenus. Nous concluons en section 4 en précisant les perspectives.

2. Travaux connexes

Le travail de Pospescu et al. (Popescu et al., 2009) a pour objectif l'extraction automatique d'informations touristiques à partir de photos de voyages Flickr. Ils s'intéressent plus particulièrement à ce que les gens visitent, la durée de leur visite et les photos panoramiques. L'extraction des ES contenues dans les titres et les tags s'effectue selon une approche lexicale qui se base sur la construction d'un gazetier à partir de Wikipédia et qui associe à chaque nom de lieu ses coordonnées géographiques et son type. (Mohamed et al., 2014) présente une recherche similaire en se focalisant surtout sur la catégorisation des entités nommées (personnes, lieux et organisations) en se basant sur les Infobox de Wikipédia. D'autres méthodes permettant l'extraction d'ES telles que la méthode TEXT2GEO (Tahrat et al., 2013) qui présente une approche hybride combinant des patterns linguistiques avec une approche d'apprentissage supervisé. La combinaison des deux méthodes améliore significativement la précision et le rappel des résultats. Cependant, le principal inconvénient de la méthode d'apprentissage supervisée réside dans l'exigence de données annotées manuellement, ce qui est plutôt une tâche coûteuse et complexe.

3. Travaux réalisés

Le corpus de départ est un ensemble de 7034 posts multilingues (italien, anglais et français), publiés entre septembre 2011 et janvier 2016, extraits du réseau social Instagram car mentionnant le tag « Via Francigena ». La méthodologie que nous

¹ <http://www.itinerairesculturels.fr>

SAGEO'2016 – Nice, 6-9 décembre 2016

appliquons dans nos travaux se décompose en 2 étapes : Extraction des étapes de la Via Francigena et des ES à partir de la ressource Wikipedia ; Désambiguïsation des ES.

Pour l'**extraction des ES**, nous avons combiné une méthode morpho-syntaxique et une approche lexicale s'appuyant sur la liste des étapes de la Via Francigena et la liste d'ES générée à partir d'articles Wikipédia. Pour cela, nous avons sélectionné dans Wikipedia les articles dont les lieux sont géocalisés dans les régions que traverse la Via Francigena. Un article contenant différentes métadonnées (commune, région, province, coordonnées, description, lien de la page) est attaché à chaque ES de la liste obtenue. Le marquage du corpus Instagram avec la liste des étapes et les données Wikipédia permet d'obtenir une liste d'ES candidates à décrire la Via Francigena. Pour chaque ES candidate, une (voir deux) étiquette est associée pour préciser son type en nous appuyant sur la description et sur le nom de l'article provenant de Wikipedia. Nous avons défini pour cela plusieurs motifs s'appuyant sur des lexiques, afin d'identifier cinq types distincts (étapes, monuments, lieux religieux, lieux naturels, autres lieux).

La **désambiguïsation** s'applique aux ES identifiées à partir de la liste Wikipédia qui ne correspondent pas à une étape de l'itinéraire. Nous retenons une ES si elle est localisée dans une région, province ou commune par laquelle passe la Via Francigena. Les coordonnées géographiques extraites des données Wikipédia nous permettent de calculer la distance entre un lieu et une étape. Au regard des expérimentations menées sur le corpus visant à identifier la distance adéquate entre les étapes de la Via Francigena et les ES candidates (5, 10, 15, 20, 30, 50, 100km), nous retenons les ES qui se trouvent à moins de 15 km d'une étape. Par exemple, l'ES « passo della cisa », identifiée dans 3 posts et caractérisée dans Wikipedia comme appartenant à une région par laquelle passe la Via Francigena, est identifiée comme pertinente car localisée à 6km de l'étape Berceto.

Des **analyses statistiques** résultent de ces premières expérimentations. Tout d'abord, à partir des 7034 posts de départ, 2598 contiennent une ou plusieurs ES pertinente(s). Aussi, 6186 occurrences d'ES sont identifiées, à savoir 3874 pour les étapes et 2312 pour les autres ES pertinentes. Les figures 1 et 2 mettent en avant les ES les plus fréquentes, respectivement les étapes et les ES provenant de Wikipedia.

Un **démonstrateur Web**, développé dans le cadre du projet², permet de (1) visualiser l'itinéraire culturel Via Francigena de façon dynamique; et (2) afficher sur la carte les informations pertinentes relatives aux ES présentes dans Wikipedia. Les liens vers la description plus complète des lieux provenant de Wikipedia sont en cours d'intégration.

5. Conclusion et perspectives

Nous avons présentés les premiers résultats des travaux menés dans le cadre du projet Itinéraire Numérique visant à valoriser l'itinéraire Via Francigena. Nous remarquons, en analysant les informations relatives à l'itinéraire présentes dans les réseaux sociaux et notamment dans Instagram, que les internautes s'expriment sur le sujet (posts et photos associées). Le démonstrateur est en cours d'amélioration pour ajouter l'accès aux articles Wikipedia, et pour le rendre plus attractif. En perspectives, nous souhaitons étendre l'analyse à Tweeter puis Facebook.

² Prochainement disponible à l'adresse suivante : <http://geriico-demo.univ-lille3.fr/ItineraireNumerique>

4 SAGEO'2016

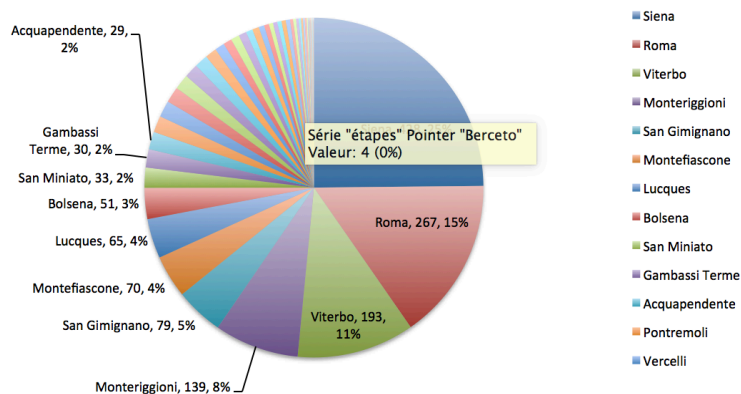


FIGURE 1. Répartition nb occurrences étapes ViaFrancigena dans les posts Instagram

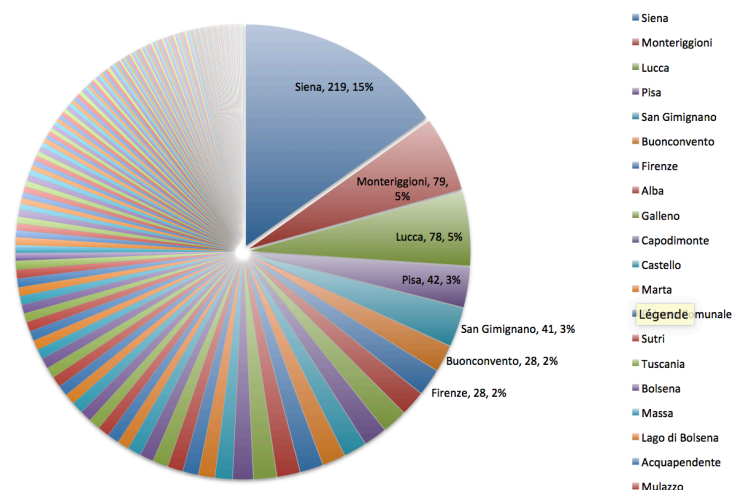


FIGURE 2. Répartition nombre d'occurrences ES Wikipedia dans les posts Instagram

Bibliographie

Lesbegueries J., Sallaberry C. & Gaio M. (2006). Associating spatial patterns to text-units for summarizing geographic information. In Proceedings of ACM SIGIR 2006. GIR, Geographic Information Retrieval, Workshop, p. 40–43 : LIUPPA.

Pospecu M., Grefenstette G., Moëllic P-A., (2009). *Mining Tourist Information from User-Supplied Collections*, CIKM '09 Proceedings of the 18th ACM conference on Information and knowledge management, p. 1713-1716.

Mohamed A., Oussalah M., (2014). *Identifying and Extract Named Entities from Wikipedia Database Using Entity Infoboxes*, International Journal of Advanced Computer Science and Applications, vol. 5, n° 7.

Tahrat S., Kergosien E., Bringay S., Roche M., Teiseire M., (2013). *Text2Geo : from textual data to geospatial information*, WIMS: International Conference on Web Intelligence, Mining and Semantics, p.4.

Une démarche de modélisation spatiale multiscalaire pour appréhender la croissance urbaine

Alexandre Ornon¹

1. CNRS, UMR 7300 E.S.P.A.C.E., Université Nice Sophia-Antipolis
98 boulevard Edouard Herriot 06204 Nice cedex 3
alexandre.ornon@unice.fr

RÉSUMÉ. La région PACA est connue pour les fortes pressions urbaines qui interagissent avec les phénomènes de métropolisation. Ces derniers à différentes échelles jouent sur la croissance et le développement des espaces urbains. Dans le cadre de l'anticipation de la croissance urbaine il est conventionnellement procédé à une analyse rétrospective pour anticiper le devenir territorial. L'objectif principal est de proposer une méthode de simulations prospectives du devenir territorial du Var et ce, par une pluralité de modèles spatialement explicites à plusieurs échelles. Il s'agit d'introduire une forme de modélisation multiscalaire dont les résultats d'un premier modèle à l'échelle régionale mettent en exergue les espaces à enjeux et conditionnent la seconde phase analytique à l'échelle métropolitaine, et dont les seconds résultats conditionnent la dernière phase prospective à l'échelle locale. La Géoprospective est le fil directeur de cette recherche, elle vise à l'anticipation des devenirs des espaces en s'appuyant sur la connaissance des dynamiques et des structures spatiales et de mettre ces projections au centre des discussions avec les acteurs du territoire dans une optique d'aide à la réflexion et à la décision.

ABSTRACT. The PACA region is known for strong urban pressures that interact with metropolisation. This impacts at different scales on urban growth and urban development. As part of the analysis of the determinants of urban growth and change of land use, a retrospective analysis is conventionally conducted to anticipate territorial future. The main objective is to propose a spatial prospective method applied on the Var department, based on several models built at different scales. It consists in introducing a multiscalar modeling where the results of a first model at a regional scale allow to highlight local strategic spaces. These results enable to formalize the second analytical step at a metropolitan scale, the results of which lead to the last prospective phase at a local scale. This research is led by the concept of Geoprospective which aims at anticipating the future of a space, based on the knowledge of spatial structures and dynamics studies so as to bring projections to the actors of territory in optics of reflection and decisionmaking support.

MOTS-CLES : Géoprospective, croissance urbaine, analyse spatiale, analyse d'images, modélisation, multi-échelles

KEYWORDS : Geoprospective, spatial analysis, multiscalar modeling, urban growth ...

1. Introduction

La région Provence Alpes Côte d'Azur connaît depuis le début des années 2000 un rythme de croissance urbaine soutenu. L'augmentation du nombre d'habitants en PACA crée des enjeux sur la dynamique du développement urbain. De 1990 à 2012, les espaces artificialisés ont enregistré une augmentation de 31% de leurs surfaces contre 20% au niveau national (Corine Landcover) principalement par l'extension des grands pôles urbains. La pression urbaine invoquée comme étant la cause de la croissance urbaine et de la pression foncière se définit généralement par le développement des espaces urbains sur les espaces naturels et agricoles environnants (Decoupigny et Passel, 2014). Afin d'anticiper la croissance urbaine sur un territoire, il est conventionnellement réalisé des modèles de simulation à base d'analyse rétrospective à une échelle précise (Liziard, 2013). Cette recherche a pour but de développer une démarche multiscale pour traiter cette problématique, en adoptant un emboîtement de différents modèles réalisés à différentes échelles pour appréhender l'ensemble des mécanismes spatiaux à l'œuvre.

2. Données

Pour le premier modèle à l'échelle régionale, les bases de données OCSOL du CRIGE PACA ont servi de base analytique et de modélisation couplées au MNT du SRTM de la région et du réseau viaire de la BD TOPO (IGN) de 2010. Pour le second modèle à l'échelle intermétropolitaine, les données relatives à l'occupation des sols ont été générées à partir de télédétection via les séries multispectrales Landsat 5-7 et 8 en fonction de l'année. Les données de la BD TOPO 2008 et 2014 ont été utilisées ainsi que l'Atlas des Zones Inondables, ZNIEFFs, Espaces Naturels Sensibles et périmètres AOC.

3. Méthode

Outre l'anticipation de la croissance urbaine dans le centre var, l'intérêt d'une approche modélisatrice à plusieurs échelles réside dans la possibilité de comprendre les déterminants de l'expansion urbaine par la simulation des processus spatiaux (Batty, 2007). Afin de traiter cette problématique de la morphogénèse urbaine, les différentes étapes de modélisation s'appuient sur l'utilisation d'un automate cellulaire. Dans le but d'observer le comportement général de la croissance urbaine dans le Centre-Var, un premier modèle est formalisé à l'échelle de la région PACA, suivant une emprise s'étendant de Marseille à Antibes. Les dynamiques rétrospectives du changement d'occupation des sols sont étudiées d'un point de vue quantitatif (matrice de transitions) et sous une approche spatiale (distances, paramètres physiques, ...) permettant de déterminer des règles de transitions. Ces règles de transition donnent lieu à des simulations rétrospectives, de $t-1$ à t_0 , la proximité entre résultats de simulation et donnée à t_0 permettant la validation théorique du modèle. Lorsque le modèle est validé par ce comparatif, il devient

SAGEO'2016 – Nice, 6-9 décembre 2016

possible de réaliser une simulation jusqu'à 2030, fondée sur les dynamiques observées et des scénarios (Voiron-Canicio et al., 2014). Ces zones sous pression de l'urbanisation permettent de recentrer la seconde échelle d'analyse autour d'une nouvelle zone d'étude et d'accroître la précision des phénomènes locaux qui conduisent à la croissance urbaine. Contrairement au premier modèle, le second, à l'échelle intermétropolitaine est plus précis tant concernant sa résolution que sur les modalités de transitions d'une occupation du sol à l'autre. Cette seconde étape tient compte également d'un potentiel morphologique à la croissance urbaine, qui localise les espaces les plus propices aux changements en fonction de la forme urbaine préexistante (Voiron-Canicio, 2009), d'un indice de pression urbaine potentielle au niveau communal (Voiron-Canicio et al., 2014), d'indices locaux d'entropie et de complexité et des protections environnementales. Des analyses et simulations rétrospectives permettent de calibrer le modèle et une fois validées, donnent lieu aux simulations prospectives. Les résultats obtenus sont destinés à formaliser une nouvelle emprise à une échelle locale, au niveau intercommunal par l'analyse des espaces aux plus fortes densités de changements potentiels (figure 1).

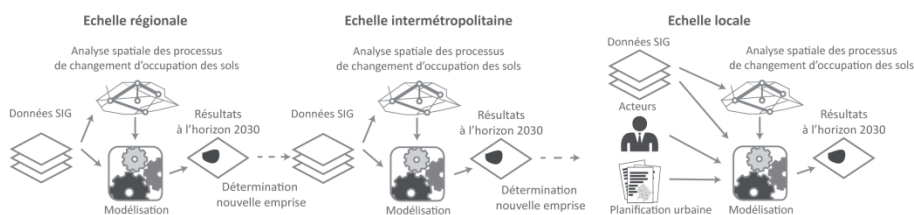


FIGURE2. Résultats cartographiques des deux premiers modèles

4. Résultats

La modélisation à l'échelle régionale vise à obtenir la localisation des changements vers les espaces urbains afin de faire ressortir différents espaces soumis à des changements potentiels. La volonté de réaliser un modèle fondé sur peu de règles a conduit à l'obtention de résultats certes généraux, mais significatifs aux vues des espaces à enjeux que le modèle permet de déceler. Le second modèle à l'échelle métropolitaine permet de localiser précisément les espaces de changements potentiels (figure 2). Il appartient au troisième niveau de modélisation (en construction) de déterminer localement des plus fortes potentialités d'urbanisation. Après présentation des résultats prospectifs aux membres de l'Agence d'Urbanisme De l'Aire Toulonnaise et du Var, les simulations de la croissance urbaine font écho avec les analyses réalisées en interne.

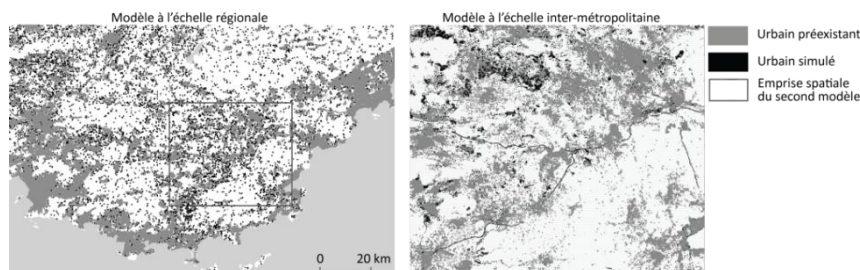


FIGURE2. Résultats cartographiques des deux premiers modèles

5. Conclusion

Il apparaît en premier lieu que la modélisation multiscalaire et spatialement explicite des dynamiques de croissance urbaine accomplit parfaitement son rôle dans la détermination des espaces à enjeux en matière d'aménagement du territoire. En effet si l'on considère que les localisations potentielles des futurs changements spatiaux conduisant à une occupation du sol urbaine sont des espaces à enjeu, la méthode proposée tient ses objectifs. Elle permet d'observer les localisations les plus probables des dynamiques de périurbanisation, densification, mitage, et ce, en fonction des processus qui ont joué de façon rétrospective pour en arriver à la date la plus récente et faire perdurer ces processus. Il a été ainsi déterminé *via* les résultats de modélisation que les espaces nécessitant une attention toute particulière quant à la localisation des changements à venir se trouvent à la jonction des aires métropolitaines marseillo-niçoise et niço-toulonnaise. Plus qu'un outil de modélisation prospective, la méthode proposée permet par le biais de simulations de comprendre de façon rétrospective et multiscalaire quels sont les processus spatiaux qui, emboîtés, conduisent à la croissance des espaces urbains. Ce chaînage de modèles est avant tout conçu comme un outil d'aide à la réflexion, préalablement à la décision. À cet égard, la réalisation du modèle au niveau local, et sa discussion avec les décideurs et gestionnaires du territoire, paraissent indispensables, dans une optique d'opérationnalité.

Bibliographie

- Batty M. (2007). *Cities and Complexity Understanding cities with cellular automata, Agent-based Models, and Fractals*. MIT Press. Cambridge
- Liziard, S. 2013. *Littoralisation de la façade Nord-Méditerranéenne : Analyse spatiale et prospective dans le contexte du changement climatique*. Thèse en géographie. Université Nice Sophia-Antipolis
- Voiron-Canicio C. (2009). *Predicting the Urban Spread Using Spatio-morphological Models*, Ceocomputation and urban planning. Studies in Computational Intelligence Vol 176
- Voiron-Canicio C., Decouigny F., Liziard S., Ornon A., Roussel D., Séry R. 2014. *PLUMES : Pression Littorale Urbaine : Modélisation, Environnement, Simulation, Application au milieu littoral marseillais*. OHM Littoral Méditerranéen, Labex DRIIHM.

Utiliser les réseaux opportunistes en cas de catastrophes naturelles : une solution pour pallier la saturation et/ou la coupure des réseaux de télécommunication ?

Douvinet J.^{1,2}, El Azouzi R.^{2,3}, Kouadio J.⁴, Hassani O.^{2,3}, Serre D.¹

1. Université d'Avignon et des Pays de Vaucluse, UMR 7300 CNRS

74 rue Louis Pasteur, Case 19, 84029 Avignon Cedex 9.

johnny.douvinet@univ-avignon.fr

2. FR Agorantic 3621 CNRS, Université d'Avignon et des Pays de Vaucluse

74 rue Louis Pasteur, Case 19, 84029 Avignon Cedex 9.

3. LIA, Laboratoire d'Informatique d'Avignon, EA 4128 CNRS,

339 Chemin des Meinajaries, 84140 Avignon.

4. Université de Reims, 9 Boulevard de la Paix, 51100 Reims.

RESUME. Le but de ce travail est de proposer une solution basée sur les réseaux opportunistes, capable de s'adapter à n'importe quel type de catastrophe naturelle et capable de collecter des données pour la gestion d'urgence sans dépendre des réseaux de télécommunication qui peuvent être saturés ou coupés. Cette technologie permet d'exploiter la mobilité des individus lorsqu'un nombre important de personnes (utilisant et/ou possédant un Smartphone) se croise sans arrêt dans une zone bien définie. Cette piste est envisageable car 60% de la population mondiale (78 % en France) possède un smartphone. Ces téléphones seront utilisés comme des appareils émetteur/récepteur d'un signal particulier, et utilisés pour localiser rapidement les porteurs au cas de besoin. Mais la mise en place d'une telle solution devra prendre en charge divers problèmes, liés notamment au routage, à l'énergie et à la capacité des réseaux.

ABSTRACT. The aim of this project is to implement a system using Delay Tolerant Networks in order to manage different type of natural disaster and to enable the collect of data for rescue services without depending on the classical telecommunication networks, potentially failed or saturated during violent crisis. Such system requires a sufficient number of smartphones in an area, whatever the people use or not their smartphone. This objective is clearly interesting as 60% of the population around the world (even 78% in France) have smartphone. We could expect to use smartphone as transmitter/receptor, but before it several problems may be solve as the routing, the energy and the network capacity.

MOTS-CLES : opportuniste, catastrophes, réseaux

KEYWORDS: DTN (Delay Tolerant Network), natural disasters, networks.

1. Contexte de recherche

Les conséquences des catastrophes naturelles sont de plus en plus importantes du fait d'une exposition accrue des populations dans des territoires à risque (près d'1/3 de la population française vit en zones inondables ; Vinet, 2010 ; Gérin, 2011) et en lien avec l'intensification des pluies à échelle fine (GIEC, 2013). Si les internautes acceptent désormais de partager leur localisation en temps réel (*via* les coordonnées GPS) et leurs déplacements (*tracking* sur Iphone), allant même jusqu'à diffuser ces données (individuelle, collective) sur des Réseaux Sociaux Numériques (RSN) plus ou moins ouverts (professionnels ou amis), ils souhaitent avoir en contrepartie accès à des réseaux urgentistes et/ou des appels de secours (Goodchild, 2007) en continu. C'est pour cette raison que les acteurs du monde humanitaire se sont très vite saisis de cette question, en développant plusieurs applications (*Ushaidhi*, 18, *HumanLife*, *Red Cross*) pour permettre à leurs agents de faire des états des lieux, des points de situation ou des évaluations sur l'étendue des dégâts après les catastrophes.

C'est le début du *Mobile Data Collection Systems* (MDCS) : sans avoir besoin d'un téléphone mobile, on peut renseigner des informations à partir d'un formulaire et les transmettre à un centre de décision (Laurila *et al.*, 2013 ; Chen et Chan, 2010). En parallèle, c'est le début de la communication à double sens, qui se fait au départ à travers une implication assistée et contrôlée des citoyens dans la collecte. Cette demande se traduit désormais par le développement d'applications spécialisées, qui sont souvent multirisques. Si les premières applications dédiées à l'alerte inondation sont apparues dès 2009 à l'échelle mondiale (*FloodAlert* puis *Flash Flood Guidance* aux Etats-Unis), il faut plutôt attendre 2013 en France pour voir, sur le marché, des solutions logicielles et matérielles sur cette thématique. Des structures publiques ou privées (communes, bureaux d'étude, services de l'Etat) se sont lancées dans cette direction et les exemples se multiplient. La « guerre » ou la « jungle » des applis est désormais lancée, tellement les solutions foisonnent (Kouadio et Douvinet, 2016).

Cependant, dans certains cas, les communications classiques comme Internet ou les téléphones (fixes ou mobiles) peuvent être hors service (pendant, voire après une catastrophe), à cause d'une panne électrique ou de la violence de l'orage qui gêne ou interrompt les liaisons et les relais (Serre *et al.*, 2012). Lors des inondations du 15 juin 2010 dans le Var (France), par exemple, les pompiers ont dû intervenir (2 357 hélitreuillages entre 16h38 et 23h15) sur de nombreux secteurs (sur 35km), alors que les transmissions étaient inexistantes (plus de téléphone filaire, plus de GSM, plus de radio analogique, et seuls quelques SMS passaient de temps de temps – d'après le colonel Eric Grohin, SDIS06 – communication orale). Plus de 200 000 personnes n'avaient plus d'électricité. Autre exemple récent : le 3 octobre 2015, de nombreux habitants sont restés injoignables pendant plus de 2 heures (entre 20h30 et 22h30), notamment dans un quartier à Biot (le Hameau du Carimaï) où les habitants ont eu jusqu'à plus de 2m d'eau dans leur maison...

Ces problèmes perturbent la réactivité des acteurs opérationnels, et notamment la mise en relation des victimes avec les secours (Boyd et Ellison, 2007 ; Corvey *et al.*, 2012). Aussi, il faut trouver des solutions plus souples et non dépendantes de ces réseaux. A ce titre, les réseaux opportunistes paraissent intéressants à explorer. Dans

SAGEO'2016 – Nice, 6-9 décembre 2016

ce cas, on parle de MANET (*Mobile Ad-Hoc Networks*) (Brownlee et Liang, 2011) ou de réseau *mesh* (Takahashi *et al.*, 2008). Ces solutions permettent de s'affranchir des réseaux de télécommunication officiels et le smartphone devient un relais pour établir un réseau pair-à-pair (*peer-to-peer*) entre plusieurs smartphones, autrement dit sans avoir un nœud central, dans l'objectif d'aboutir à un réseau de plus en plus « résilient ». Des systèmes de routage, comme BATMAN (*Better Approach To Mobile Ad-hoc Networking*), ou AODV (*Ad hoc On-demand Distance Vector*) par exemple, ont justement été créés pour être opérationnels quand les réseaux sont hors service (Wirtz *et al.*, 2011).

2. Différents scénarios à imaginer

Différents scénarios sont actuellement explorés pour tester la résistance, voire la robustesse, de l'architecture du réseau cellulaire, notamment en cas de crise. On peut ici citer quelques exemples, sans prétendre à un bilan exhaustif, loin de là.

ProximAid (*Proximal adhoc networking to Aid emergency response*) est une application développée par des chercheurs de l'université de Chypre (Koumidis *et al.* 2015) dans le cadre du projet *COncORDE*. Elle se base sur un mécanisme de balises wifi, pour une solution de réseaux *ad hoc* distribuée localisée (*HelpNet*) pour l'aide aux interventions d'urgence. En plus de permettre la découverte de périphérique, les balises émettent deux métriques liées au réseau, une pour connecter les appareils (*Local Connectivity*), et une autre pour les connecter via un pont réseau (*Exogenous Connectivity*), notamment utilisés par l'ADS (*Alert Dissemination Strategy*). Il s'agit en fait d'une stratégie intelligente de diffusion des alertes, basée sur l'expédition des messages vers les parties du réseau contenant le plus d'appareils localisés en vue de réduire la consommation d'énergie des batteries.

En situation post-catastrophe, il est aussi vital de restaurer les communications le plus vite possible. Actuellement, les systèmes sans fils mobiles sont utilisés comme solutions de secours. Mais ces systèmes demandent du temps de configuration, ont une couverture spatiale limitée et font appel à des accès satellites très onéreux. Les solutions alternatives s'appuient, de leur côté, sur les potentiels d'auto-configuration des périphériques, à travers la création de réseaux *mesh*. Les travaux de Ghafoor *et al.* (2015) portent sur la réalisation d'un système radio en vue d'évaluer le potentiel de ces nouveaux dispositifs. Basé sur l'exploitation des ondes wifi IEEE 802.11s pour le contact avec le périphérique, et sur le réseau GSM pour des accès extérieurs, le prototype de ce système intègre la gestion des appels ainsi que la transmission des données. Les premiers retours d'expérience relèvent que dans des conditions idéales, le système supporte un très grand nombre de connexions et d'appels simultanés mais qu'au-delà, la qualité des communications s'amenuise.

D'autres scénarios sont également imaginés lorsque l'infrastructure supportant le réseau cellulaire est encore fonctionnel. Les appareils mobiles pourraient servir de distribution des opérations de sauvetage et de secours dans les zones sinistrées, mais on doit s'attendre, dans ce cas là, à une saturation du réseau, donc à une réduction de la qualité des communications. En effet, en pareille situation, la connectivité totale

ne peut pas être garantie en continu, notamment entre la source et la destination d'un point à l'autre du réseau. Dans le cadre d'un tel scénario, Do-Duy et Vázquez-Castro (2015) ont proposé une architecture réseau pour pallier les difficultés que peuvent engendrer une telle situation. Elle se base sur les appareils mobiles existant dans les réseaux cellulaires classiques, où la livraison de contenu à partir du centre de commandement de la gestion des catastrophes est garantie à tous les utilisateurs *via* une connectivité intermittente. Par ailleurs, l'ensemble des ressources du réseau cellulaire peut être répertorié en utilisant des techniques de coopération entre les liens wifi et le réseau cellulaire.

On peut aussi imaginer que les opérationnels et les services de secours soient sur le terrain, dotés de smartphones et qu'ils aient besoin d'informations en provenance du centre de commandement pour être plus efficaces. L'architecture proposée permet alors de convertir les terminaux en appareil de multidiffusion interconnecté, pouvant directement communiquer entre eux. Le modèle de réseau se compose alors de deux groupes. Le premier est composé des utilisateurs directement connectés à la station de base avec des antennes relais. Toutefois, en raison des effets de l'environnement sur la qualité du signal et des obstacles présents, certaines interruptions peuvent survenir et donner lieu à une connectivité saccadée et plus ou moins intermittente. Pour pallier ce problème, le second groupe d'utilisateurs est utilisé comme source indirecte pour relayer le contenu multimédia demandé, sur la base de liens wifi, créant ainsi un réseau *ad-hoc*. Le réseau proposé peut alors être adapté par le remplacement de la station de base avec une station sans fil portable, qui est équipée d'une liaison satellite et radio (de courte portée).

3. Objectifs et méthodes mises en œuvre dans cette recherche

Cette recherche, débutée en juin 2016, se structure autour de plusieurs objectifs. Il s'agit, du point de vue méthodologique, d'utiliser et d'améliorer une plateforme, appelée *SecurePhone*, qui permettra de localiser les victimes en utilisant le réseau des smartphones pour organiser, de façon simple et rapide, les moyens de secours dans les meilleures conditions. Ce système sera composé de téléphones intelligents et de collecteurs/diffuseurs (véhicule de secours, drone, point d'accès), qui peuvent servir de points d'accès à Internet afin d'acheminer les messages de détresse vers le centre des secours, et de diffuseurs, utilisés pour activer le mode "Secure" dans les smartphones. Pour répondre à cet enjeu, il est nécessaire de mener un travail entre plusieurs disciplines, d'où la collaboration entre informaticiens et géographes.

Du point de vue opérationnel, il s'agit ensuite de mener l'étude de la faisabilité de la plateforme au niveau local, régional et national, pour mieux comprendre les spécificités des terrains en termes de vulnérabilités et de capacités (El Azouzi *et al.*, 2013). Il s'agit de voir comment ces outils participent à la diffusion en temps réel d'informations ou de comportements à adopter face à des risques naturels, mais aussi d'expérimenter des applications urgentistes pour intégrer les attentes des populations locales et formuler des recommandations à destination des élus et de la population. Ces outils pourraient devenir des leviers de résilience face aux dysfonctionnements des réseaux classiques de communication pendant une catastrophe.

SAGEO'2016 – Nice, 6-9 décembre 2016

L'objectif technologique vise à concevoir un réseau « DTN » (*Delay Tolerant Network*), en commençant par une étude qui permettra de développer les algorithmes requis et les protocoles de routage. Cette recherche propose ainsi de :

1) récolter un ensemble de traces pendant une catastrophe (la mobilité joue un rôle important sur la conception des protocoles dans les réseaux opportunistes) ; sur ce point, deux pistes seront explorées : a) faire la synthèse des enseignements tirés des études menées sur les comportements (Ruin *et al.*, 2007 ; Cavalière, 2015) et établir un profil type pour sélectionner un échantillon de personnes vulnérables ; b) définir un protocole de recrutement et d'équipement de volontaires en amont des catastrophes (en ciblant des territoires potentiellement exposés) ;

2) créer des protocoles de routage avec l'hypothèse que le réseau reste homogène (en termes de densité, mobilité, communication, batterie, capacité de stockage.). La nature des applications existantes (Kouadio et Douvinet, 2016) amène à comprendre comment les protocoles de routage dans les DTN peuvent être affectés lorsque la densité des nœuds est en grande partie non homogène, ou quand les nœuds ont des caractéristiques et des modèles divers de mobilité ;

3) étudier les limites liées à la consommation d'énergie. Les utilisateurs d'un smartphone sont conscients du fait que les batteries de leurs téléphones portables ont une quantité d'énergie tout à fait limitée. En proposant une solution à basse énergie, cela pourrait optimiser la transmission du contenu souhaité.

De manière plus spécifique, nous étudierons le mécanisme de coopération basé sur l'analyse du comportement de nœuds, et adapterons des schémas venant à la fois des développements de l'algorithmique distribuée, de la théorie des jeux et des algorithmes d'apprentissage (El Azouzi *et al.*, 2013). Les solutions seront évaluées par un simulateur, appelé ONE¹, en utilisant les traces réelles observées pendant une catastrophe naturelle. Le simulateur sera adapté pour couvrir différentes catastrophes en commençant tout d'abord par étudier les apports et les limites des DTN dans des zones sensibles aux crues rapides, ces inondations ayant des critères bien connus et affectant des secteurs prédisposés (Gaume et Borga, 2013 ; Douvinet *et al.*, 2014).

4. Conclusion et discussions

Cette recherche, débutée en 2016, pose les bases des développements attendus et ce poster sera l'occasion de présenter les différentes solutions existantes et celles qui sont imaginés pour la suite de ce travail. Il s'agit aussi d'évoquer les points critiques qui n'ont peut-être pas été imaginés, et de poursuivre les expérimentations en cours. En tous cas explorer ces solutions en cas de coupure ou de saturation des réseaux de télécommunication est une piste contextuelle et scientifique. D'un côté, bon nombre de retours d'expérience ont confirmé ces problèmes lors de crues récentes (Higonet *et al.*, 2014) ou lors de phénomènes à courte cinétique temporelle. D'un autre côté, les principaux opérateurs de télécommunication refusent, toujours, pour le moment,

¹ https://www.netlab.tkk.fi/tutkimus/dtn/theone/pub/the_one_simutools.pdf

d'appliquer les normes internationales du CBC (*CellBroadCast*). Alors que plusieurs discussions ont eu lieu entre les opérateurs et le Ministère de l'Intérieur (dès 2005 !), ces projets ont vite été annulés suite à l'arrivée, en 2009, d'un nouvel opérateur en France. Le CBC ne nécessite ni la 3G ni la 4G pour son déploiement ; il permet de diffuser massivement des SMS multilingues sur tous les téléphones, provenant de tous les pays et servis par tous les opérateurs, sur des zones cibles en cas de crise. Plusieurs systèmes existent dans le monde (*Wireless Emergency Alerts, Emergency Alert System, Emergency Public Warning System, Flash Flood Guidance Systems*) mais pas en France : les opérateurs attendent que ce soit l'Etat qui prenne en charge le coût nécessaire à la mise aux normes de leurs antennes relais. Ce choix politique peut changer dans les années à venir (comme le propose Luc Bellot, qui a déposé un projet de loi allant dans ce sens en septembre 2016) mais cela explique en tous cas la profusion actuelle d'acteurs (surtout privés) dans le développement des applications, et l'intérêt porté aux Réseaux Sociaux Numériques (RSN). De toute façon, même si ce projet de loi aboutit à la création de SMS diffusés en masse (ce qui serait une solution très complémentaire), il est désormais plus que temps d'harmoniser tous ces systèmes pour définir une politique cohérente et intégrée, en faisant une information en temps réel multi-source et multicanaux.

Bibliographie

- Boyd D., Ellison N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13,1.
- Brownlee B., Liang Y. (2011). *Mobile Ad Hoc Networks: An Evaluation of Smartphone Technologies*. Royal Military Collection Of Canada Kingston (Ontario).
- Cavalière C. (2015). *Des tweets pour l'observation des réponses sociales aux crues rapides de l'automne 2014 dans le sud-est de la France : quelles informations, quels traitements ?* Mémoire de Master « Sciences du Territoire », Université Joseph Fourier de Grenoble, 88p.
- Chen B., Chan M. (2010). *Mobicent: a credit-based incentive system for disruption tolerant network*, *Proceedings of IEEE INFOCOM*, 2010, 1–9.
- Corvey, J.W., Verma, S. Vieweg, S. Palmer, M. & Martin, J.H. (2012). Foundations of a multilayer annotation framework for Twitter communications during crisis events. LREC 2012, 8 *International Conference on Language Resources and Evaluation*, Istanbul, Turquie.
- Do-Duy T., Vázquez-Castro M. A. (2015). Efficient communication over cellular networks with network coding in emergency scenarios, *Proceeding of the 2nd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Rennes, 2015, 71-78. doi: 10.1109/ICT-DM.2015.7402048
- Douvinet J., Denolle A.S., Defossez S., Anselme A. (2011). Les maires face aux Plans de Prévention du Risque Inondation (PPRI). *L'espace géographique*, Paris, 1, 31–46.
- Douvinet J., Van de Wiel M, Delahaye D., Cossart E. (2014). A flash flood hazard assessment in dry valleys (northern France) by cellular automata modeling. *Natural Hazards*, 75 (3), 2905-2929 (doi: 10.1007/s11069-014-1470-3)

SAGEO'2016 – Nice, 6-9 décembre 2016

El-Azouzi R., De Pellegrini F., Sidi H., Kamble V. (2013). Evolutionary forwarding games in Delay Tolerant Networks: equilibria, mechanism design and stochastic approximation, *Computer Networks*.

Gaume E., Borga M. (2013). État des connaissances récentes acquises sur les crues éclair en Europe : bilan du projet de recherches européen HYDRATE (2006-2010). *La Houille Blanche* (2), 24-30.

Ghafoor S., Brown K. N., Sreenan C. J. (2015). Experimental evaluation of a software defined radio-based prototype for disaster response cellular network. *Proceeding of 2nd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Rennes, 2015, 57-63. doi: 10.1109/ICT-DM.2015.7402034

Gérin S. (2011). *Une démarche évaluative des Plans de Prévention des Risques dans le contexte de l'assurance des catastrophes naturelles: Contribution au changement de l'action publique de prévention*. Paris: Université Diderot Paris 7, thèse de doctorat, 300 p. (consulter). <http://www.tel.archives-ouvertes.fr/tel-00582692>

GIEC (2013). *Changements climatiques – Les éléments scientifiques*, Publications du Groupe d'Experts Intergouvernemental sur l'évolution du climat, 222p.

Goodchild M. F. (2007). *Citizens as sensors: the world of volunteered geography*. National Center for Geographic Information and Analysis, Nex York, 180 p.

Higonet, H, Moureau, P., Roy, F. & Weiss, W. (2014). Comment utiliser les réseaux so-ciaux en gestion de crise ? Promotion CGPT 2014/O1 n°28 (Mémoire de FAE de l'ENSOSP et certificat d'Études Politiques de Science Po).

Kouadio J.S., Douvinet J. (2016). Diffuser une alerte aux crues rapides via une application Smartphone en France : de la théorie à la mise en pratique, *Revue Ingénierie des systèmes d'information, sous presse (juin 2016)*.

Koumidis K., Kolios P., Panayiotou C., Ellinad G. (2015). ProximAid: Proximal adhoc networking to aid emergency response, *Proceeding of 2nd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Rennes, 2015, 65-72-63.

Laurila, J. K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Minh Tri Do, M., Dousse, O., Eberle, J. & Miettinen, M. (2013). From big smartphone data to worldwide research : the mobile data challenge. *Journal Persuasive and Mobile Computing*, 9(6),752-771.

Palen L., Liu S. B. (2007). *Citizen communications in crisis : Anticipating a future of ICT-supported participation*. Communication présentée à la CHI Conference, San Jose, CA : USA (p 727-736). New York, USA, ACM Press.

Ning T., Yang Z., Xie X., Wu H. (2011). Incentive-aware data dissemination in delay-tolerant mobile networks, *Proceedings of IEEE Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 539–547.

ORSEC (2013). *Alerte et information des populations, tome G.4*. Direction Générale de la Sécurité Civile et de la Gestion des Crises (DGSC) du Ministère de l'Intérieur. Paris, France.

Palmer, N., Kemp, R., Kielmann, T., Bal, H. (2012). Raven: Using smartphones for collaborative disaster data collection. *Proceedings of the International ISCRAM Conference*. Vancouver, Canada: L. Rothkrantz, J. Ristvej and Z. Franco, eds.

Ruin I., Gaillard J.-C. et Lutoff C. (2007). How to get there? Assessing motorists flash flood risk perception on daily itineraries. *Environmental Hazards*, 7, 235-244.

Saha S., Shah V.K., Verma R., Mandal R., Nandi S. (2012). Is It Worth Taking a Planned Approach to Design Ad-hoc Infrastructure for Post Disaster Communication?, Proceedings of the ACM CHANTS'12, co-located with MobiCom 2012, August 22, 2012, Istanbul, Turkey.

Serre D., Barroca B., Laganier R. 2012. *Resilience and Urban Risk Management*, CRC Press Balkema, Taylor & Francis Group, ISBN 978-0-415-62147-2.

Takahashi D., Hong X., Xiao Y. (2008). On-demand anonymous routing with distance vector protecting traffic privacy in wireless multi-hop networks. *Proceeding of the 4th International Conference on Mobile Ad-hoc and Sensor Networks*, 145-151.

Vinet F. (2010). *Le risque inondation, diagnostic et gestion*. Paris, France : Lavoisier.

Wirtz H. Heer T., Backhaus R., Wehrle K. (2011). Establishing mobile ad-hoc networks in 802.11 infrastructure mode. *Proceedings of the 6th ACM workshop on Challenged networks* 49-52.

Amélioration de la cartographie des ambiances sonores urbaines

Kamal Serrhini¹, Gaëtan Palka¹, Catherine Lavandier², Catherine Dominguès³, Saúl Gomez³, Dominique Andrieu⁴

1. UMR CNRS 7324 Citeres, Université François-Rabelais de Tours

35 Allée Ferdinand de Lesseps - 37000 Tours, France

kamal.serrhini@univ-tours.fr

2. Laboratoire ETIS, UMR 8051, ENSEA, Université de Cergy-Pontoise, CNRS, F-95000, Cergy

3. Univ. Paris-Est, LASTIG COGIT, IGN, ENSG, F-94160 Saint-Mandé, France

4. Maison des Sciences de l'Homme (MSH) Val de Loire - USR 3501

RÉSUMÉ. Au-delà de la phase de concertation du public, la cartographie réglementaire du bruit en milieu urbain du Plan de Prévention du Bruit dans l'Environnement (PPBE) repose sur une démarche linéaire de type top-down n'intégrant pas les besoins ni préférences des destinataires potentiels. Par ailleurs, en tant qu'outil de communication, la cartographie du bruit a fait l'objet de peu de travaux. Cette contribution s'intéresse à la cartographie du bruit en prenant en considération la perception d'usagers. Dans le cadre d'une démarche circulaire, une enquête par questionnaire a été menée auprès de 57 sujets volontaires aboutissant à une construction participative d'une représentation des ambiances sonores. Le but de cet article réside dans la proposition de recommandations en matière de cartographie du bruit en milieu urbain intégrant les perceptions des lecteurs.

ABSTRACT. Beyond the legal public participation, urban noise presented in the Noise Prevention Plan is based on a top-down approach (linear process) without a real integration of the needs and preferences of end users. Moreover, little research focused on noise mapping as an efficient communication tool. This contribution focuses on noise mapping by taking into account the perception of various users. As part of a circular process, a survey was conducted among 57 volunteers aiming at the construction of a participatory representation of soundscapes. The goal of this paper is to propose recommendations for mapping urban noise that integrate the perceptions of readers.

MOTS-CLÉS : intensité et ambiance sonores, cartographie efficace, besoins et préférences des usagers, densité d'information et complexité, communication cartographique.

KEYWORDS: global loudness, sound pleasantness, users' needs and preferences, information density and complexity, cartographic communication.

1. Introduction

Le bruit fait l'objet d'une réglementation importante (Rozec et *al.*, 2003). Si les premières lois (loi n° 92-1444 de 1992, loi n° 99-198 de 1999) se sont attachées à la prévention et à la maîtrise des impacts du bruit par limitation des phénomènes physiques, les lois plus récentes (Directive 2002/49/CE, Circulaire du 07/06/2007) affichent une nette évolution en faveur de la publication de cartographies réglementaires du bruit (cartes stratégiques de bruit et PPBE¹) et leur communication à destination du public. La DCE 2002/49/CE du 25 juin 2002 relative à l'évaluation et à la gestion du bruit dans l'environnement souligne que : « ces informations doivent être claires, compréhensibles et accessibles ».

Par ailleurs, un rapide examen des cartes actuelles du bruit montre certaines insuffisances (Chesneau, 2004). En tant qu'outil adapté pour la communication en aménagement du territoire, la carte est construite en suivant une « grammaire » : la sémiologie graphique (Bertin, 1977). Bien que le dégradé de couleurs respecte la norme NF S 31-130, au regard de la sémiologie bertinienne, la carte des niveaux de bruit routier (Figure 1 ci-dessous) est incorrecte puisqu'un dégradé de plusieurs teintes de couleurs est utilisé pour représenter une seule variable visuelle, le bruit : « *Quand il s'agit de représenter avec la couleur un caractère ordonné, une carte correcte devrait utiliser un dégradé dans une seule couleur* » (Pumain, 1997).

De plus le nombre de paliers différenciables ne devrait pas dépasser six à sept valeurs d'une même teinte de couleur y compris le blanc et le noir (ici 8 paliers).



FIGURE 1 : Carte du bruit routier (extraite du site Internet de la Mairie de Paris)²

¹ Plans de Prévention du Bruit dans l'Environnement.

² <http://capgeo.maps.arcgis.com/apps/MapSeries/index.html?appid=941f1e5ce3494ee79c276258909afee0&webmap=4a1cd31d83bd4380ae3c22ce050df819>

SAGEO'2016 – Nice, 6-9 décembre 2016

Par ailleurs, la carte (Figure 1) du niveau sonore moyen pondéré sur 24h (L_{DEN} en dB(A)) dû au bruit du transport routier utilise une échelle globale dont la signification et l'unité sont peu comprises par les aménageurs ou les riverains auxquelles elles sont destinées : « *Experiences and user interview show that an understanding of the absolute sound pressure level value is quite difficult* » (Schiewe et al., 2013).

Selon la théorie de la communication s'appuyant sur le langage visuel (Bertin, 1977), la carte, en tant que canal de la communication (MacEachren, 1995 ; Board 1972), est généralement construite selon un modèle le plus souvent linéaire : de l'émetteur vers le destinataire ou récepteur sans prise en compte de ses besoins et préférences, alors que ce dernier point est un facteur important de son efficacité (Cauvin et al., 2007). Pour visualiser des données sonores urbaines, Servigne et al. (2000) proposent une sémiologie adaptée (courbes isophoniques, plages colorées, symboles, etc.) sur une représentation de l'espace.

La carte de la Figure 1 est accessible à un public varié comme le demande la directive européenne 2002/49/CE (élus, aménageurs, techniciens, riverains) mais ces catégories de destinataires n'ont pas nécessairement les mêmes attentes ni les mêmes cultures (Guski et al., 1999). Servigne et al., (2000) distinguent « *la représentation de l'espace sonore observé ou estimé par un observateur averti de celle vécue par l'habitant* ». Ainsi, différents travaux ont montré que la perception visuelle est fonction d'une part de l'objet (l'image) observé et d'autre part des caractéristiques socio-culturelles de l'observateur (lecteur) : la perception est iconographique et anthropique (Fuchs et al., p. 66, 2009 ; Serrhini 2008).

Dans cet article, la méthode proposée opte pour un modèle circulaire (destinataire – concepteur – destinataire) qui intègre notamment les perceptions cognitives du lecteur (Fuchs et al., 2005 ; Meyer et al., 2012 ; Palka, 2015).

Le but de ce travail réside donc dans l'amélioration de l'information portée par les cartes de bruit, et indirectement de la prévention (cartographie du bruit en tant que risque environnemental dans les PPBE) et de la gestion (modification d'un Plan de Déplacement Urbain, insertion d'une nouvelle infrastructure de transport), afin de faciliter le dialogue entre acteurs (décideurs, riverains, services techniques...). L'hypothèse de ce travail est que le niveau sonore moyen (Figure 1) n'est pas l'indicateur idéal pour favoriser le dialogue entre les acteurs. Une carte qui s'appuierait sur des valeurs perceptives compréhensibles par tous, par exemple la qualité de l'environnement sonore, la force sonore perçue, ou la composition de l'environnement sonore urbain avec identification des différentes sources présentes dans celui-ci serait sans doute plus adaptée.

Dans une première étape et après avoir précisé le cadre de l'étude, l'ensemble du protocole expérimental sera détaillé (choix des volontaires, questionnaire). Dans une seconde étape, les résultats de l'enquête cognitive seront analysés et discutés. Enfin des recommandations seront proposées en conclusion.

2. Contexte de la recherche

Ce travail est réalisé dans le cadre du projet de recherche Cart_Asur³ « Représentations cartographiques de la qualité des ambiances sonores urbaines : acceptabilité des cartes » soutenu par l'Ademe⁴ et réunissant cinq laboratoires publics (le laboratoire MRTE et l'UMR ETIS de l'Université de Cergy Pontoise ; l'équipe COGIT de l'IGN ; l'UMR CITERES de l'Université de Tours ; le laboratoire AI-Lab de Vrije Universiteit Bruxelles) plus deux partenaires institutionnels, la ville de Paris et Bruitparif⁵.

L'objectif de Cart_Asur est d'élaborer un outil cartographique représentant le paysage sonore urbain à différentes échelles spatiales (grands axes routiers, quartiers) et temporelles (moments de la journée imposés par la Directive européenne). Cet outil s'appuie sur trois types de données : acoustiques, perceptives et spatiales ou géographiques. Pour recueillir les données perceptives et acoustiques, et afin de les représenter sur cet outil cartographique, une campagne a été menée entre septembre 2013 et septembre 2014.

2.1. Site d'étude et collectes des données

Les données perceptives et acoustiques ont été recueillies grâce à des téléphones mobiles (distribués à 57 sujets parisiens) équipés de microphones calibrés (Delaitre et al., 2014 ; Lavandier et al. 2013) et d'une application Cart_Asur dédiée (extension de l'application NoiseTube, www.noisetube.net). Les 57 participants devaient tout au long de l'année caractériser une centaine de situations différentes, 28 lieux des 13^{ème} et 14^{ème} arrondissements de Paris, à différents moments de la journée, et pendant les quatre saisons. L'application permettait à chaque participant d'accéder à ses propres objectifs fixés par le protocole général du projet (les lieux qu'il devait visiter, les moments des mesures ainsi que sa position géo localisée). Lorsque le participant était prêt à évaluer la situation, il mesurait d'abord en continu l'évolution temporelle du niveau sonore pendant dix minutes (stockage du niveau toutes les secondes) puis complétait un questionnaire élaboré et testé précédemment (Brocolini et al., 2010). Ainsi, à l'issue de chaque mesure, cette application interrogeait le sujet sur : (1) les conditions météorologiques (journée froide, chaude, etc.), (2) la perception globale de l'environnement sonore évaluée sur des échelles sémantiques en 11 points (de silencieux à bruyant, de sans vie à animé, d'enveloppant à peu enveloppant, de désagréable à agréable), (3) l'intensité sonore de six sources événementielles (mobyettes, voitures, poids lourds / bus, métro / RER, klaxons / sirènes, activités urbaines) qui pouvait être qualifiée de : faible - forte, (4) le temps de présence de six sources constituant le bruit de fond (circulation routière, voix, pas, oiseaux, bruit d'eau, bruit de vent) sur une échelle

³ <http://noisetube.net/cartasur/>

⁴ Agence De l'Environnement et de la Maîtrise de l'Énergie.

⁵ Bruitparif est un observatoire du bruit en Ile-de-France, est une association créée à l'initiative de la région, à la demande des associations de défense de l'environnement. <http://www.bruitparif.fr/>

SAGEO'2016 – Nice, 6-9 décembre 2016

allant de rarement présent à continuellement présent ; (5) enfin l'application interrogeait le sujet sur les éventuelles difficultés rencontrées (erreurs de localisation par GPS, transfert des mesures, etc.) puis demandait au participant de prendre autant de photographies qu'il le souhaitait pour caractériser visuellement la situation. Les données étaient validées lorsqu'il n'y avait plus aucune valeur manquante (Aumond P. *et al.*, 2016). Les 13^e et 14^e arrondissements de Paris avaient été choisis car offrant des lieux différenciés où les sources de bruit visées sont bien représentées. Des lieux types avaient été identifiés : boulevard, rue piétonne, place, marché, parc, etc. Au total, 3409 mesures perceptives et acoustiques ont été relevées à différents créneaux horaires sur une année (Lavandier *et al.*, 2016b). Ces mesures ont été stockées automatiquement au moment du recueil sous un format de bases de données relationnelles (PostgreSQL) sur un serveur informatique dédié (Linux Slackware).

Un premier prototype de carte a été élaboré pour représenter les grandeurs perceptives qui ont le plus d'influence dans la construction de la qualité de l'environnement sonore. Suite à une étude statistique sur la relation des variables perceptives entre elles (Ricciardi *et al.*, 2015) il a été montré que l'intensité sonore perçue a une forte influence sur la qualité sonore, mais que l'identification des sources doit être prise en compte pour ajuster le lien qui existe entre ces variables. Il a donc été envisagé dans un premier temps de représenter sur une même carte la qualité de l'environnement sonore et l'intensité sonore perçue, puis dans un deuxième temps, de représenter les sources qui composent l'environnement (Figure 2). Ces deux types de prototypes ont donc été évalués via un questionnaire.



FIGURE 2 : Exemples de prototypes cartographiques (à gauche : agrément et intensité sonores en semaine, à droite : principales sources de bruit)

2.2. Structure générale du questionnaire

D'une longueur de 23 pages, le questionnaire est articulé autour de deux thématiques complémentaires : l'évaluation perceptive et cognitive des prototypes et la comparaison du contexte sonore en termes de représentativité d'un panel d'icônes (feux tricolores, commerces, etc.) par rapport à des sources de bruit (circulation routière, voix, etc.). Il est composé de 4 parties :

- 5 pages comprenant la présentation de l'enquête, la définition des termes clés (bruit et agrément sonore) : « *Le bruit se définit comme un son constituant une gêne pour l'homme, pouvant même nuire à son état de santé (irritabilité, fatigue, stress, insomnie...).* En revanche, l'agrément sonore rend compte de la qualité de l'environnement sonore (tous les sons) entourant l'individu. Ces sons sont ceux perçus par un individu, en un point, dans une zone appelée zone d'horizon acoustique. Cette zone est limitée par les éléments de la ville que l'on considère comme des obstacles principaux tels que les bâtiments. Elle est inscrite dans un cercle de 300 m. », l'identifiant du volontaire et un test préliminaire de la qualité de la perception visuelle des couleurs (cécité chromatique) extrait de celui d'Ishihara (Ishihara, 2010). Tous les volontaires percevaient convenablement les couleurs.

- 8 pages en vis-à-vis comprenant une carte (pages 6, 8, 10 et 12) et les questions associées (pages 7, 9, 11 et 13). Cette partie correspond à l'évaluation des prototypes cartographiques en termes de propriétés globales (complexité, aspect esthétique, utilité) de la carte et de compréhension des symboles représentant l'agrément et/ou l'intensité sonores (disques pleins, demi-disques, cônes et cadrans).

- 8 pages comprenant des extraits de cartes ou les principales icônes utilisées. Cette partie correspond à la comparaison du contexte sonore en termes de représentativité de huit icônes (feux tricolores, jeux, commerces, etc.) par rapport à sept sources de bruit (circulation routière, oiseaux, deux-roues, bruits de pas, voitures, voix, poids-lourds/ bus).

- 2 pages visant à préciser à la fois quelques éléments (âge, sexe, catégorie socio-professionnelle, ...) permettant de mieux caractériser le répondant et le positionnement de la thématique traitée par rapport à son activité et sa fréquence d'utilisation de la carte en général.

Des extraits (questions) types des différentes parties du questionnaire sont présentées dans ce qui suit montrant le caractère aéré et lisible de ce questionnaire. La durée de réalisation du questionnaire par un volontaire est généralement comprise entre 30 et 45 minutes.

2.3. Les sujets tests

Les sujets tests retenus sont tous des volontaires tourangeaux qui n'ont pas participé à la phase expérimentale de relevés de terrain avec les téléphones mobiles. Puisque les caractéristiques socio-culturelles des lecteurs interviennent dans la perception des cartes (Fuchs et *al.*, p. 66, 2009 ; Serrhini 2008), le panel tourangeau a cherché à respecter au mieux la composition des volontaires parisiens qui ont effectué les relevés sur lesquels les prototypes de cartes ont été construits.

SAGEO'2016 – Nice, 6-9 décembre 2016

TABLE 1. *Différentes catégories socio-professionnelles des participants*

Catégories socio-professionnelles	Effectifs (site de Tours)	Panel initial de Cart_Asur (Paris)
Artisans, commerçants, chef d'entreprises	2 (3%)	0
Cadres, professions intellectuelles supérieures	13 (23%)	15 (26%)
Professions intermédiaires	10 (17%)	1 (2%)
Employés	6 (10%)	2 (3%)
Étudiants	25 (44%)	37 (65%)
Autres	1 (2%)	2 (3%)
Total	57	57

3. Évaluation des prototypes cartographiques

Le but précis de cet article réside dans l'amélioration cartographique en passant par l'évaluation de prototypes cartographiques représentant l'intensité et/ou l'agrément sonore par des sujet tests par l'intermédiaire d'un questionnaire.

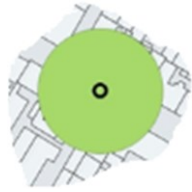



3.1. Perception et compréhension des cartes globales

Rappelons que les cartes globales soumises à l'évaluation offrent une vue d'ensemble de plusieurs lieux de mesures (Figure 3). Elles représentent une zone urbaine de 740 x 525 m. Ces cartes représentent les évaluations d'agrément et/ou d'intensité perçue par les volontaires parisiens aux lieux imposés par le protocole expérimental.

Les quatre symboles de base ci-dessous (Table 2) ont servi à l'élaboration de 4 cartes d'agrément et/ou d'intensité sonores qui ont été évaluées par les lecteurs tourangeaux en termes de compréhension et de propriétés globales.

Les modèles 1, 2 et 4 proposent une lecture du centre vers la périphérie, tandis que le modèle 3 propose une lecture inversée qui suggère que les sons sont des ondes qui proviennent de l'extérieur au point d'écoute. Le modèle 4 respecte les périodes de la journée qui sont proposées dans la directive européenne 2002/49/CE, c'est-à-dire la journée entre 7h et 19h, le soir lorsque les riverains rentrent du travail entre 19h et 23h, puis la nuit correspond à la période où les riverains sont le plus sensible au bruit. Ces périodes ont été respectées lors de la campagne de mesures des données perceptives.

TABLE 2. Les quatre modèles de symboles

Modèle 1 :	Modèle 2 :	Modèle 3 :	Modèle 4 :
			
Disque plein	Demi-disques	Cônes	Cadrans
Niveau de l'agrément sonore pendant la journée : 1 information et 1 période	Niveau de l'agrément sonore pendant la journée et la soirée : 1 information et 2 périodes	Agrément et intensité sonores pendant la journée et la soirée : 2 informations et 2 périodes	Agrément et intensité sonores pendant la journée, la soirée et la nuit : 2 informations et 3 périodes

3.1.1. Compréhension des symboles

La compréhension revient à utiliser des éléments de la légende et du titre pour répondre à deux questions sur le niveau de l'agrément sonore en cinq lieux (points 1 à 5) pour les deux modèles de disques pleins et des demi-disques (Figure 3). À titre d'exemple, pour le modèle de disques pleins (Figure 3 (a)) les deux questions posées sont :

1. Questions concernant la compréhension de la carte

a. En quel(s) point(s) l'environnement sonore est-il le plus agréable (agrément sonore) ?

Point 1 Point 2 Point 3 Point 4 Point 5 Ne sais pas Pas sur la carte

b. En quel(s) point(s) le bruit est-il le plus fort (intensité sonore) ?

Point 1 Point 2 Point 3 Point 4 Point 5 Ne sais pas Pas sur la carte

Les réponses attendues aux questions ci-dessus sont : *Point 5* pour la question *a.* et *pas sur la carte* pour *b.* En effet, sur la Figure 3 (a), représentant l'agrément sonore en semaine pendant la journée, l'environnement sonore est le plus agréable au *point 5* (disque plein en vert). En revanche, comme il ne s'agit pas d'une carte d'intensité sonore, la réponse à la question *b.* qui porte sur le bruit le plus fort est : *pas sur la carte.*

Pour l'ensemble des 57 sujets, on constate un pourcentage d'erreurs très important sur la compréhension des deux premiers modèles : pour le modèle de disques pleins, **74%** de mauvaises réponses et **75%** pour le modèle des demi-

SAGEO'2016 – Nice, 6-9 décembre 2016

disques, notamment sur la seconde question qui traite de l'intensité sonore, qui est très largement confondue avec l'agrément sonore. Ainsi, ces deux premiers modèles ne sont pas satisfaisants.



FIGURE 3. Représentation de l'agrément sonore en semaine, pendant la journée avec le modèle de disques pleins (a), pendant la journée et la soirée avec le modèle des demi-disques (b)

La même démarche a été appliquée aux deux derniers modèles des cônes et des cadrans (Figure 4) avec respectivement quatre questions spécifiques sur l'agrément ou l'intensité sonores puisque ces symboles représentent chacun deux informations (agrément et intensité) à deux (pour le modèle des cônes) ou trois (pour le modèle des cadrans) périodes de la journée.

Le pourcentage de bonnes réponses est beaucoup plus important avec **39 %** de bonnes réponses pour le modèle des cônes et **68 %** pour celui des cadrans (contre respectivement 26% et 25% pour les disques pleins et les demi-disques). L'amélioration progressive des bonnes réponses depuis le modèle de disques pleins jusqu'à celui des cadrans, peut s'expliquer en partie par un phénomène d'apprentissage par habitude (Cordier et *al.*, 2010). En effet, si les modèles disques pleins et demi-disques ont fait l'objet de deux questions sur la compréhension du message, en revanche les deux modèles suivants (cônes et cadrans) ont fait l'objet de quatre questions spécifiques et différentes les unes des autres nécessitant une certaine vigilance.

De plus, en complément aux questions sur la compréhension des différents symboles, les sujets sont amenés à se prononcer sur leurs préférences en répondant à la question suivante : « parmi les 4 symboles suivants –disque plein, demi disques...-, lequel préférez-vous ? ». **52%** des participants déclarent préférer le modèle des cadrans (28% pour les cônes, 16% pour les demi-disques et seulement 4% pour les disques pleins).



FIGURE 4. Représentation de l'agrément et de l'intensité sonores en semaine avec le modèle des cônes (a), avec le modèle des cadrans (b)

Parmi les quatre symboles, celui des cadrans est le plus complexe en termes de densité d'informations car il représente à la fois deux données (intensité et agrément sonores) et trois périodes (journée, soirée et nuit). Malgré cette complexité, ce symbole obtient le meilleur score sur la compréhension globale du message cartographique (68%) et est préféré par plus de la moitié des sujets (52%). Ce résultat peut être expliqué par le fait que ce symbole est le seul qui permette d'apprécier d'une manière localisée l'environnement sonore (agrément et intensité) à trois moments différenciés (journée, soirée et nuit) où des activités spécifiques (travail, déplacement, étude, sommeil) nécessitent un certain confort acoustique.

Hegarty et al. (2009) montrent que, dans le cas des cartes de climat, les représentations réalistes et en 3D sont généralement préférées à la fois par les novices et les spécialistes. Pourtant, ces représentations sont moins efficaces notamment quant à la compréhension du contenu du message en comparaison avec une représentation plus simple (principe de la cartographie). Dans cette enquête, les sujets préfèrent aussi le symbole le plus complexe, celui des cadrans, qui est aussi celui le mieux compris.

3.3.2. Propriétés globales : complexité, utilité et attractivité des symboles

Il s'agit d'évaluer les trois critères de complexité, d'attractivité et d'utilité de chacun des quatre modèles cartographiques. Pour chaque critère, le volontaire dispose d'une échelle horizontale et continue de cinq nuances de gris, qui facilite l'appréciation du critère.

Les moyennes des trois critères par modèle cartographique et pour les 57 sujets montrent que (i) le coefficient de variation de la complexité est de 43% alors qu'il est seulement de 4% et 7% pour l'attractivité et l'utilité. (ii) l'attractivité est presque similaire pour les quatre modèles et (iii) l'utilité est légèrement plus élevée par

SAGEO'2016 – Nice, 6-9 décembre 2016

rapport à l'attractivité (Table 3). Malgré une complexité plus élevée les cadrans sont donc les symboles les plus attractifs et les plus utiles.

TABLE 3. *Coefficients de variations des trois critères de complexité, d'attractivité et d'utilité*

	Valeur moyenne par critère et symbole		
	Complexité	Attractivité	Utilité
Disques pleins	1,23	2,91	3,18
Demi-disques	1,81	2,84	3,23
Cônes	3,19	2,96	3,51
Cadrans	3,33	3,14	3,65
Moyenne	2,39	2,96	3,39
Écart-type	1,04	0,13	0,23
Coef. de variation	43%	4%	7%

Par ailleurs, la matrice de corrélations (Bravais-Pearson) des critères deux à deux montre que pour les quatre symboles, ce sont les variables d'attractivité et d'utilité qui possèdent les liens les plus forts. À titre d'exemple, le coefficient de corrélation linéaire r de 0,729 indique une assez forte liaison entre attractivité et utilité pour le modèle des cadrans (Figure 5). Le coefficient de détermination R^2 correspondant atteint 53%.

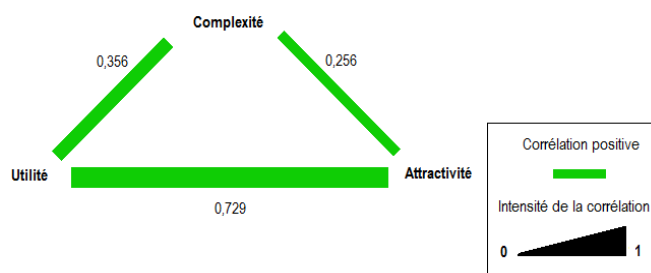


FIGURE 5. *Grphe des corrélations des critères deux à deux, cas du symbole des cadrans*

La complexité semble entretenir statistiquement un lien faible (0,356) voire très faible (0,256) avec l'utilité et l'attractivité respectivement (Figure 5). Ainsi, les trois critères étudiés peuvent donc être répartis en deux sous-groupes homogènes : la

complexité d'une part, l'utilité et l'attractivité d'autre part (Serrhini et al., p. 61, 2010).

3.4. Évaluation de la représentation du contexte sonore des cartes de détail

3.4.1. Représentativité des icônes pour différentes sources de bruit

La carte a pour objectif de représenter l'environnement sonore le long d'un boulevard ou d'une place urbaine. Différentes représentations ont été construites. Cette partie de l'enquête s'intéresse à l'évaluation de la représentativité de huit icônes pour un ensemble de sept sources sonores (Figure 6). Ces icônes sont inspirées de celles de Google-Map© avec certaines adaptations (l'icône bar-restaurant est l'agrégation des icônes de deux commerces. L'icône panier est préférée à celle de caddie pour mieux représenter les commerces de proximités, etc.). Les icônes ont été choisies pour représenter les diverses sources qui ont fait l'objet d'une évaluation perceptives (voir paragraphe 2.1), sauf les oiseaux pour lesquels la représentation des espaces verts a été préférée (Delaitre et al. 2014).

Sur une échelle allant de 0 (symbole non représentatif) à 5 (symbole parfaitement représentatif), le sujet doit évaluer la représentativité des huit icônes (feux tricolores, balançoire, commerce, laverie, métro, bar-restaurant, théâtre, arrêt de bus) vis-à-vis de sept sources différentes de bruit urbains évaluées (circulation routière, deux-roues, bruits de pas, voitures, voix, poids-lourds / bus).

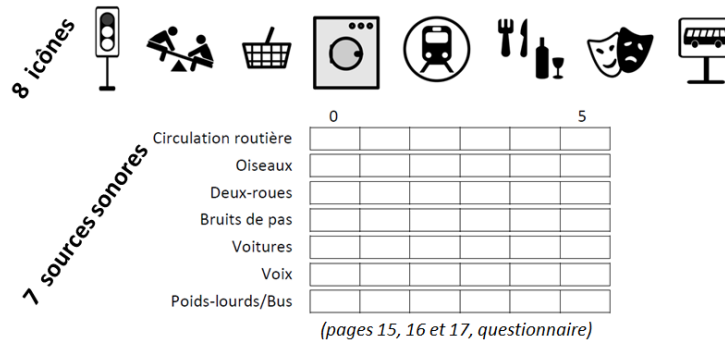


FIGURE 6. Question permettant d'évaluer la représentativité des huit icônes vis-à-vis des sept sources de bruit (environnement sonore)

Ainsi, par icône, pour les sept sources de bruit et par individu, on obtient une note moyenne. Ensuite on calcule une note moyenne pour l'ensemble des sujets : moyenne des 57 moyennes (individus) par icône et pour sept sources de bruit. Cette dernière caractérise la représentativité globale de l'environnement sonore (Figure 7).

D'une manière qualitative, la Figure 7 montre que les icônes les plus représentatives pour les 57 enquêtés sont :

SAGEO'2016 – Nice, 6-9 décembre 2016

– Feux tricolores (moyenne 2,46 / 5) : cette icône obtient généralement des notes très élevées pour des sources de bruit en rapport avec le transport terrestre (circulation routière, voitures et poids lourds / bus) ;

– et arrêt de bus (moyenne 2,17 / 5) : circulation routière, voitures et poids lourds / bus.

À l’opposé, les icônes laverie, masques, commerce, etc. sont très peu représentatives des sept sources de bruit proposées (moyenne 0,61 / 5 ; 1,09 / 5... Figure 7).

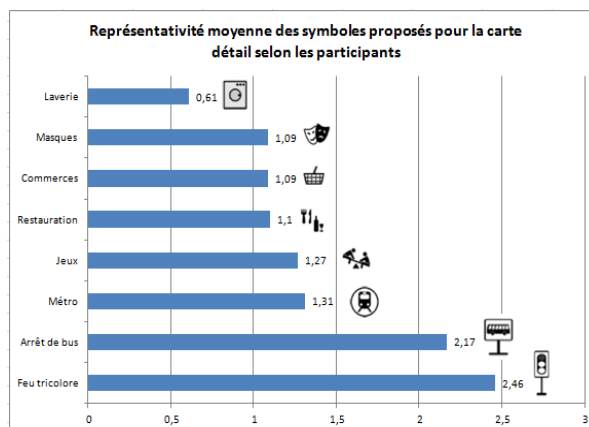


FIGURE 7. Comparaison de la représentativité des huit icônes vis-à-vis des sept sources de bruit

Le précédent résultat est confirmé par le test statistique non paramétrique de Kruskal-Wallis/Friedman qui permet la comparaison de plusieurs moyennes (paramètre de position) observées d’échantillons dépendants (ici un seul échantillon de 57 sujets > 30) avec un risque d’erreur de 5% (Table 4).

TABLE 4. Test non paramétrique de Kruskal-Wallis/Friedman

Echantillon	Effectif	Somme des rangs	Moyenne des rangs	Groupes
Laverie	57	116,000	2,035	A
Restauration	57	209,500	3,675	B
Commerces	57	215,000	3,772	B
Masques	57	215,500	3,781	B
Jeux	57	241,000	4,228	B
Métro	57	248,000	4,351	B
Arrêt bus	57	397,000	6,965	C
Feux	57	410,000	7,193	C

Le test statistique (Table 4) construit ainsi 3 classes d’icônes plus ou moins représentatives des sept sources de bruit :

– Classe A : l’icône laverie seule (la moins représentative) ;

– Classe B : regroupe bar-restauration, commerces, théâtre, jeux et métro (représentativité intermédiaire) ;

– Classe C : regroupe les deux icônes feux tricolores et arrêt de bus.

Après avoir analysé la représentativité globale (moyenne de moyennes) des icônes vis-à-vis simultanément de sept sources de bruit, on s'attachera ensuite à analyser d'une manière plus fine la représentativité de chaque icône (huit) par sources (sept) de bruit et pour l'ensemble des sujets (matrice de la Figure 8 ci-dessous) à l'aide d'une carte de chaleur.

La carte de chaleur de la Figure 8 représente le score obtenu pour chaque couple d'icône et de source de bruit pour les 57 sujets.

3.4.2. Représentativité des icônes par source de bruit

Rappelons que le participant doit évaluer la représentativité d'une icône vis-à-vis d'une source de bruit sur une échelle allant de 0 (symbole non représentatif) à 5 (symbole parfaitement représentatif). Ainsi, la réponse ou note moyenne neutre (voir Figure 6) est comprise entre 2 et 3 (le symbole n'est ni « non représentatif » ni « parfaitement représentatif »). Dans un contexte cartographique, une note inférieure à 2,5 suggère donc pour l'ensemble des participants que l'icône considérée représente mal le bruit associé.

Comme le montre le cercle en tirets de la Figure 7, les icônes feux tricolores et arrêt de bus obtiennent les notes les plus élevées ($\geq 3,72$ et $\leq 4,63$) pour les trois sources de bruit en rapport avec le transport routier (circulation, poids lourds / bus et voitures). Cependant, le couple arrêt de bus-voiture enregistre une note faible ($2,21 < 2,5$). Peu représentative de la voiture, l'icône arrêt de bus est surtout associée aux deux autres sources de bruit, poids lourds / bus ($4,63$) et circulation ($3,72$).

Ces deux icônes, feux tricolores et arrêt de bus, sont donc particulièrement intéressantes pour une communication cartographique efficace (notes les plus élevées en dehors du couple arrêt de bus - voiture) sur un environnement sonore urbain dominé par des sources de bruit des transports routiers.

Les icônes jeux, bar / restaurant, théâtre et commerces représentent des bruits de voix (Figure 8, rectangle tirets) et dans une moindre mesure des bruits de pas (Figure 8, rectangle pointillé).

L'icône jeux réalise le score le plus élevé à la fois pour le bruit de voix ($3,32$) et celui des oiseaux ($1,86$). Même si ce dernier score est inférieur à 2,5 suggérant une faible association entre l'icône jeux et le bruit des oiseaux, néanmoins cette association reste la plus forte en comparaison avec les scores obtenus par les autres icônes avec ce bruit. Cette association est probablement due au fait que les jeux urbains sont majoritairement situés dans les espaces verts qui eux-mêmes sont associés, dans nos représentations cognitives, au chant des oiseaux.

Ainsi, l'ensemble de ces icônes doit être utilisé adéquatement en fonction des principales sources de bruit urbaines présentes pour une meilleure représentation de l'environnement sonore spécifique à chaque lieu de mesures (carte de détail).

SAGEO'2016 – Nice, 6-9 décembre 2016

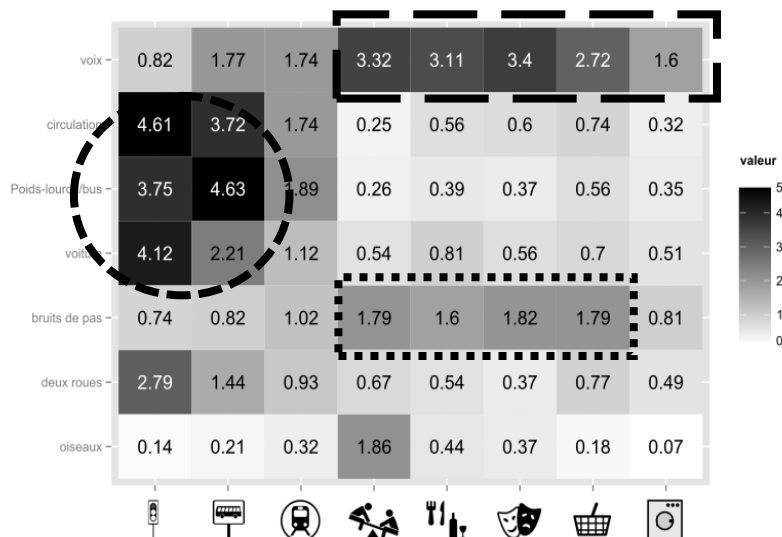


FIGURE 8. Carte de chaleur des notes moyennes obtenues (56) par couples d'icônes (huit) et de sources de bruit (sept) et pour tous les sujets (57)

4. Conclusion

Au-delà de la cartographie réglementaire du bruit (PPBE), peu de travaux concernent la cartographie du bruit en milieu urbain et encore moins prennent en compte les besoins et les préférences des usagers.

Ce projet montre qu'il est possible d'améliorer les représentations cartographiques de l'ambiance sonore, dans une démarche circulaire (destinataire – concepteur – destinataire). Les résultats statistiques mettent en avant que la représentation ponctuelle de l'ambiance sonore urbaine à privilégier tant dans sa perception que dans sa compréhension repose sur un symbole centré sur le point d'écoute, détaillant les deux composantes, agrément et intensité sonores, et montrant les périodes de la journée comme un disque d'horloge.

En complément de cette représentation, certaines icônes sont plus représentatives de la source sonore. Les icônes feux tricolores et arrêt de bus expriment le mieux le bruit de la circulation tandis que l'icône jeux d'enfants traduit le mieux les voix et les bruits d'oiseaux. Inspirées par Google-Map©, les icônes étudiées (au nombre de huit) représentent ainsi relativement bien certaines des sources de bruit considérées (au nombre de sept) en termes de perceptions et d'attentes des lecteurs et permettent de construire une cartographie plus adéquate de l'environnement sonore.

Qu'en est-il de l'adéquation entre l'environnement sonore urbain et les nombreuses autres icônes utilisées par les outils de la cartographie ? Google-Map© utilise l'icône sac-à-main pour représenter indifféremment une multitude de commerces (téléphonie, optique, super marché, fleuriste, magasin de luxe, enseignes

de voitures, etc.) et quasiment la même icône pour représenter un cabinet médical ou une pharmacie. Dans le cadre d'une cartographie de l'environnement sonore urbain, une sélection plus large de telles icônes pourrait être évaluée en vue d'une éventuelle exploitation plus rigoureuse en lien avec le message escompté.

En termes de bruit, les groupements des icônes testées suggèrent une géographie des espaces urbains en trois types :

- les lieux de quiétude tels que les parcs et jardins urbains où dominent les bruits de voix et d'oiseaux ;
- les lieux de vie collective tels que les restaurants et les commerces où dominent les bruits de voix et de pas, synonymes de présence humaine ;
- et enfin les lieux où prédominent les infrastructures de déplacements (réseau de transport routier).

Lieux	Icônes	Sources de bruits
de quiétude	Jeux	Voix et oiseaux
de vie collective	Restaurants, commerces, masques, laveries	Voix et bruits de pas
à prédominance d'infrastructure de transport	arrêt de bus, métro et feux tricolores	Poids lourds / bus, voiture, deux roues, circulation

En complément à l'étude de la perception et de la cognition (questionnaire), ces prototypes cartographiques de bruit seront aussi évalués en termes de perception visuelle (mouvements oculaires) à l'aide d'un eye-tracker. Enfin, l'association de l'image (la vue) au son (l'ouïe) pour une meilleure restitution de l'environnement sonore urbain (Slocum et al., 2009, p. 481) pourra être étudiée dans le cadre d'une interface Web avec des cartographies dynamiques de l'environnement sonore urbain (Servigne et al., 1999, 2000).

Bibliographie

- Aumond P., Lavandier C., Ribeiro C., Gonzalez Boix E., Kambona K., D'Hondt E., Delaitre P., *A study of the accuracy of mobile technology for measuring urban noise pollution in large scale participatory sensing campaigns*. Appl Acoust (2016), <http://dx.doi.org/10.1016/j.apacoust.2016.07.011>
- Bertin J. (1977). La graphique et le traitement graphique de l'information, Flammarion, Paris, 277 p.
- Board C. (1972). Cartographic communication. Cartographica: The International Journal for Geographic Information and Geovisualization, vol. 18, n°2, p. 42-78.
- Brocolini et al. (2010). Méthodes de prédiction de la qualité du paysage sonore urbain : régression linéaire multiple ou réseau de neurones ?, *CD-ROM des proceedings du 10^{ème} Congrès Français d'Acoustique*, Lyon.
- Buard E., Ruas A. (2007). Evaluation of colour contrasts by means of expert knowledge for on demand mapping, *Actes de la conférence de cartographie internationale ICA 2007*, Moscou.

SAGEO'2016 – Nice, 6-9 décembre 2016

- Cauvin C. et al. (2007). *Cartographie Thématique 2 - des transformations incontournables*, Éditions Lavoisier, Paris.
- Cordier F., Gaonac'h D. (2010). *Apprentissage et mémoire*. Sous dir. d'Alain Blanchet, 2^e édition, Armand Colin. 128 p.
- Delaitre P. et al. (2014), "Évaluation de l'agrément sonore en milieu urbain à l'aide de téléphones mobiles", CD-ROM des actes du 12^{ème} Congrès Français d'Acoustique, 6 pages, Poitiers.
- Defreville B. (2005). *Caractérisation de la qualité de l'environnement sonore urbain : une approche physique et perceptive basée sur l'identification des sources sonores*. Thèse en Acoustique, Université de Cergy-Pontoise.
- Fuchs S. et al. (2009). Evaluating cartographic design in flood risk mapping. *Environmental Hazards*, vol. 8, p. 52-70.
- Fuchs S. et al. (2009). Evaluating cartographic design in flood risk mapping, *Natural Hazards Earth System. Science*, vol. 8, n° 1, p. 52–70.
- MacEachren A. M. (1995). *How maps work: representation, visualization, and design*. Guilford Press.
- Meyer V. et al. (2012) Recommendations for the user-specific enhancement of flood maps, *Natural Hazards Earth System. Science*, vol. 12, n° 5, p. 1701–1716.
- Mignerot J.G. et al. (1980). *Acoustique urbaine*, Edition Masson les presses de l'Université de Québec, Québec.
- Montès C. (2012). La ville, le bruit et le son, entre mesure policière et identités urbaines. *Géocarrefour*, vol. 78, n° 2, p. 91-94.
- Hegarty M. et al. (2009). Naïve cartography: how intuition about display configuration can hurt performance, *Cartographica*, vol. 44, n° 3, p. 171-186.
- Lavandier C. et al. (2013). Urban sound quality assessment with mobile technology : The Cart_Asur project, *ACOUSTICS2013*, New Delhi.
- Lavandier C. et al. (2016a). Cartographie de l'agrément sonore de l'espace public urbain à partir de données géo-référencées, *Congrès Français d'Acoustique*, Le Mans.
- Lavandier C. et al. (2016b). *Représentations cartographiques de la qualité des ambiances sonores urbaines : acceptabilité des cartes*, Rapport final du contrat de recherche PREDIT, 2014.
- Chesneau E. (2004). Propositions pour une cartographie du risque, *Cartes & Géomatique*, vol. 81, p. 55-70.
- Ishihara, S. (2010). *Ishihara's Tests for Colour Deficiency: 24 Plates*, Kanehara.
- Delaitre P. et al. (2014). Évaluation de l'agrément sonore en milieu urbain à l'aide de téléphones mobiles, *CDROM des actes du 12^{ème} Congrès Français d'Acoustique*, Poitiers.
- Palka G. (2015). *Amélioration de la cartographie d'inondation par l'intégration des besoins et préférences des destinataires*. Thèse en Aménagement de l'espace, Université François Rabelais de Tours.
- Pumain D., Saint-Julien T. (1997). *L'analyse spatiale. Localisations dans l'espace*, Armand Colin, Paris.

- Ricciardi P. et al. (2015). Sound quality indicators for urban places in Paris cross-validated by Milan data, *Journal of the Acoustical Society of America*, vol. 138, n° 4, p. 2337–2348.
- Rozec V., Ritter P. (2003). Les avancées et les limites de la législation sur le bruit face au vécu du citoyen. *Géocarrefour*, vol. 78, n° 2, p. 111-119.
- Serrhini K., Palka G. (2010). *Improving Flood Risk Maps as a Means to Foster Public Participation and Raising Flood Risk Awareness: Toward Flood Resilient Communities*. Rapport de recherche ERA-NET CRUE Funding Initiative du projet RISKMAP, septembre 2011
- Serrhini et al. (2008). Sémiologie graphique expérimentale et cartographie du risque d'inondation, *Actes du 26ème Congrès INFORSID CEMAGREF*, Fontainebleau.
- Servigne S., Kang M-A. SIG pour l'environnement sonore urbain : De la représentation à la visualisation animée. *Revue internationale de Géomatique*. 2000. Vol. 10, n°3-4. Pp. 411-432.
- Servigne S., Laurini R., Kang M.-A., Li K.-J. (1999). First specifications of an information system for urban soundscape. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems '99*, Florence, June 7-11, 1999 Volume II, pp. 262-266.
- Schieve J., Weninger B. (2013). Visual Encoding of Acoustic Parameters – Framework and Application to Noise Mapping. *The Cartographic Journal*, vol. 50, n° 4, p. 332-344.
- Slocum et al. (2009). *Thematic Cartography and Geovisualization* (chapter 26 “Using Senses other than Vision to Interpret Spatial Patterns”).
- Weninger B. (2013). The Effects of Colour on the Interpretation of Traffic Noise in Strategic Noise Maps, *Actes de la conférence de cartographie internationale ICA 2013*, Dresde.

Vers une modélisation de la sémantique des trajectoires de navigation maritime côtière

Wissame LADDADA¹, Eric SAUX¹

Institut de Recherche de l'Ecole Navale, Lanvéoc-Poulmic, CC600, F-29240 Brest Cedex 9, France

{wissame.laddada,eric.saux}@ecole-navale.fr

RÉSUMÉ. La modélisation des systèmes de routage est sujette à de plus en plus de défis. D'une part en raison de la nette avancée dans les théories spatio-temporelles et d'autre part en raison du besoin d'aide à la navigation dans différents environnements. Dans ce contexte, notre contribution consiste à formaliser le processus de wayfinding qui caractérise les différents raisonnements liés au routage. Plus précisément, nous proposons un modèle formel fondé sur la logique de description pour l'aide à la navigation maritime côtière à des fins d'exploitation de la sémantique qu'il offre par les systèmes d'information afin d'en déduire une trajectoire.

ABSTRACT. Routing system modeling is becoming more challenging. This is due to the significant advances in spatio temporal theories and the need of navigation assistance in different environments. Within this context, our contribution is to formalize the wayfinding process which characterizes the different reasoning related to routing. More specifically, we propose a formal model based on the description logic to assist coastal navigation. Using the semantic of the model enables information systems to infer a trajectory.

MOTS-CLÉS : wayfinding, itinéraire maritime, formalisation, logique de description, système d'information géographique.

KEYWORDS: wayfinding, maritime trajectory, formalization, description logic, geographic information system

1. Contexte et objectifs

La navigation maritime nécessite une analyse spatiale, temporelle et sémantique de l'environnement car le navigateur doit trouver son itinéraire en se focalisant sur un raisonnement lié au sens que peut refléter cet environnement. En d'autres termes, la navigation fait appel au processus de wayfinding qui prend en considération son aspect cognitif. Il est alors important d'intégrer un modèle de wayfinding dans les systèmes d'aide à la navigation maritime. Cette dernière s'effectue sur un espace ouvert non contraint par un réseau de circulation comme en navigation urbaine. Ceci rend la description du processus de wayfinding plus complexe puisque l'espace de recherche n'est pas réduit.

Dans ce cadre, nous nous intéressons à la formalisation du processus de wayfinding en navigation maritime côtière qui a pour spécificité de fonder ses créations sur la notion de repères. Ces derniers doivent de ce fait être pris en compte dans le modèle formel afin d'inférer une trajectoire relative à la sémantique de l'environnement. Ainsi, nos ambitions s'inscrivent dans la perspective d'intégrer ce modèle dans des systèmes d'aide à la navigation afin d'exploiter sa sémantique et les méthodes de raisonnement pour en déduire une trajectoire maritime côtière.

2. Modélisation

La modélisation du processus de navigation a fait l'objet de plusieurs travaux. Afin de mieux représenter ce processus, les recherches se sont orientées vers l'analyse des descriptions d'itinéraire linguistique (écrite, orale) ou graphique (schémas, croquis). Les contributions dans la première description ont intéressé principalement Denis (Denis, 1997) qui construit un cadre général fondé sur la relation action/repère. Dans un autre contexte, ces descriptions verbales ont contribué à différentes analyses d'itinéraire pour une meilleure compréhension spatiale dans un milieu intérieur « indoor » (Raubal, Egenhofer, 1998), ou dans un milieu naturel (Brosset, 2008; Egorova *et al.*, 2015). La description graphique quant à elle permet d'inférer des relations spatiales sur des entités et d'exprimer les déplacements en prenant ces dernières comme référence. Les chorèmes sont d'ailleurs un outil de représentation cartographique adapté à la visualisation d'une dynamique de l'espace. D'un autre point de vue, les contributions associées à la modélisation du processus de wayfinding peuvent être classifiées selon leur prise en compte ou pas de certains paramètres, comme l'aspect temporel (Tsatcha, 2014), le raisonnement (Kiefer *et al.*, 2015) et les repères (Kiefer *et al.*, 2015). La distinction entre les travaux peut aussi être faite sur le type du modèle élaboré : modèle graphique (Brosset, 2008), modèle avec grammaire (Kiefer *et al.*, 2015) ou modèle ontologique (Tsatcha, 2014).

Dans la démarche faite en vue d'atteindre notre objectif et suivant cette littérature, notre travail s'appuie d'abord sur l'analyse d'une description mixte

SAGEO'2016 – Nice, 6-9 décembre 2016

(linguistique et graphique). Ensuite, nous établissons notre modèle fondé sur une formalisation avec la logique de description (DL). Ce choix est motivé par le fait que DL fournit une hiérarchisation conceptuelle d'un domaine d'application, ce qui nous permet d'intégrer les différents paramètres imbriqués (temps, raisonnement et repère) et concevoir par la suite un modèle ontologique.

3. De la description à la formalisation

La logique de description s'appuie sur la définition de plusieurs concepts grâce auxquels sont faites les inférences. Pour définir ces concepts nous avons procédé dans l'ordre à l'analyse linguistique et à l'analyse graphique comme développé ci-dessous.

3.1. Analyse linguistique

Afin de planifier les trajectoires maritimes en environnement côtier, nous devons tenir compte des concepts de la navigation en ce milieu. Pour ce faire, nous analysons dans un premier temps les descriptions textuelles faites par des experts (ex : *route au 254 pour 3.13 Nq. La route est sur un relèvement constant par l'avant sur le phare du Petit Minou*). Ceci nous permet de définir les principes et la terminologie de cette navigation, comme le fait que la *trajectoire* est composée d'une succession de routes, que chaque changement de route est marqué par un *point de décision* et que chaque route est orientée selon un *relèvement* ou un *alignement* par rapport à des *amers*. Une fois la terminologie acquise, nous en déduisons les paramètres qui influencent la trajectoire, à savoir, les entités spatiales représentées par l'aménagement de l'environnement et les repères (amers) ainsi que l'influence temporelle qui à son tour a un impact sur l'aspect spatial (ex : la visualisation des amers qui diffère de nuit ou de jour).

3.2. Analyse graphique

Pour une conceptualisation plus intuitive et une meilleure compréhension spatiale, nous effectuons également une analyse graphique des principes de la navigation ainsi que de ses paramètres. A titre d'exemple, la trajectoire est subdivisée en plusieurs *segments* ([DeparturePoint,DecisionPoint], [DecisionPoint,DecisionPoint], [DecisionPoint,DecisionPoint], [DecisionPoint,DestinationPoint]) qui représentent des routes. Chacune d'entre elles est sur une *ligne* de relèvement/alignement déduite à partir d'une *ligne de visibilité*. Les relations existantes entre les lignes font référence aux relations topologiques (intersection, overlapping, etc.). La forme géométrique de chaque principe ou paramètre est définie quant à elle en se fondant sur les concepts de la formalisation geo-spatiale de l'OGC (Point, Line, LineString, etc.).

3.3. Définition des concepts pour la formalisation

A partir de ces deux analyses, nous définissons les différents concepts (route, point de départ, alignement, ligne de visibilité, etc.) de notre modèle formel pour en déduire une trajectoire à partir de la sémantique que renvoie l'environnement dans un contexte de navigation maritime. A titre d'exemple, il en résulte la définition suivante qui formalise la notion de point de décision : $DecisionPoint \equiv \forall Point \sqcap equals.((Bearing \sqcap intersection.(LeadingLine \sqcup Bearing)) \sqcup (LeadingLine \sqcap intersection.LeadngLine))$.

4. Conclusion

Le processus de wayfinding en environnement maritime côtier est associé à une sémantique très riche. Ceci nous a incité à établir des analyses afin de formaliser la navigation. La formalisation proposée nous a permis d'inclure une sémantique dans les différents concepts identifiés. Les travaux en cours consistent à définir l'ontologie relative à ce modèle formel dans laquelle certains de ses concepts nécessitent une alimentation automatique qui s'effectue grâce un algorithme fondé sur la recherche spatiale et le raisonnement.

Bibliographie

- Brosset. (2008). *Description d'itinéraire en milieu naturel : modèle intégré de description verbale et de représentation spatiale au sein des sig*. Thèse de doctorat, École Nationale Supérieure d'Arts et Métiers (ENSAM Paris).
- Denis. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, vol. 16, p. 409-458.
- Egorova E., Tenbrink T., Purves R. S. (2015). Where snow is a landmark: Route direction elements in alpine contexts. In *Spatial information theory - 12th international conference, COSIT 2015, santa fe, nm, usa, october 12-16, 2015, proceedings*, p. 175-195.
- Kiefer P., Scheider S., Giannopoulos I., Weiser P. (2015). A wayfinding grammar based on reference system transformations. In *Spatial information theory - 12th international conference, COSIT 2015, santa fe, nm, usa, october 12-16, 2015, proceedings*, p. 447-467.
- Raubal, Egenhofer. (1998). Comparing the complexity of wayfinding tasks in built environments. *Environment & Planning B: Planning and Design*, vol. 25, n° 6, p. 895-913.
- Tsatcha. (2014). *Contribution à l'extraction et à la représentationsentation des connaissances de l'environnement maritime. proposition d'une architecture dédiée aux applications de navigation*. Thèse de doctorat, Université de Brest.

Visualisation et analyse spatio-temporelle de trajectoires maritimes

Mattia Bunel¹, César Ducruet¹, Claire Lagesse², Éric Mermet^{3,4} et Sélina Petit¹

¹CNRS UMR 8504 *Géographie-Cités*

²CNRS UMR 6049 *ThéMa, Université de Bourgogne – Franche-Comté*

³CNRS UPS 3611 *Institut des Systèmes Complexes – Paris Île-de-France*

⁴CNRS UMR 8557 *Centre d'Analyse et de Mathématique Sociales*

le 28 octobre 2016

MOTS-CLÉS : Web-SIG, Cartographie dynamique, Analyse de graphe, Analyse spatio-temporelle, Flux maritimes

KEYWORDS : Web-SIG, Dynamic cartography, Graph analysis, Spatio-temporal analysis, Maritime flows

Introduction

La dernière décennie a vu se multiplier les visualisations en tous genres de flux maritimes traversant le globe. En effet, la mise à disposition de données nouvelles sur la position en temps réel des navires, qu'elles soient fournies par antenne radar (on parlera de données AIS ou Automated Identification System) ou par satellite (AVHRR - Advanced Very High Resolution Radiometer) (Etienne et al., 2015) a permis des représentations innovantes. De plus, en vertu du rôle de premier plan du transport maritime dans l'économie mondiale (environ 90 % du commerce mondial en volumes est assuré par voie maritime), ces flux se sont révélés être de bons marqueurs de l'intensité des activités humaines et par-là de leur impact sur l'environnement. D'aucuns considèrent pourtant que l'on manque, encore actuellement, d'une fine connaissance de la répartition des navires (Tournadre, 2014). Si les premières visualisations connues de ces flux maritimes mondiaux, fournies par des géographes dès les années 1940, n'avaient finalement pas beaucoup à envier aux représentations plus actuelles (voir une synthèse dans Ducruet, 2016), ces dernières sont par ailleurs marquées par une grande diversité d'objectifs et de méthodes. Sans prétendre à l'exhaustivité, on peut citer par exemple les travaux dans le champ de la paléoclimatologie (pro-

jets NOAA¹, ICOADS², Old Weather³, et CLIWOC⁴) utilisant des données historiques ou *ship logs* (carnets de bord) ayant abouti à des représentations fidèles des routes maritimes mondiales à différentes époques, par la cartographie des trajectoires des navires (Garcia-Herrera et al., 2005), et parfois à des visualisations dynamiques⁵.

L'essentiel des autres contributions consiste en la production de *heatmaps* selon différentes techniques d'agrégation des trajectoires, via la mesure de l'impact du *shipping* sur le milieu naturel (Halpern et al., 2008), les émissions carbone potentielles (cf. équipe Shipviz⁶), ou simplement pour servir la veille économique (projets IDV Solutions⁷ et ExactEarth⁸). Des visualisations sous ces angles ont pu être proposées, le plus souvent en 2D (projet Shipmap⁹ ou encore d'autres travaux abrités par Getty Images¹⁰), la 3D étant abordée plus rarement mais de façon encore statique (Shen et al., 2013). Enfin d'autres représentations ont préféré transformer ces trajectoires en graphe afin de mesurer la connectivité du réseau maritime, par rapport à des problématiques d'invasions biologiques marines (Kaluza et al., 2010) mais aussi de hiérarchies portuaires en lien ou non avec les sociétés littorales (Ducruet, 2015). Des geoportails dans le domaine historique et archéologique ont également vu le jour, dédiés à un contexte particulier comme les routes romaines antiques (Scheidel, 2013), celles de *l'Incanto* vénitien (Fournier, 2015) ou encore les flux de commerce français dans le monde et en Méditerranée au 18^{ème} siècle (Marzagalli, 2015). Le National Maritime Museum de Greenwich, quant à lui, travaille au « Great Map Project »¹¹ pour fournir au public une visualisation pédagogique des routes maritimes historiques à partir de données d'archives.

Dans ce contexte, le projet ERC World Seastems, initié en 2013, a pour objectif d'étudier les flux maritimes mondiaux dans leur dimension spatiale mais aussi temporelle, grâce à la découverte et l'extraction de données inédites permettant de reconstituer les routes maritimes mondiales sur 130 années. Ce projet ambitionne ainsi de compléter l'approche dominante focalisée sur les données récentes par une prise de recul mettant en valeur la profondeur historique de ces flux, et ce qu'ils peuvent nous apprendre sur le passé et le présent en termes de mondialisation et de développement régional. Nous commencerons par présenter rapidement les données ainsi que leur potentiel et leur utilisations futures. La base de données recueillie par le projet ne fournissant pas des géométries directement exploitables, il a donc été nécessaire de procéder à différentes étapes allant de la normalisation des entrées à leur enrichissement. Nous décrirons rapidement ces opérations en nous attardant particulièrement sur une étape cruciale,

1. National Oceanic and Atmospheric Administration, U.S. Department of Commerce, voir <http://aeinews.org/archives/2226>

2. <https://www.gislounge.com/mapping-ship-logs/>

3. <https://www.oldweather.org/>

4. <http://pendientedemigracion.ucm.es/info/cliwoc/>

5. <http://www.theguardian.com/news/datablog/interactive/2012/oct/01/first-world-war-royal-navy-ships-mapped>

6. <https://www.ucl.ac.uk/research-it-services/case-studies-pub/2016-03-rsdg-shipviz>

7. <http://uxblog.idvsolutions.com/2011/10/shipping-mix.html>

8. <http://www.exactearth.com/products/exactais-density-maps>

9. <https://www.shipmap.org/>

10. <http://www.gettyimages.fr/detail/vid%C3%A9o/earth-connections-europe-aerial-maritime-ground-routes-and-film/485519208>

11. <http://mw2014.museumsandtheweb.com/paper/the-great-map/>

SAGEO'2016 – Nice, 6-9 décembre 2016

la reconstruction des routes maritimes. Au delà du traitement des données, un portail web de visualisation cartographique, GeoSeastems, est développé au sein du projet. Il offre de puissants outils *ad hoc* de visualisation et d'analyse spatio-temporelle des données. Son objectif est d'être un support d'exploration adapté à la spécificité de notre problématique, et qui pourra, à terme, être décliné en un outil pédagogique adapté aux exigences du grand public.

1 La base de données

Le projet World Seastems repose sur l'exploitation du corpus de données de la Lloyd's List, premier assureur maritime mondial¹² regroupant l'ensemble des mouvements des navires assurés par la compagnie¹³. Nous avons construit une base de données à partir d'un échantillonnage du fonds d'archives sur la période 1977–2008. À l'heure actuelle, nous sommes en possession du tiers des données sur cette période. À terme, la base de données portera sur l'ensemble du fonds, soit la période 1890–2008, avec une extension en cours pour les flux de porte-conteneurs sur la période 2009–2016. Ces données fournissent des informations pour chaque navire : comme son nom, le nom de l'armateur, le pavillon, l'assureur, le tonnage ainsi que sur son trajet. Ce dernier est caractérisé par un port et une date de départ, ainsi qu'un port d'arrivée et éventuellement un port d'escale. Ce type de variable ne permet qu'une spatialisation partielle du tracé, contrairement à des sources de données plus récentes, telles que les données AIS¹⁴. Les variables recueillies ne nous informent pas sur la forme réelle des routes suivies. Il est par conséquent nécessaire d'interpoler la géométrie des routes pour espérer exploiter pleinement les données recueillies.

2 La méthodologie d'agrégation de trajectoires

La reconstitution des routes maritimes a nécessité d'enrichir les données de la Lloyd's avec des données exogènes, telles que les coordonnées des ports. Nous avons donc sémantisé, spatialisé et normalisé la table des ports extraite des données de la Lloyd's. Ce qui nous a permis de construire un graphe spatialisé, dont les nœuds sont les ports et les liens sont les routes maritimes. Toutefois se pose le problème de l'irréaliste linéarité des arcs du graphe ainsi obtenu. Nous avons résolu ce problème en créant un maillage anisotrope, modélisant au mieux l'ensemble des routes (Lagesse et al., 2016). C'est à partir de cette grille qu'ont été extraits les plus courts chemins entre ports, *i.e.* les routes maritimes.

La grille a été construite suivant une méthodologie en quatre étapes. Tout d'abord, une grille régulière sur la surface des océans a été constituée. Celle-ci comporte un pas de 5°, à laquelle nous avons ajouté un tirage aléatoire de 75 000 points. Cet effet de normalisation et de tirage aléatoire permet d'obtenir à la fois une grille dont la densité est homogène, mais également de parer aux chemins qui prendraient des formes trop rectilignes ou encore en forme d'escaliers. Le tirage aléatoire des points sur la surface des océans du monde entier ne permettant

12. La Lloyd's assure environ 80 % de la flotte mondiale.

13. Auxquels peuvent s'ajouter certains navires assurés par des compagnies concurrentes (Ducruet et al., 2015).

14. Par ailleurs consultables sur des portails cartographiques comme MarineTraffic.

pas d'obtenir une connexité du graphe sur des zones dont les passages sont très étroits (Gibraltar, mer Baltique, baie de Hudson, mer du Japon, etc.), nous avons choisi d'effectuer un second tirage aléatoire de 75 000 points le long des côtes pour densifier les zones potentiellement déconnectées. Ceci a également pour effet d'obtenir un lissage des trajectoires avec un effet de cabotage pour les connexions entre les ports proches les uns des autres. Nous obtenons alors un support réseau sur les mers et océans comportant un peu plus de 160 000 nœuds (incluant les ports) et 460 000 liens. Le réseau ainsi créé est déposé et géré par une base de données PostGreSQL/PostGIS. La partie dédiée aux calculs de chemins s'appuie quant à elle sur la librairie PgRouting.

Une dernière étape importante fut de parer aux problèmes concernant la projection utilisée. En effet, la projection choisie, bien qu'elle garantisse l'échange et l'interopérabilité des données produites avec le Web, a l'inconvénient de pas lier nativement l'est de l'océan Pacifique à l'ouest. À cette étape, la logique de graphe nous permet d'effectuer ces connexions virtuelles sur la ligne de changement de date. Les liens ajoutés comporteront une géométrie nulle afin de ne pas bruyter les chemins. Pour corriger les trajectoires insolites, de légers traitements manuels ont été effectués et impactent un peu moins d'un quart des liens (environ 100 000). Le premier traitement concernait des contraintes qui ont été ajoutées sur les passages par les zones de banquises. Le poids des liens situés au dessus de 70°N et au dessous de 70°S ont été augmentés de 20 %. Le second traitement manuel a été la saisie manuelle des canaux de Suez et de Panama ¹⁵. D'autres sont prévus comme l'intégration d'une couche virtuelle sur les voies navigables permettant d'inclure les ports fluviaux et lacustres dans le réseau maritime.

Dès lors que les poids sont correctement renseignés, il est possible d'interroger la base de données et de calculer des itinéraires de port à port. Avec la configuration actuelle, le temps de calcul pour un itinéraire est de l'ordre de 800 ms à 1 s. Les calculs d'itinéraires entre ports aboutissent à la création d'une géométrie visualisable, constituant l'ensemble des géométries des liens parcourus par les chemins dont le réseau maillé est le socle.

3 Le géoportail

Parallèlement au traitement des données est développé un portail de visualisation cartographique conçu comme une réponse à une triple demande : visualisation de la donnée, valorisation de l'information et vulgarisation des analyses. Des géoportails portant sur les flux maritimes et remplissant au moins un de ces objectifs existent. On pourra notamment citer le site ORBIS ¹⁶ (Scheidel 2013b), ou bien MarineTraffic, site web à visée commerciale. Toutefois aucune des solutions existantes ne répond de façon satisfaisante à nos besoins. En effet, les produits développés dans un cadre académique sont trop spécifiques pour être adaptés à d'autres problématiques. Les solutions commerciales utilisent quant à elles des données AIS, qui ne sont disponibles que sur la période récente. De plus, aucune de ces solutions ne prévoit l'importation de données externes, l'uti-

15. Ceux-ci existent pour la période sur laquelle nous travaillons, le canal de Panama sera retiré du réseau pour les dates antérieures à 1914 ; quant au canal de Suez, il est en service depuis 1867.

16. <http://orbis.stanford.edu/>

SAGEO'2016 – Nice, 6-9 décembre 2016

lisation des données collectées par le projet nécessite donc une autre solution. C'est dans ce cadre que la conception d'un nouveau portail cartographique a été envisagée. Son objectif principal étant de réunir et de mettre à disposition les données historiques recueillies, tout en permettant leur analyse et l'extraction d'informations cohérentes et pertinentes. L'interface du portail doit être à la fois pédagogique (pour le grand public), mais également permettre des analyses poussées pour les spécialistes du domaine.

Même si le recours à des solutions SaaS (ArcGisOnline, CartoDB) permettrait de gagner un temps de développement non négligeable, les éditeurs de ces solutions ont une politique de gestion des données incompatible avec nos impératifs. Cela nous conduit donc vers un développement intégral du géoportail, nécessairement plus long, mais également plus flexible.

Le portail comporte une partie de gestion des couches et de création de cartes (dessin, mesure, texte). Un ensemble d'outils d'analyses statistiques permet de lire les données sous forme de graphiques dynamiques, coordonnés avec la carte (linking and brushing). Des indicateurs spécifiques (portant aussi bien sur les ports que sur les trajectoires) offrent la possibilité d'analyser en profondeur les objets sélectionnés. Une timeline est liée aux outils, proposant aux utilisateurs d'effectuer des études statiques (sélection de date) ou diachroniques (sélection de période). L'ensemble des résultats produits (cartes, graphiques, analyses) sont exportables. Le portail est en cours de développement. Il offrira à terme un outil cartographique avancé de visualisation et d'analyse des graphes, appliqué aux flux maritimes. Sa finalisation est prévue pour la fin du mois d'août, pour une diffusion au sein de l'UMR ; début 2017 pour une ouverture au grand public.

4 Perspectives

Nous poursuivons l'élaboration de la base de données, en la complétant par de nouvelles numérisations de données maritimes mais également par des données viaires. Les outils de visualisation et d'analyse du géoportail sont développés pour permettre une analyse fine des graphes de trajectoires, en deux ou trois dimensions. Si le portail est avant tout conçu comme un outil adapté aux spécificités des problématiques de la recherche, une version pédagogique adaptée au grand public est un aboutissement important du projet. Cet objectif s'intègre dans un partenariat avec la Cité de la Mer de Cherbourg, qui souhaite disposer d'un outil permettant aux visiteurs de découvrir l'évolution du trafic maritime mondial au cours des siècles passés. Une des perspectives de ce travail est de permettre le croisement entre visualisation et analyse des flux maritimes et celles de réseaux de transport terrestres (réseaux routiers ou ferrés). Cette analyse conjointe pourra être réalisée à différentes dates grâce aux données d'OpenStreetMap et de VMAP0, en écho à l'analyse de Nelson (2008) et du Joint Research Centre (JRC) de la Commission Européenne sur le calcul d'une accessibilité multimodale mondiale des villes. Ces questions urbaines sont appuyées par la création, au sein du projet World Seastems, d'une base de données temporelles et harmonisées couvrant la période 1890-2010 sur la population des villes et agglomérations du monde.

Bibliographie

Ducruet C. (2015) *Maritime Networks : Spatial Structures and Time Dynamics*, Routledge Studies in Transport Analysis.

Ducruet C. (2016) *La spatialité des réseaux maritimes. Contributions maritimes à l'analyse des réseaux en géographie. Mémoire d'Habilitation à Diriger des Recherches*, Université de Paris 1 Panthéon-Sorbonne.

Ducruet C., Haule S., Ait-Mohand K., Marnot B., Kosowska-Stamirowska Z., Didier L., Coche M.A. (2015a) *Maritime shifts in the contemporary world economy : Evidence from the Lloyd's List corpus, 18-21 c.*, In : Ducruet C. (Ed.), *Maritime Networks : Spatial Structures and Time Dynamics*, Routledge Studies in Transport Analysis, pp. 134-160.

Etienne L., Alincourt E., Devogele T. (2015) *Maritime network monitoring : From position sensors to shipping patterns*. In : Ducruet C. (Ed.), *Maritime Networks : Spatial Structures and Time Dynamics*, Routledge Studies in Transport Analysis, pp. 190-209.

Eyring V., Isaksen I.S.A., Berntsen T., Collins W.J., Corbett J.J., Endresen O., Grainger R.G., Moldanova J., Schlager H., Stevenson D.S. (2010) *Transport impacts on climate and atmosphere : shipping*. *Atmospheric Environment*, 44(37) : 4735- 4771.

Fournier M. (2015) *Venetian supremacy through time. A visualization experiment*. In : Ducruet C. (Ed.), *Maritime Networks : Spatial Structures and Time Dynamics*, Routledge Studies in Transport Analysis, pp. 77-91.

Halpern B.S., Walbridge S., Selkoe K.A., Kappel C.V., Micheli F., D'Agrosa C., Bruno J.F., Casey K.S., Ebert C., Fox H.E., Fujita R., Heinemann D., Lenihan H.S., Madin E.M.P., Perry M.T., Selig E.R., Spalding M., Steneck R., Watson R. (2008) *A global map of human impact on marine ecosystems*. *Science*, 319(5865) : 948-952.

Kaluza P., Kölzsch A., Gastner M.T., Blasius B. (2010) *The complex network of global cargo ship movements*. *Journal of the Royal Society Interface*, 7(48) : 1093- 1103.

Lagesse, C., Mermet, E., Petit, S., Ducruet, C. (2016), *Geovisualisation of the worldwide maritime network over 30 years. Challenges and preliminary results*, *Proceedings of the 2nd International workshop on Maritime Flows and Networks (WIMAKS'16), Session 1 : Géovisualization of Maritime Networks*, Paris 25-26 april 2016.

Marzagalli S. (2015) *Navigocorpus database and eighteenth-century French world maritime networks*. In : Ducruet C. (Ed.), *Maritime Networks : Spatial Structures and Time Dynamics*, Routledge Studies in Transport Analysis, pp. 92-111.

Nelson A. (2008) *Travel time to major cities : A global map of accessibility*. *Global Environment Monitoring Unit, Joint Research Centre of the European Commission*, Ispra, Italy.

SAGEO'2016 – Nice, 6-9 décembre 2016

Scheidel W. (2013) The Shape of the Roman World. Working Paper, Stanford University.

Scheidel W. (2013b) Explaining the Maritime Freight Charges in Diocletian's Prices Edict, *Journal of Roman Archaeology* 26.

Shen G., Wang C., Pulat P.S. (2013) An exploratory analysis and visualization of U.S. global trade patterns through maritime freight movement between U.S. ports and world ports : 1997-2007. International Forum on Shipping, Ports and Airports, Hong Kong, 3-5 juin.

Tournadre J. (2014) Anthropogenic pressure on the open ocean : The growth of ship traffic revealed by altimeter data analysis. *Geophysical Research Letters*, 41(22) : 7924-7932.

Formalisation d'une méthode de filtrage graphique pour simplifier la cartographie des interactions spatiales

Serge Lhomme 1¹

1. Université Paris-Est Créteil (UPEC), Lab'Urba (EA 3482)
61 avenue du Général de Gaulle, 94010 Créteil, France
serge.lhomme@u-pec.fr

RESUME. Cartographier des interactions spatiales se révèle souvent problématique. En effet, les nombreux chevauchements entre les lignes à représenter peuvent rendre la représentation finale illisible. Pour faire face à cette problématique, la cartographie de flux a depuis longtemps recours à des méthodes de filtrage visant à réduire le nombre de liens à cartographier. Plus récemment, des algorithmes de visualisation de graphe, qui proposent des solutions graphiques visant à déplacer ou déformer les arcs, ont aussi été appliqués pour rendre ces représentations plus lisibles. Néanmoins, peu de travaux ont entrepris « d'hybrider » ces deux approches. C'est précisément l'objectif de la méthode de filtrage graphique développée ici qui propose de filtrer l'information à représenter au sein d'une carte en utilisant les valeurs des liens et le nombre d'intersections générées par leur représentation.

ABSTRACT. The visualization of spatial interactions may be problematic. Indeed, several issues limit pictures readability. For instance, too many lines overlaps generate unreadable cartography. It is the spaghetti effect. To address this specific issue, flow mapping is based on filtering methods in order to reduce the number of links to visualize, whereas graph drawing algorithms are based on graphics solutions in order to move or deform the links to map. Only a few works try to merge these two approaches. That is why a graphic filtering method has been developed here. This filtering method is based on links values and number of intersections induced by their visualization. A comparison between results obtained with this new method and those obtained with other filtering methods has been done in order to assess relevance of the graphic filtering method.

MOTS-CLES : Interactions spatiales ; Cartographie de flux ; Visualisation de graphes ; Analyse de réseaux.

KEYWORDS: Spatial interactions; Flow mapping; Graph visualization; Network analysis.

SAGEO'2016 – Nice, 6-9 décembre 2016

1. Contexte de la recherche

Bien qu'elle joue un rôle central dans la définition de la géographie contemporaine, la notion d'interaction spatiale est difficile à définir tant les définitions qui ont pu en être proposé sont variées. Cette notion est alors souvent réduite aux échanges effectifs (humains ou matériels) entre des lieux. Dans ce cadre, les interactions spatiales sont majoritairement cartographiées par des flux. Par simplification mathématique, ces interactions définissent alors un graphe valué. Ces graphes valués ont la particularité d'être spatialisés. Or, lorsque l'on cherche à les visualiser sur une carte, la géolocalisation des lieux (des sommets) peut se révéler contraignante. En effet, comme le nombre de lignes reliant ces sommets est fonction du nombre de lieux, si le réseau est dense, les chevauchements entre ces lignes peuvent rendre la représentation illisible : c'est l'effet spaghetti (Bonin et al., 2013 ; Bahoken, 2015).

Pour illustrer la problématique traitée, cette recherche s'appuie sur les données de mobilité professionnelle produites et diffusées par l'Insee (2010). Ces données recensent les flux domicile-travail entre les communes françaises, lorsque ces flux sont supérieurs à 100 personnes. A partir de ces données, le volume des flux recensés pour chaque couple de communes a été calculé. Le résultat obtenu peut alors être modélisé sous la forme d'un graphe spatial, valué et non orienté, dont la seule visualisation des liens (réduits à de simples lignes droites reliant les centrides des communes) produit un résultat très insatisfaisant (Fig. 1).

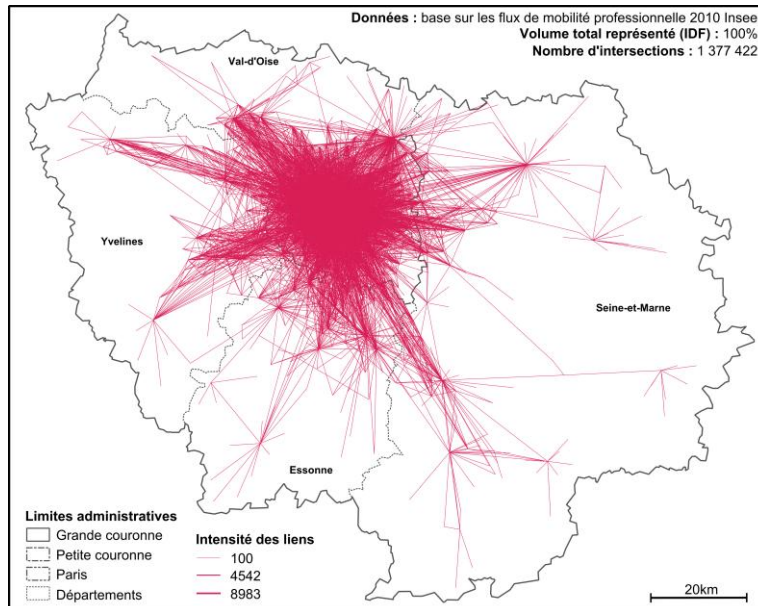


FIGURE 1. Cartographie des interactions spatiales entre les communes d'Ile-de-France à partir des flux domicile-travail.

Afin de faciliter la représentation des interactions spatiales, il est par conséquent pertinent d'utiliser des méthodes qui limitent l'effet spaghetti. Dans la pratique, il est alors souvent nécessaire d'avoir recours à des techniques de filtrage qui ont pour finalité de limiter le nombre de liens à afficher. Ces techniques de filtrage sont particulièrement utilisées dans le domaine de la cartographie des flux et ont comme point commun d'utiliser les valeurs des liens pour filtrer l'information. Les liens caractérisés par des valeurs fortes doivent alors être affichés, tandis que les liens caractérisés par des valeurs faibles doivent plutôt être masqués. La méthode des flux dominants de Nystuen et Dacey (1961) est par exemple l'une des méthodes les plus utilisées. Le principal défaut de cette méthode constitue aussi son principal atout : cette méthode simplifie énormément la structure relationnelle étudiée.

Ainsi, d'autres méthodes de filtrage ont été développées afin notamment d'assouplir les contraintes de la méthode des flux dominants. Par exemple, les méthodes de type « flux majeurs » retiennent les flux dont l'intensité représente plus de x% du volume total des flux (méthode 1) ou les x premiers flux par sommet (méthode 2). Néanmoins, se fonder seulement sur un pourcentage du volume total crée le plus souvent un grand nombre de sommets isolés, car les réseaux spatiaux se caractérisent fréquemment par une surreprésentation des liens de valeurs faibles (Fig. 2) et retenir x liens par sommet engendre très vite des cartes peu lisibles (Fig. 3). C'est pourquoi, des approches statistiques originales de transformation des données ont parfois été appliquées pour résoudre les problèmes liés à la surreprésentation des valeurs faibles (Feng-Chong, 2012).

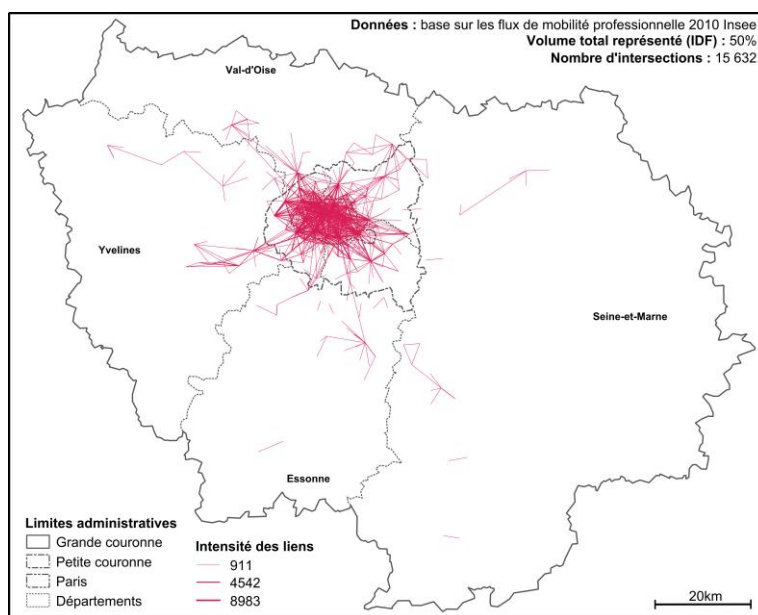


FIGURE 2. *Méthode de filtrage utilisée : Conservation de 50% du volume total porté par les liens.*

SAGEO'2016 – Nice, 6-9 décembre 2016

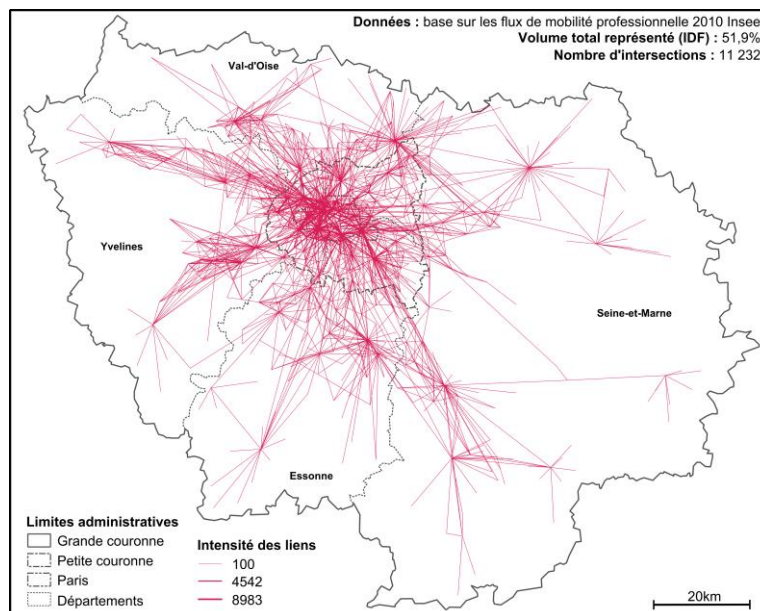


FIGURE 3. Méthode de filtrage utilisée : 5 liens maximum par sommet.

Par ailleurs, des algorithmes ont été développés pour faciliter la visualisation des graphes valués. Ces algorithmes sont *a priori* peu adaptés pour cartographier des interactions spatiales. En effet, ces algorithmes de visualisation sont pour la plupart fondés sur différentes méthodes de placement des sommets (Harel et Koren, 2002). Pour un graphe spatial, l'application de ces algorithmes conduira donc à une disposition différente des sommets et des arcs par rapport à leur localisation géographique initiale. Quoi qu'il en soit, cette famille de représentations connaît actuellement un succès indéniable. Ce succès est principalement dû à la qualité graphique des représentations obtenues. Cette qualité graphique est le résultat des critères de lisibilité retenus dont un semble communément partagé : les chevauchements entre les liens doivent être limités. De surcroit, dans ce cadre, les méthodes d'optimisation sont centrales, car elles permettent de déterminer concrètement la meilleure disposition des sommets et des arcs (Harel et Koren, 2002).

Quelle que soit la puissance heuristique des représentations obtenues par le déplacement des sommets et des arcs, ces représentations peuvent rester insatisfaisantes. C'est pourquoi, des méthodes cherchent à rendre ces représentations encore plus lisibles en déformant les arcs (Holten, 2006). Les liens ne sont alors plus représentés par de simples lignes droites, mais par des courbes. La méthode probablement la plus utilisée est celle du « *edge bundling* » qui s'appuie sur un principe d'agrégation des liens (Holten, 2006). Cette technique du « *edge bundling* » est particulièrement intéressante pour cartographier des flux puisqu'elle préserve la

géolocalisation initiale des lieux et conserve l'ensemble des liens (Lambert et al., 2010 ; Bonin et al., 2013). Néanmoins, le problème majeur du « *edge-bundling* » est probablement le suivant : pour un même jeu de données, les résultats obtenus diffèrent sensiblement selon l'algorithme choisi. Par conséquent, les interprétations données à une même problématique peuvent varier sans qu'il soit possible de déterminer simplement quelle est la représentation la plus pertinente. Les différentes représentations obtenues dans le cadre de l'étude des migrations américaines illustrent parfaitement cette problématique (Cui et al., 2008 ; Lambert et al., 2010 ; Ersoy et al., 2011).

2. Formalisation d'une méthode de filtrage graphique

L'analyse des approches existantes montre qu'il est pertinent de chercher des méthodes originales pour faciliter la cartographie de graphes valués représentant des interactions spatiales. En l'occurrence, deux principes issus des méthodes de visualisation de graphe sont retenus dans cette recherche : pour obtenir une représentation lisible les intersections (superpositions, chevauchements) entre les liens doivent être minimisées (ce principe est aussi partagé par la cartographie de flux) ; la recherche de la meilleure représentation repose sur des méthodes d'optimisation. De surcroît, un principe issu du domaine de la cartographie de flux est aussi retenu : il est pertinent de limiter le nombre de liens à représenter (*i.e.* filtrer les liens) tout en cherchant malgré tout à représenter le volume total le plus important. L'idée de déformer les liens est ici laissée de côté pour être éventuellement utilisée une fois l'opération de filtrage exécutée. Les liens sont donc considérés comme de simples lignes droites reliant deux sommets géolocalisés.

Formellement, la méthode de filtrage développée s'applique à la représentation des interactions spatiales pouvant être modélisées sous la forme d'un graphe G caractérisé par un ensemble V de sommets et un ensemble E d'arcs : $G = (V, E)$. Ce graphe doit être valué, chaque arc reliant un sommet i à un sommet j doit ainsi être caractérisé par une valeur non nulle $w(i,j)$. Le nombre total d'arcs du graphe G est égal à l et la somme des « valuations » de ces arcs vaut $W(G)$. La méthode développée vise alors à identifier un sous-graphe G' de G tel que l'ensemble des sommets V' de G' soit identique à V et l'ensemble des arcs E' de G' soit inclus dans E : $G' = (V', E')$ avec $V' = V$ et $E' \subset E$. Le nombre d'arcs de G' est égal à l' (avec $l' \leq l$) et la somme des poids du sous-graphe G' vaut $W(G')$ (avec $W(G') \leq W(G)$). La « précision » induite par la méthode de filtrage peut dès lors s'évaluer ainsi :

$$P(G) = \frac{W(G')}{W(G)} \quad (1)$$

Concrètement, plus le nombre d'arcs représentés sera élevé, plus la représentation obtenue sera considérée comme « précise ». Néanmoins, cette précision génère de nombreuses intersections rendant la représentation illisible. C'est pourquoi, la méthode développée repose sur la définition d'une valeur N correspondant au nombre d'intersections générées par la représentation de G' :

SAGEO'2016 – Nice, 6-9 décembre 2016

$\text{Plot}(G') = N$. Le « bruit » induit par ces intersections peut alors être quantifié à l'aide d'une loi de puissance du type :

$$B(G') = 1 - \frac{1}{(N+1)^\alpha} = 1 - (N+1)^{-\alpha} \quad (2)$$

Plus N est élevé, plus la représentation sera « bruitée » et par conséquent considérée comme étant peu lisible. Le coefficient α est quant à lui compris entre 0 et 1. Plus ce coefficient est élevé, plus les intersections seront considérées comme générant du « bruit ». L'objectif de la méthode de filtrage développée est donc de maximiser la précision $P(G')$ tout en minimisant le bruit $B(G')$. Pour cela, il est possible de définir une « fonction objectif » correspondant à l'information apportée par une représentation de G' :

$$I(G') = P(G') - B(G') \quad (3)$$

La meilleure représentation G^* correspond *in fine* au sous-graphe G' de G qui maximise l'information apportée :

$$I(G^*) = \text{Max}[P(G') - B(G')] \quad (4)$$

3. Application de la méthode

Pour trouver G^* (pour une valeur α donnée), il est possible de s'appuyer sur un algorithme gloutin qui, partant d'un ensemble vide, va à chaque itération retenir l'arc qui apporte le plus d'information à la représentation. A aucun moment l'algorithme ne remettra en question la sélection d'un arc. L'algorithme s'arrête dès que l'ajout d'un arc engendre nécessairement une réduction de l'information apportée à la représentation. Une analyse statistique des résultats obtenus montre que pour un même volume total représenté, le nombre d'intersections est sensiblement plus faible avec la méthode développée qu'avec les autres méthodes (Fig. 4). Selon ce critère, les représentations obtenues avec la méthode de filtrage graphique développée doivent donc apparaître plus lisibles que celles obtenues avec les autres méthodes. L'analyse du nombre d'arcs retenus au sein des représentations semble montrer que la méthode développée réalise un certain compromis entre les deux méthodes de filtrage existantes étudiées ici (Fig. 5).

Plus précisément, le choix du coefficient α constitue l'enjeu principal de la méthode développée. Par exemple, si le coefficient α est supérieur au taux du volume total correspondant à la plus forte valuation d'un arc, alors aucune intersection ne sera représentée. Dans cette recherche, la plus forte valeur d'un arc correspond à l'intensité de la relation existant entre le 8ème et le 15ème arrondissement de Paris : cette valeur est de 8983, ce qui représente 0,31% du volume total. Ainsi, pour une valeur α égale à 0,003 la représentation obtenue ne contient aucune intersection et tous les sommets sont reliés au réseau par au moins un arc. Au contraire, lorsque la valeur α tend davantage vers zéro, tous les liens sont

représentés. La meilleure représentation est probablement obtenue lorsque la valeur de α est comprise entre ces deux extrêmes.

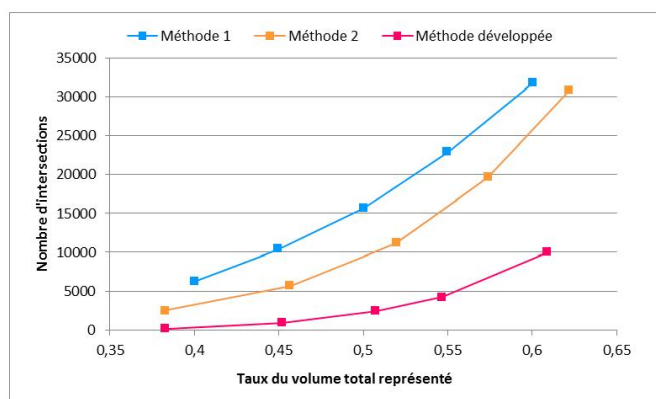


FIGURE 4. Comparaison du nombre d'intersections générées par chaque méthode.

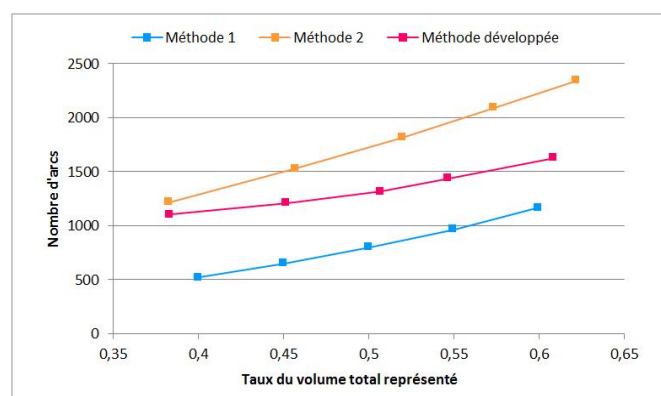
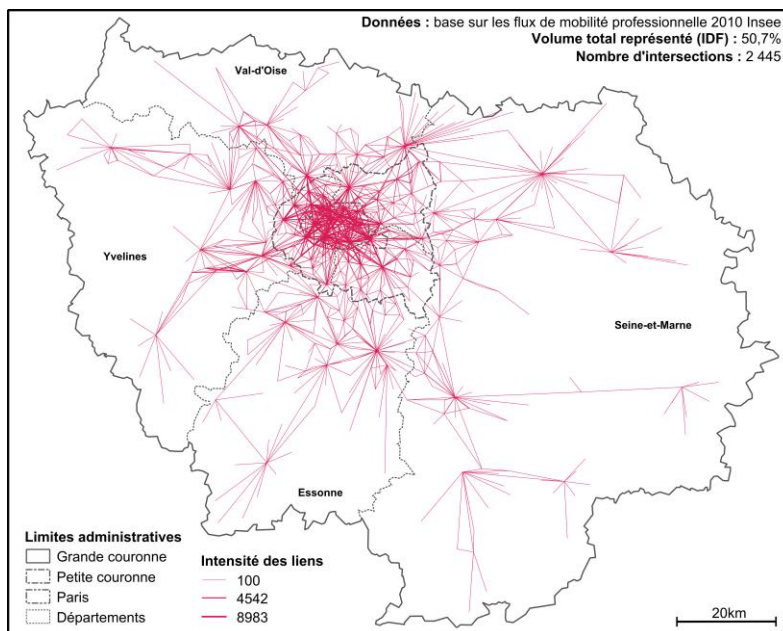


FIGURE 5. Comparaison du nombre d'arcs retenus.

Dès lors, pour juger l'efficacité de la méthode développée, il est possible de comparer visuellement les représentations obtenues par la méthode de filtrage graphique avec celles obtenues par d'autres méthodes de filtrage. Pour cela, il convient de comparer les représentations dont les volumes totaux représentés sont comparables (ici environ 50% du volume total) (Fig. 2, 3, 6). Il apparaît alors que la méthode développée produit des représentations qui semblent réaliser un certain compromis entre les méthodes qui retiennent les arcs dont l'intensité représente plus de $x\%$ du volume total (méthode 1) et celles qui retiennent les x premières intensités par sommet (méthode 2).

SAGEO'2016 – Nice, 6-9 décembre 2016

FIGURE 6. Exemple d'application de la méthode de filtrage graphique ($\alpha = 0,00017$).

Ainsi, la méthode développée permet d'obtenir une représentation cartographique du réseau étudié qui ne se réduit pas à Paris et à sa petite couronne, contrairement à la méthode 1. Dans le même temps, cette méthode limite le nombre d'intersections, ce qui permet d'identifier que Paris et sa petite couronne concentrent les interactions, ce que la méthode 2 masque en partie puisqu'elle surreprésente les liens de faibles valeurs.

4. Limites et perspectives

La méthode de filtrage développée permet de limiter le nombre d'intersections tout en maximisant le volume total représenté. Néanmoins, la détermination de la meilleure représentation obtenue (et du coefficient α correspondant) repose encore sur une appréciation visuelle subjective. La comparaison entre les représentations obtenues pour différentes valeurs de α montre bien qu'il est difficile de déterminer quelle représentation est la meilleure (Fig. 7). De surcroît, la méthode proposée tend à surreprésenter les liens courts au détriment des liens les plus longs, la représentation de ces derniers étant susceptible de générer beaucoup d'intersections. C'est pourquoi une analyse des liens qui n'ont pas été représentés est pertinente. Néanmoins, les autres méthodes de filtrage tendent aussi à surreprésenter les liens courts, sinon elles deviennent vites illisibles (Fig. 7).

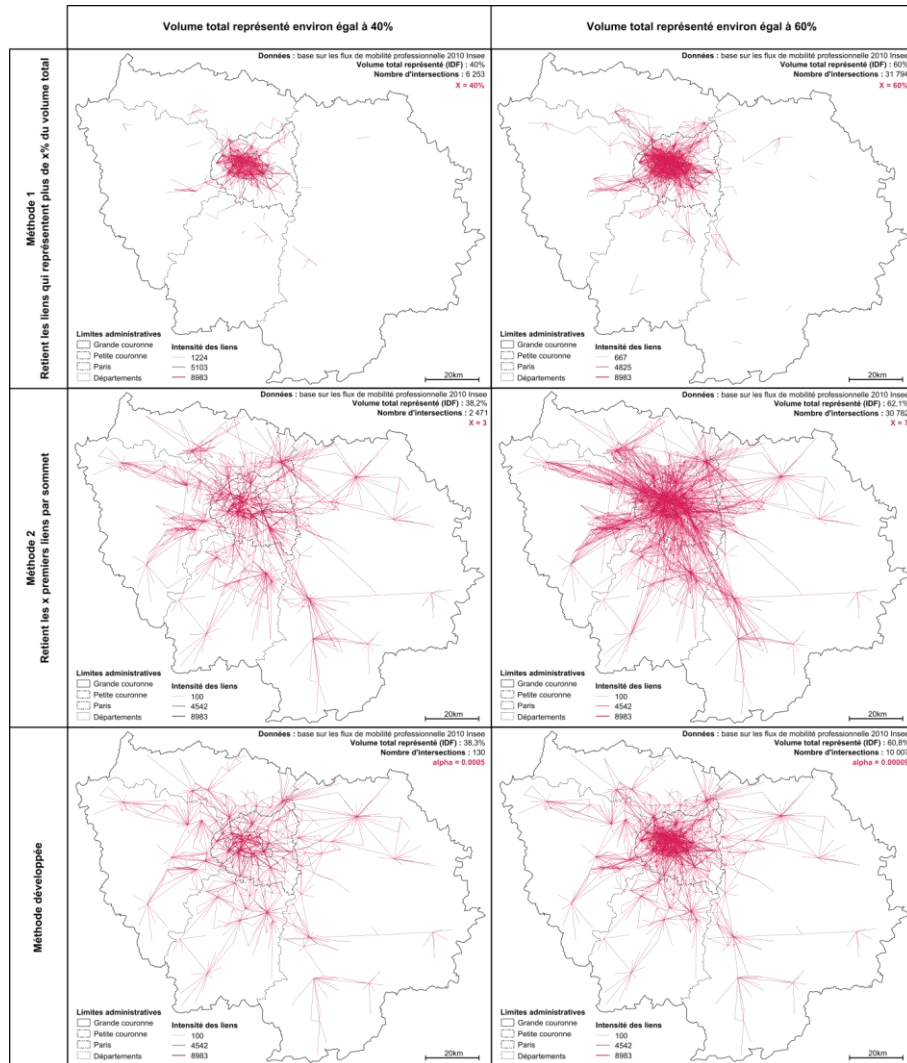


FIGURE 7. Comparaison de différentes méthodes de filtrage en fonction du volume total représenté.

En outre, mélanger des aspects graphiques (le nombre d'intersections) et statistiques (la valeur des liens) nécessite de prendre du recul vis-à-vis des résultats obtenus. En effet, des structures réticulaires similaires n'auront pas la même signification dans un espace dense et dans un espace peu dense. L'interprétation des représentations est par conséquent plus complexe.

SAGEO'2016 – Nice, 6-9 décembre 2016

Les deux derniers points évoqués constituent les principales perspectives de cette recherche, l'objectif étant bien entendu d'atténuer ces biais.

Enfin, les représentations obtenues peuvent apparaître encore insatisfaisantes car trop peu lisibles. Il faut préciser ici que représenter l'intensité des flux domicile-travail entre les communes d'Ile-de-France est un défi qui n'est pas trivial : la représentation de départ comporte en effet plus de 6000 arcs. Cette problématique questionne naturellement le géographe sur la taille des unités géographiques et celle de la zone analysée. Dans ce cadre, la pertinence des approches multi-niveaux (Amiel et al., 2005) ou des visualisations interactives (Jégou et Maisonobe, 2013) ne sont plus à démontrer. Néanmoins, développer des méthodes de filtrage graphique qui reposent sur des critères intelligibles constitue un champ de recherches original à investiguer.

Références

- Amiel M., Mélançon G., Rozenblat C., (2005). Réseaux multi-niveaux : l'exemple des échanges aériens mondiaux de passagers. *M@ppemonde*, n°79.
- Bahoken F. (2015). Eléments pour une représentation (carto)graphique des matrices de flux. *M@ppemonde*, n° 115.
- Bonin O., Bahoken. F., Mermet S., (2013). Cartod/mapod : Visualiser la demande locale de transport. *M@ppemonde*, n° 112.
- Cui W., Zhou H., Qu H., Wong P. C., Li X., (2008). Geometry based edge clustering for graph visualization. *Transactions on Visualization and Computer Graphics*, vol. 14, n° 6, pp. 1277-1284.
- Ersoy O., Hurter C., Paulovich F. V., Cantareira G., Telea A., (2011), Skeleton-based edge bundling for graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, n° 12, pp. 2364-2373.
- Fen-Chong J. (2012). *Organisation spatio-temporelle des mobilités révélées par la téléphonie mobile en Ile-de-France*. PhD thesis, Université Panthéon-Sorbonne - Paris I.
- Harel D., Koren Y., (2002). Fast drawing of large graphs. *Journal of Graph Algorithms and Applications*.
- Holten D., (2006). Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, Vol. 12, pp. 741-748.
- Jégou L., Maisonobe M., (2013). Les collaborations scientifiques mondiales, 1999-2008, une application web de géovisualisation. *M@ppemonde*, n° 112.
- Lambert A., Bourqui R., Auber D., (2010). Winding Roads: Routing edges into bundles. *Computer Graphics Forum*, Wiley, Vol. 29, n° 3, pp. 853-862.
- Nystuen J. D., Dacey M. F., (1961). A graph theory interpretation of nodal regions. *Regional Science*.

Responsables de la publication :

Karine Emsellem, Diego Moreno Sierra, Christine Voiron-Canicio, Didier Josselin

Éditeur : UMR 7300 ESPACE – Avignon, Case 41, 74 rue L. Pasteur, 84029 Avignon Cedex, France

Collection : Actes Avignon - ISBN: 978-2-9105-4511-3 - Décembre 2016

Design et Mise en page : N. Brachet (UMR ESPACE – CNRS)

<https://sageo2016.sciencesconf.org>