



HAL
open science

Actes de la conférence Spatial Analysis and GEOmatics (SAGEO) 2023

Thierry Badard, Jacynthe Pouliot, Matthieu Noucher, Marlène Villanova

► **To cite this version:**

Thierry Badard, Jacynthe Pouliot, Matthieu Noucher, Marlène Villanova. Actes de la conférence Spatial Analysis and GEOmatics (SAGEO) 2023. Spatial Analysis and GEOmatics (SAGEO 2023), 2023, 978-2-9822061-0-6. halshs-04484449

HAL Id: halshs-04484449

<https://shs.hal.science/halshs-04484449>

Submitted on 29 Feb 2024

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.

Copyright



Actes de la conférence Spatial Analysis and GEOmatics (SAGEO) 2023.

Éditeurs : Thierry Badard, Jacynthe Pouliot, Matthieu Noucher, Marlène Villanova-Oliver

ISBN : 978-2-9822061-0-6

Conférence coorganisée par le Centre de recherche en données et intelligence géospatiales (CRDIG) de l'Université Laval et le groupement de recherche CNRS MAGIS



Préambule

La conférence Spatial Analysis and GEomatics (SAGEO) 2023 s'est tenue à Québec, Qc, Canada du 5 au 9 juin 2023, sur le campus de l'Université Laval. SAGEO est la conférence internationale francophone en géomatique soutenue par le groupe de recherche CNRS MAGIS (<https://gdr-magis.cnrs.fr/>). L'objectif de SAGEO est de présenter les derniers travaux de recherche en géomatique et en analyse spatiale.

Cette édition de SAGEO a été organisée par le Centre de Recherche en Données et Intelligence Géospatiales (CRDIG) de l'Université Laval.

En plus de conférences, SAGEO a permis la tenue de 7 ateliers thématiques et l'organisation de l'assemblée générale du GDR MAGIS. Des visites dans 2 institutions québécoises (Communauté Métropolitaine de Québec et Centre de simulation et d'expertise maritime) en lien avec les technologies géospatiales ont été organisées.

Thierry Joliveau était le conférencier invité de cette édition de la conférence et a présenté une synthèse des travaux qu'il a réalisés durant sa riche et intense carrière au sujet des « techniques, pratiques et imaginaires de la géo-numérisation. Les objets culturels en géomatique ».

La liste des thèmes d'intérêt pour cette édition de la conférence était :

- Capteurs et Internet des objets
- Acquisition de données spatiales : topographie, positionnement et navigation par satellites, géo-senseurs...
- Géographie participative (VGI), Collaboration, participation, médiation, modélisation d'accompagnement
- Modélisation conceptuelle de données spatiales et temporelles
- Structure de données spatiales et algorithmique géométrique, modélisation des relations spatiales
- Modélisation, représentations et dynamiques urbaines
- Modélisation et simulation de systèmes complexes, couplage de modèles
- Intégration, fusion multi-échelle, généralisation
- Incertitude, modélisation et gestion de l'incertitude, propagation des erreurs, analyse de sensibilité
- Web sémantique et ontologies spatio-temporelles
- Données spatiales massives, Spatial Big Data
- Bases et entrepôts de données spatiales, fouille/forage de données, (S)OLAP...
- Géospatial et apprentissage machine
- Indexation, traitement et optimisation de requêtes spatiales et temporelles
- Infrastructures de données spatiales, interopérabilité, métadonnées
- Services web géographiques
- Raisonnement spatial, raisonnement temporel

- Fouille de données spatiales et temporelles
- Extraction et recherche d'information géographique
- Calculs hautes performances en géomatique
- Algorithmes distribués et parallèles pour l'Information géographique
- Géo-informatique en nuage
- Géolocalisation et analyse des déplacements
- Référentiels spatiaux : géodésie, gravimétrie, projections cartographiques...
- Télédétection, photogrammétrie et traitement d'images, LiDAR, observation de la Terre
- Analyse d'images et de vidéos, traitement de flux d'informations géographiques
- Géo-visualisation et cartographie dynamique, langages visuels, Webmapping
- Cognition spatiale : perception et appropriation de l'espace
- Mobilité des personnes, des objets et des systèmes, analyse de trajectoires, transports et flux
- Applications mobiles spatio-temporelles et services géolocalisés
- SIG 3D, jumeaux numériques
- SIG temps réel
- Analyse spatiale, géo-statistiques, modélisation de surfaces...
- Analyse d'erreurs, d'incertitudes et de sensibilité des données spatiales et temporelles
- Aide à la décision spatiale
- Dynamiques spatiales et Interactions
- Observatoire territorial, observatoire scientifique, observatoire de l'information géographique
- Approches critiques des sciences de l'information géographique
- etc.

Les domaines d'application concernaient :

- Écologie du paysage, environnement, gestion des ressources naturelles
- Aménagement et Gouvernance du Territoire
- Villes et communautés intelligentes
- Développement durable
- Urbanisme
- Energie
- Transport et mobilité
- Société et usages de l'information géographique
- Planification et géo-prospective
- Mers et océans, Gestion des zones côtières et des littoraux
- Prévention des risques et vulnérabilités
- Santé, Épidémiologie
- Changements climatiques
- Enseignement de la géomatique et de l'analyse spatiale

Tenue en format hybride (en présence et accès à distance possible), la conférence a rassemblé 98 personnes (80 en présence, 18 à distance), ce qui est un beau succès pour une édition de la conférence organisée à l'étranger.

Le comité de lecture composé de 45 experts de France et de Québec a eu à évaluer les 30 articles qui ont été soumis. Chaque article a été évalué par 2 membres du comité de lecture et 18 articles ont finalement été acceptés (soit un taux de 60% d'acceptation).

Les auteurs des 8 meilleurs articles ont été invités à soumettre une version étendue de leur papier pour considération dans un volume dédié aux Sciences de l'information géographique et les mesures environnementales du Journal of Interdisciplinary Methodologies and Issues in Science (JIMIS) : <https://jimis.episciences.org/page/vol-12>.

Les détails et le programme de la conférence SAGEO 2023 sont disponible en ligne à l'adresse <https://crdig.ulaval.ca/sageo2023/>.

Thierry Badard, Jacynthe Pouliot, Matthieu Noucher, Marlène Villanova-Oliver
Responsables du comité scientifique de SAGEO 2023

Comité d'organisation

- Éric Guilbert
- Frédéric Hubert
- Sonia Rivest
- Alexandre Gourret
- Guetchine Gaspard

Comité scientifique

- Thierry Badard, Université Laval
- Jacynthe Pouliot, Université Laval
- Matthieu Noucher, CNRS
- Marlène Villanova-Oliver, Université Grenoble Alpes

Comité de programme

- Nathalie Abadie (IGN, LATSIG)
- Mireille Batton-Hubert (LIMOS, Ecole Nationale Supérieure des Mines de Saint-Etienne)
- Yacine Bouroubi (Université de Sherbrooke)
- Carmen Brando (EHESS)
- Anne-Christine Bronner (CNRS, SAGE)
- Christophe Claramunt (Ecole navale de Brest, IRENav)
- Rodolphe Devillers (IRD, Espace-DEV)
- Jean Dubé (CRDIG, Université Laval)
- Bertrand Duménieu (EHESS)
- Pauline Dusseux (Université Grenoble Alpes, PACTE)
- Laurent Etienne (L@BISEN, Yncrea Ouest, Brest)
- Grégoire Feyt (Université Grenoble Alpes, PACTE)
- Pierre Gautreau (Université Paris 1 Panthéon Sorbonne, PRODIG)
- Jérôme Gensel (Université Grenoble Alpes, LIG)
- Mickaël Germain (Université de Sherbrooke)
- Marc Gervais (CRDIG, Université Laval)
- Charles Gignac (CRDIG, Université Laval)
- Arnaud le Guilcher (IGN, LATSIG)
- Françoise Gourmelon (CNRS, LETG)
- Baptiste Hautdidier (INRAE, ETTIS, Bordeaux)
- Pierre-Alexis Herrault (Université de Strasbourg, LIVE)
- Saeid Homayouni (INRS)
- Thierry Joliveau (Université St Etienne, EVS)
- Didier Josselin (CNRS, ESPACE, Avignon)

- Christian Larouche (CRDIG, Université Laval)
- Maud Loireau (IRD, Espace-DEV, Montpellier)
- Eric Masson (Université de Lille, TVES)
- Hélène Mathian (CNRS, EVS)
- Sehl Mellouli (CRDIG, Université Laval)
- Boris Mericskay (Univ. Rennes 2, ESO)
- André Miralès (INRAE, TETIS, Montpellier)
- Ludovic Moncla (INSA Lyon, LIRIS)
- Mir Mostafavi (CRDIG, Université Laval)
- Ana-Maria Olteanu-Raimond (IGN, LATSIG)
- Christine Plumejeaud (CNRS, Migrinter, Poitiers)
- François Pinet (INRAE, TSCF)
- Juste Rimbault (IGN, LATSIG)
- Sébastien Rey-Coyrehourcq (CNRS, IDEES)
- Stéphane Roche (CRDIG, Université Laval)
- Francis Roy (CRDIG, Université Laval)
- Anne Ruas (Université Gustave Eiffel, COSYS)
- Sylvie Servigne (INSA Lyon, LIRIS)
- David Sheeren (Institut National Polytechnique de Toulouse, DYNAFOR)
- Myriam Servières (Ecole centrale de Nantes, AAU)
- Brice Trouillet (Nantes Université, LETG)

Tables des matières

HedgeTools : une boîte à outils pour caractériser automatiquement les haies en milieu agricole. <i>David Sheeren, Gabriel Marques, Lewis Villierme, Jean-Baptiste Boissonnat, Guillaume Guebin, Marc Lang et Claude Monteil</i>	9
Caractérisation spatiale multiscalaire des agricultures contribuant à l’approvisionnement alimentaire de proximité. Propositions méthodologiques pour l’accompagnement de l’action publique de planification alimentaire. <i>Esther Sanz Sanz, Claude Napoléone et Michel Mouléry</i>	23
Proposition méthodologique pour identifier la limite spatiale ville-agriculture : Caractérisation de la rugosité du contour. <i>Michel Mouléry, Didier Josselin, Esther Sanz Sanz et Claude Napoléone</i>	38
ORL : une Ontologie des Relations de Localisation pour l’aide à la localisation de victimes en montagne lors d’un appel aux secours. <i>Mattia Bunel, Cécile Duchêne and Ana-Maria Olteanu-Raimond</i>	52
Qaujikairit, un outil d’information et d’alerte précoce sur les événements météorologiques pour les communautés nordiques du Nunavik. <i>Charles Gignac, Thierry Badard and Stéphane Voyer-Boutet, Jacynthe Pouliot et Richard Fortier</i>	64
Developing an Ontology-Based Framework to Select Geospatial Data for Natural Hazards Early Warning Systems (Case study: Nunavik, Québec, Canada). <i>Amirhossein Vahdat, Jacynthe Pouliot, Thierry Badard and Richard Fortier</i>	80
Cartographie lidar des biotopes terrestres, marins et intertidaux de l’atoll de Tetiaroa. <i>Benoit Stoll, Poeiti Tuheiava, Moana Badie and Lila De Oliveira</i>	95
Cartographie de la diversité spécifique forestière des milieux tempérés à partir d’imagerie hyperspectrale. <i>Marc Lang, Max Ferriere, Florian De Boissieu, Xavier Briottet, Sophie Fabre, David Sheeren and Jean-Baptiste Féret</i>	107
Cartographie des trajectoires maritimes incertaines du XVIIIème siècle. <i>Christine Plumejeaud-Perreau and Bernard Pradines</i>	122
Variation d’estimation de température selon les échelles administratives. <i>Didier Josselin, Matthieu Vignal, Nicolas Viaux, Delphine Blanke and Céline Lacaux</i>	137
Identifier le domaine public et son évolution : méthodologie d’analyse spatio-temporelle par traitements géomatiques en creux du cadastre. <i>Thibault Lecourt, Laure Casanova Enault and Didier Josselin</i>	152
Optimiser la localisation des services publics : Développement méthodologique et application à la Région SUD. <i>Quentin Godoye and Cyrille Genre-Grandpierre</i>	165
CentipedeRTK Le Real Time Kinematic collaboratif, low-cost et open source. <i>Ancelin Julien, Ladet Sylvie and Heintz Wilfried</i>	184

Unlocking Insights into Business Trajectories with Transformer-based Spatio-temporal Data Analysis. <i>Muhammad Arslan and Christophe Cruz</i>	198
Estimer un potentiel de regroupement de mobilité à partir de données spatio-temporelles. <i>Louissette Garcin, Didier Josselin, Sonia Chardonel, Mathilde Vernet and Jean-Baptiste Chesneau</i>	209
Cartographier pour le web avec la bibliothèque JavaScript `bertin`. <i>Nicolas Lambert, Timothee Giraud, Matthieu Viry and Ronan Ysebaert</i>	221
La partie immergée de l'information géographique : Analyse critique à travers le cas de la planification spatiale marine. <i>Juliette Davret and Brice Trouillet</i>	237
Pour un dialogue entre la géomatique et les humanités numériques : propositions pour contribuer à des pratiques géo-numériques critiques. <i>Juliette Morel and Baptiste Hautdidier</i>	251

HedgeTools : une boîte à outils pour caractériser automatiquement les haies en milieu agricole

David Sheeren¹, Gabriel Marquès^{1,2}, Lewis Villierme^{1,3}, Jean-Baptiste Boissonnat^{1,2}, Guillaume Guébin¹, Marc Lang¹, Claude Monteil¹

¹ Université de Toulouse, INRAE, UMR DYNAFOR
Av. de l'Agrobiopôle, BP 32607, Auzeville Tolosane, 31326 Castanet Tolosan
prenom.nom@toulouse-inp.fr

² TerraNIS
12 Av. de l'Europe, 31520 Ramonville-Saint-Agne

³ Direction du Système d'Information de Polynésie Française
BP 2551, 98713 Papeete Tahiti, Polynésie Française

RÉSUMÉ. La haie occupe une place centrale dans les approches paysagères de l'agroécologie à travers les fonctions de production, de régulation et de protection qu'elle assure. Toutefois, l'opérationnalité de ses fonctions dépend des caractéristiques de la haie, au-delà de son existence dans un territoire. S'il existe différentes techniques permettant d'extraire automatiquement les haies dans des images, il n'existe pas de boîte à outils dédiée à l'évaluation de leur multifonctionnalité. Dans cet article, nous présentons HedgeTools, un nouveau plug-in sous QGIS conçu pour caractériser les haies selon différents points de vue et différents niveaux d'analyse.

ABSTRACT. Hedgerow plays a key role in agro-ecological landscape approach through the production, regulation and protection functions they provide. However the operability of these functions depends on the characteristics of the object, beyond its existence in the landscape. Although various techniques exist to extract automatically hedgerows from imagery, there is no tools dedicated to the evaluation of their multifunctionality. In this paper, we present HedgeTools, a new QGIS plug-in designed to characterize hedgerows from different points of view and different levels of analysis.

MOTS-CLÉS : Haie, point de vue, représentation multiple, topologie, analyse spatiale, SIG.

KEYWORDS: Hedgerow, viewpoint, multiple representation, topology, spatial analysis, GIS.

1. Introduction

Les haies jouent un rôle clé dans les paysages agricoles. Au-delà de leur fonction de corridor écologique au sein de la trame verte, elles constituent un important refuge pour la biodiversité et rendent de nombreux services écosystémiques : protection des cultures et animaux d'élevage, régulation des ravageurs, rôle épurateur de l'eau, lutte contre l'érosion, protection contre le vent (Weninger *et al.*, 2021). Dans le contexte actuel de changement climatique, les structures bocagères représentent un puit de carbone non négligeable, à travers leur biomasse aérienne et racinaire et leur capacité à séquestrer du carbone dans les sols (Axe *et al.*, 2017). Elles constituent un des leviers reconnus d'atténuation du réchauffement climatique avec des marges de progression importantes (intensification des plantations et amélioration des pratiques d'entretien, usage en agroforesterie, usage comme bois d'œuvre ou bois énergie en remplacement des énergies fossiles et avec un approvisionnement local...). Elles contribuent aussi à complexifier les paysages agricoles et augmenter leur hétérogénéité qui est une caractéristique nécessaire pour mettre en œuvre des pratiques agroécologiques (Estrada-Carmona *et al.*, 2022). Cependant, la surface des haies a largement diminué depuis les années 1950 avec la modernisation de l'agriculture. En France, la perte est estimée à près de 70 % avec un rythme d'érosion d'environ 23 500 km/an sur la période 2017-2021 (d'après un calcul effectué par l'AFAC-Agroforesteries). Par ailleurs, leurs caractéristiques structurelles et fonctionnelles sont encore mal inventoriées ce qui rend difficile de quantifier et de spatialiser sur de larges étendues les services environnementaux qu'elles assurent.

Aujourd'hui, les politiques publiques environnementales européennes et nationales soutiennent la protection et la replantation de haies mais il n'existe pas de solution permettant d'évaluer leur impact. Le premier verrou réside dans la capacité à fournir des données à jour et qui puissent rendre compte des évolutions du bocage sur les territoires. Le deuxième verrou réside dans la capacité à dériver des indicateurs qui rendent compte de la structure et du rôle des haies. Par exemple, l'estimation du potentiel de séquestration du carbone d'une haie dépend de nombreux facteurs comme sa hauteur, sa largeur, sa densité, sa biomasse, son âge, sa composition. Il y a ainsi un besoin d'affiner la cartographie des haies mais aussi, de suivre leur évolution et de qualifier leurs différentes propriétés.

En France, l'institut national de l'information géographique et forestière (IGN) et l'Office Français de la Biodiversité (OFB) ont récemment mis en place le dispositif national de suivi des bocages en créant un nouveau référentiel : la BD Haie (Morin *et al.*, 2019). Cette base de données cartographiques est obtenue en combinant les objets haies de la BDTopo v2 IGN et celles des surfaces non agricoles du Registre Parcellaire Graphique (RPG). Elle vise à lever le premier verrou évoqué en apportant une meilleure connaissance sur le bocage actuel, avant de suivre son évolution.

Malgré tout l'intérêt de ce nouveau dispositif, la représentation cartographique proposée issue de sources hétérogènes et proposant une géométrie linéaire ne répond pas toujours aux besoins des professionnels de l'arbre hors forêt. Par ailleurs, le

nombre d'indicateurs qualifiant les haies est très limité ce qui ne permet pas d'appréhender et de spatialiser la multifonctionnalité des objets (second verrou évoqué).

L'examen de la littérature scientifique fait apparaître l'existence de plusieurs méthodes pour extraire automatiquement les haies à partir d'images (Sheeren *et al.*, 2012 ; Fauvel *et al.* 2014 ; Merciol *et al.*, 2019 ; Lucas *et al.*, 2019). Il montre aussi l'absence d'outils dédiés spécifiquement à leur caractérisation. Différents développements ont vu le jour ces dernières années pour modéliser les réseaux écologiques. C'est le cas notamment de Conefor Sensinode (Saura et Torné, 2009) ou Graphab (Foltête *et al.*, 2021). Néanmoins, ces outils sont surtout spécialisés dans le calcul de métriques de connectivité entre habitats et non dans celui d'indicateurs qualifiant le rôle et la structure des haies. C'est ce qui a motivé le développement de l'outil HedgeTools.

Dans cet article, nous présentons ce nouvel outil en cours de développement, les principes qui ont guidé sa conception, les différentes fonctionnalités actuelles et deux exemples qui illustrent leur mise en œuvre.

2. Les variables attendues pour caractériser les haies

Une haie peut être décrite de multiple façons, à partir d'un grand nombre de variables. Celles-ci renseignent à la fois sur la composition et la structure de l'objet. Elles visent aussi à qualifier la qualité de l'habitat et approcher les rôles fonctionnels de la haie et les services écosystémiques qu'elle fournit. La porosité d'une haie peut ainsi aider à spatialiser son rôle brise-vent qui a un impact sur la production agricole. Sa biomasse aérienne peut permettre d'estimer une partie de la quantité de carbone qu'elle stocke. Son orientation par rapport à la pente dominante, la présence de talus ou fossés à proximité peuvent aider à évaluer son rôle hydrologique (favorisant l'infiltration de l'eau dans les sols) et sa capacité à limiter l'érosion et les pollutions diffuses. Les caractéristiques de la végétation informent sur les ressources florales disponibles pour les pollinisateurs et sur la présence potentielle d'auxiliaires des cultures et donc, sur la régulation possible des bioagresseurs. Par ailleurs, selon sa structure, la haie va jouer le rôle de corridor au sein de la matrice agricole, permettant la circulation d'un habitat à l'autre en fonction de la capacité de dispersion des espèces.

Une première étude bibliographique nous a permis de recenser les principales variables attendues par les différents acteurs qui s'intéressent à cet objet d'étude pour approcher ses caractéristiques structurelles et fonctionnelles (Villierme, 2013). Cette liste de variables a ensuite été revue et enrichie dans le cadre d'un partenariat avec l'AFAC-Agroforesteries, l'IGN, l'Office Français de la Biodiversité, le CNES et TerraNIS (projet SCO EagleHedges). La liste prend en compte une partie des indicateurs proposés dans le plan de gestion durable des haies (PGDH) dont l'objectif est d'harmoniser la description des haies à l'échelle nationale (AFAC 2021).

L'établissement de cette liste de variables (Table 1) a été un préalable au développement de HedgeTools qui vise à fournir des méthodes pour automatiser leur calcul. Néanmoins, l'outil n'a pas la prétention d'être capable de tout calculer. Certains indicateurs sont inaccessibles sans relevés terrain, soit par manque de données spatiales adéquates, soit par manque de méthodes opérationnelles. C'est par exemple le cas du mode d'entretien des haies. C'est aussi le cas de la composition en essences qui pourrait potentiellement être obtenue en utilisant de l'imagerie multispectrale ou hyperspectrale mais la reconnaissance automatique des essences, surtout dans le cas d'objets fins et hétérogènes comme les haies, reste très délicate. Le tableau précise ainsi les variables considérées comme accessibles aujourd'hui ou qui le seront à terme après des développements spécifiques. Certaines variables peuvent être obtenues avec des outils standards disponibles sous SIG (ex. la longueur d'un linéaire). Dans ce cas, HedgeTools a surtout vocation à les rassembler et les centraliser au sein d'une même boîte à outils, sans « réinventer la roue ». D'autres variables sont plus complexes et des méthodes spécifiques ont été mises au point (ou le seront à terme).

TABLE 1. Principales variables d'intérêt pour caractériser les haies (d'après Baudry et Jouin, 2003 ; Villierme, 2013 ; AFAC, 2021 adapté) avec leur niveau d'accessibilité par traitement automatique de données spatiales multi-sources.

Niv. d'analyse	Catégorie	Propriété	Accessibilité
	Typologie	Taillis simple, taillis mixte, futaie régulière, futaie irrégulière, haie en devenir, taillis sous futaie	à l'étude
	Morphologie (2D)	Longueur, largeur du houppier, forme (élongation, complexité), orientation	oui
		Densité, proportion de trouées (ou continuité des houppiers, en %)	oui
		Perméabilité (en%)	à l'étude
	Physionomie (3D)	Volume et biomasse	à l'étude
		Hauteur de la canopée (dominante, moy, var, min, max)	oui
		Présence des étages de végétation (herbacé, 0,3-2m, 2-7m, > 7m)	à l'étude
		Recouvrement par étage (0 à 100%)	à l'étude
		Présence de très gros bois ($\varnothing > 70\text{cm}$)	à l'étude
		Espèce dominante, secondaires, diversité et richesse spécifique, persistance du feuillage, espèces invasives	non
		Age moyen des arbres de futaie, des cépées d'arbres et d'arbustes	non

Objet	Composition	Présence de bois mort sur pied, arbres à cavités, lianes, terrier	<i>non</i>
	Type de sol	Superficiel, moyen, profond	<i>non</i>
	Mode d'entretien et facteurs de dégradation	Traces d'épareuse, lamier, état sanitaire dégradé, espèce invasive, abrouissement ou piétinement par le bétail, brûlage des rémanents, tailles inadéquates, fossé drainant profond, clôture.	<i>non</i>
	Fonction / rôle	Brise-vent, hydrologique, anti-érosif, potentiel d'accueil de la biodiversité	<i>à l'étude</i>
Contexte	Topographique	Pente dominante, position sur versant, orientation p/r à la pente	<i>oui</i>
	Géographique	Fossé et talus adjacent	<i>à l'étude</i>
	Embase	A plat, en creux, sur talus	<i>à l'étude</i>
	Interface	Type d'occupation du sol de part et d'autre de la haie	<i>oui</i>
Réseau	Connectivité	Type de nœud (O, L, T, X, M)	<i>oui</i>
		Indices (β , γ , μ , α)	<i>oui</i>
		Nombre de sous-réseaux, nombre d'arêtes, dimension des sous-réseaux	<i>oui</i>
Paysage	Indices synthétiques	Linéaire de haies, densité, indice de cohérence, indice de linéarité	<i>oui</i>

3. Principes de conception de HedgeTools et modélisation associée

L'analyse des différentes variables a rapidement fait émerger la nécessité d'avoir recours à différentes formes de modélisation de la haie. HedgeTools a ainsi été conçu pour générer des représentations multiples de l'objet à partir d'une couche d'entrée initiale (une carte de haies supposée disponible) en ayant recours à différentes géométries, différents niveaux d'analyse, et à une structure de données adaptée.

3.1. Des définitions multiples

De nombreuses définitions de la haie coexistent. L'acceptation la plus commune est celle d'un alignement d'arbres ou d'arbustes marquant la limite entre deux parcelles ou entre deux propriétés (Larousse, 2013). La définition de l'écologue, du géographe ou de l'agronome est généralement plus complexe, incluant des contraintes sur la physionomie, la composition interne, la gestion ou le contexte paysager. La haie peut ainsi être considérée comme une unité de gestion, au même titre qu'une parcelle agricole (Baudry et Jouin, 2003). Pour une représentation

cartographique, il est également nécessaire de s'accorder sur des dimensions minimales ou maximales à respecter, que ce soit sur la longueur ou la largeur de la haie ou encore, la taille des trouées. L'IGN précise ainsi qu'une haie, dans la BDTopo®, est une « *formation linéaire comportant des arbres, arbustes ou arbrisseaux sur au moins 25 m de long, sans interruption de plus de 20 m, sur une largeur inférieure à 20 m, et sa hauteur potentielle est supérieure à 1,30 m* ». Une seconde définition est adoptée dans la BD Haie, lorsque l'origine des données est celle des surfaces non agricoles du Registre Parcellaire Graphique.

Etant donné la pluralité des définitions, nous avons fait le choix de concevoir un outil qui ne soit pas dépendant d'une définition spécifique. Nous avons également intégré des mécanismes permettant de transformer une représentation vers une autre en tenant compte des spécifications géométriques respectives quand cela s'avérait possible.

3.2. Des géométries multiples

La haie est un objet longiligne et le recours à une modélisation géométrique multiple est nécessaire pour calculer la diversité des variables. La représentation *linéaire* est adaptée au calcul de longueur ou d'orientation mais pas à l'estimation de la largeur. La représentation *surfactive* permet de calculer l'emprise et de caractériser la morphologie ou la physionomie. Les nœuds aux extrémités (représentation *ponctuelle*) sont utiles aux mesures de connectivité dans le graphe. Chaque géométrie (emprise, axe, extrémités) participe ainsi à la modélisation de la haie et l'outil est conçu pour pouvoir générer et naviguer d'une géométrie à l'autre.

3.3. Des points de vue multiples

Le calcul des indicateurs structurels et fonctionnels des haies requiert différents découpages qui sont autant de points de vue sur l'objet (Figure 1). Le point de vue *topologique* (la haie est une arête du graphe planaire) est le point de vue fondamental dans l'outil. À partir de celui-ci, la création d'autres découpages est possible pour tenir compte d'une caractéristique interne à l'objet (ex. l'orientation des tronçons qui composent l'arc topologique ; la strate dominante de végétation) ou externe, en relation avec le voisinage de l'objet (ex. redécoupage en tenant compte des limites du parcellaire agricole d'un côté ou de l'autre de la haie) ou en référence au gestionnaire impliqué. La génération de points de vue est prévue dans HedgeTools à l'aide d'outils spécifiques.

3.4. Différentes échelles d'analyse

Selon les caractéristiques à extraire, différentes échelles d'analyse et différentes données en entrée doivent être considérées. Pour qualifier la structure d'une haie par exemple, l'échelle de l'objet lui-même suffit mais il est nécessaire d'opérer sur le réseau de haies (graphe topologique) correspondant à une autre échelle d'analyse,

pour caractériser le degré de connectivité des objets. De même, le voisinage de la haie est à prendre en compte pour qualifier le contexte topographique ou géographique et le paysage dans lequel elle s'inscrit. Dans ce dernier cas, des données supplémentaires peuvent être mobilisées. Il peut s'agir d'un modèle numérique de terrain, du parcellaire agricole, des réseaux routier et hydrographique ou d'un modèle numérique de hauteur. Des développements sont également en cours pour intégrer le calcul de certaines variables à partir d'un nuage de points LiDAR (utile pour les étages de végétation notamment).

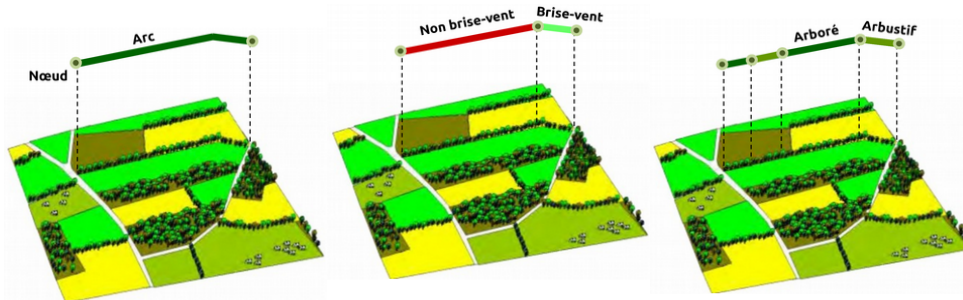


FIGURE 1. Des points de vue multiples sur la haie impliquant différents découpages géométriques.

3.5. Modèle conceptuel de données

Les différents principes évoqués ont permis de proposer une structure de données générique, adaptée aux géométries multiples et à la diversité des points de vue à prendre en compte (Figure 2).

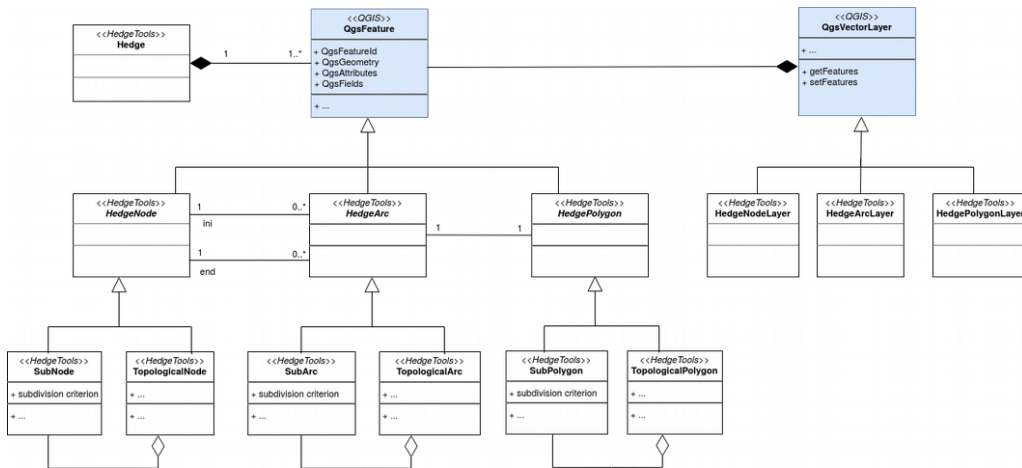


FIGURE 2. Diagramme de classes UML de l'outil HedgeTools. Les attributs et méthodes de classes ne sont pas représentés pour faciliter la lecture.

Les classes de HedgeTools s'appuient sur des classes de l'API QGIS (logiciel cible) et notamment *QgsFeature* et *QgsVectorLayer* qu'elles spécialisent. La haie est ainsi un agrégat des trois géométries via les classes *HedgeNode*, *HedgeArc* et *HedgePolygon* qui sont reliées entre elles selon une structure topologique. Les points de vue sont instanciés à travers les classes *SubArc*, *SubPolygon* et *SubNode* pour lesquelles le critère de découpage est précisé grâce à l'attribut *subdivision criterion*. Ces classes entretiennent une relation de composition avec le point de vue topologique. De cette façon, un tronçon représentant l'axe médian de la haie peut être modélisé par un arc topologique du graphe planaire (objet de la classe *TopologicalArc*) mais aussi être redécoupé en fonction d'un rôle particulier ou d'une physionomie souhaitée via la classe *SubArc* (les objets *subArc* composant alors l'arc topologique). Le même principe s'applique aux emprises de la haie et ses sommets.

4. Implémentation

La version initiale de HedgeTools a été implémentée en langage Python dans le logiciel ArcGIS Desktop (version ArcInfo) en s'appuyant sur la bibliothèque ArcPy (Villierme, 2013 ; Sheeren *et al.*, 2015). Jamais diffusée, cette version était restée à l'état de prototype jusqu'ici. La version actuelle, en cours de développement sous la forme d'un *plug-in*, a été portée sous le logiciel QGIS pour favoriser l'ouverture. Les outils disponibles et opérationnels dans cette version (qui n'est pas encore diffusée) sont organisés en différentes catégories (Figure 3). Celles-ci reflètent la logique d'utilisation de HedgeTools qui passe par l'enchaînement d'une succession d'étapes depuis la mise en structure des données jusqu'au calcul des différentes variables.

- ▼ Hedge tools
 - * Categorize wooded area
 - ▼ 1 - Data preparation
 - * 1 - Create median axis
 - * 2 - Create topological nodes
 - * 3 - Create topological arc
 - * 4 - Create polygons from median axis
 - * 5 - Modify median axis (optional)
 - ▼ Data transformation
 - * Split by distance
 - * Split by interface
 - * Split by network
 - * Split by orientation
- ▼ Geographic context
 - * Relative orientation inside the main slope
 - * Shortest distance to forest
 - * Topographic position
 - ▼ Hedges morphology
 - * Length
 - * Orientation
 - * Shape metrics
 - * Width
 - ▼ Hedges network
 - * Connexion to a forest
 - ▼ Hedges physiognomy
 - * Height metrics from CHM
 - * Strata proportion from CHM

FIGURE 3. Contenu actuel de la boîte à outils HedgeTools en cours de développement sous la forme d'un *plug-in* QGIS.

La première catégorie d'outils concerne la préparation des données. À partir d'une couche de haies surfaciques en représentation vectorielle, les données sont

transformées selon la structure de HedgeTools en créant le point de vue topologique et les géométries multiples. La seconde catégorie permet de générer d'autres points de vue qui sont au nombre de quatre actuellement : un redécoupage selon une distance fixe définie par l'utilisateur, un redécoupage selon les limites des objets adjacents à la haie, un redécoupage selon l'orientation des tronçons, et un redécoupage qui tient compte du réseau de routes avec une agrégation possible d'objets déconnectés en fonction de la distance qui les sépare (i.e. passage à une nouvelle définition de l'objet haie). Pour chaque point de vue, l'ensemble des géométries est mis à jour en redéfinissant les identifiants des objets, tout en conservant la relation avec le point de vue topologique de plus haut niveau (agrégat). Le type de nœud est également évalué (O pour une impasse, L, T ou X pour des relations avec les arcs en 2 à 4-connexité, et M si le degré de connexité dépasse 4).

Les autres catégories d'outils disponibles actuellement permettent de calculer un ensemble d'indicateurs morphologiques à l'échelle de l'objet (longueur, orientation, indices de forme et d'élongation, largeur) ou de préciser la physionomie de chaque haie en s'appuyant sur un modèle numérique de canopée (CHM). Ces variables sont ajoutées après calcul à la table attributaire de la couche de haies utilisée en entrée, correspondant à un point de vue choisi (topologique ou autre). Il est ainsi possible de calculer un même lot de variables pour différents points de vue, en choisissant à chaque fois une couche d'entrée correspondante. Les autres outils permettent de caractériser le contexte de l'objet : position topographique en 10 classes selon les catégories de géomorphons (Jasiewicz et Stepinski, 2013), orientation par rapport à la pente dominante, distance à la forêt la plus proche. Des indices de connectivité dans le graphe sont également en cours d'implémentation.

HedgeTools est avant tout dédié à la caractérisation des haies et non à leur extraction automatique à partir d'images. Toutefois, un outil d'extraction est également proposé. Il s'appuie en entrée sur deux sources disponibles à l'échelle nationale : la BDOrtho IRC de l'IGN et un Modèle Numérique de Surface (photogrammétrique ou LiDAR). L'extraction, fondée sur un seuillage empirique, permet de dériver une couverture détaillée des éléments arborés (forêts et arbres hors forêts) sans faire appel à des techniques avancées de traitement d'images ou d'apprentissage machine. La couverture arborée est ensuite retraitée par analyse morphologique de façon à identifier et séparer les objets forêts et les objets haies constituant des agrégats en sortie d'extraction (outil *Categorize wooded area* proposé en amont de la préparation des données, Figure 3).

5. Premiers cas d'utilisation

Afin d'illustrer l'usage de l'application et les résultats de quelques méthodes développées, nous détaillons ci-dessous deux cas d'utilisation. Le premier concerne la préparation des données selon la structure de HedgeTools, depuis une couche d'éléments arborés (tous types confondus) jusqu'à la création du point de vue

topologique. Le second cas d'usage illustre le calcul de variables morphologiques et physionomiques.

5.1. Cas 1 : préparation des données avec mise en géométrie et topologie

La couche d'entrée fournie par l'utilisateur est ici supposée correspondre à la couverture arborée de la zone d'intérêt, sans distinction des forêts et des arbres hors forêts (Figure 4a). Après application de l'outil *Categorize wooded area* qui s'inspire de l'approche proposée par (Touya *et al.*, 2010), les haies sont séparées des autres éléments (Figure 4b). La couche de haies est ensuite utilisée pour en extraire l'axe médian (outil *Create median axis*). Cette étape s'est avérée plus complexe que prévue dans son développement. La difficulté réside dans le post-traitement, selon l'approche adoptée, pour obtenir un nettoyage adapté des *arcs pendants* après squelettisation mais aussi, pour conserver la connectivité avec les extrémités de l'objet initial (emprise) selon l'orientation principale, et pour simplifier la géométrie linéaire qui en découle. Deux méthodes au choix sont proposées dans l'outil : l'une en vecteur et l'autre en raster. En vecteur, un lissage des emprises par dilatation-érosion est d'abord opéré pour simplifier les contours. Une squelettisation est ensuite appliquée via la méthode de Voronoï, en s'appuyant sur les points du contour (fonction *v.voronoi* du module GRASS de QGIS d'après Fortune, 1987). Cette étape est problématique actuellement car la fonction utilisée présente des instabilités numériques qui génèrent des artefacts. Une réécriture de l'algorithme est en cours. En raster, l'axe médian est obtenu par un amincissement itératif des surfaces initiales (fonction *r.thin* du module GRASS de QGIS d'après Jang et Chin, 1990). La méthode produit un raster avec les axes médian d'une épaisseur correspondant à un pixel, accompagnés d'arcs pendants. Cette sortie est ensuite vectorisée. Un post-traitement est alors opéré pour corriger les artefacts et supprimer les arcs pendants.

À ce stade, un squelette de la couche arborée initiale est disponible. La seconde étape consiste à extraire les nœuds aux extrémités des haies après avoir fusionné au préalable les tronçons d'axes médian qui partagent une relation d'adjacence (outil *Create topological nodes*). On dispose ainsi des axes et de leurs extrémités sans relation explicite entre eux (Figure 4c). Ensuite, la structure topologique de graphe planaire est produite de façon à associer les nœuds à leurs arcs respectifs (outil *Create topological arcs*). Cela permet d'obtenir les relations de connectivité et de qualifier chaque nœud selon la nature de la jonction (i.e. nœuds O,L,T,X,M selon le nombre d'arcs connectés).

Il reste à assurer une cohérence entre les représentations linéaire et surfacique. Pour la représentation linéaire, chaque haie est bien individualisée selon le point de vue topologique. Ce n'est pas le cas des emprises fournies par la couche de départ (représentation surfacique) dans laquelle les haies forment des agrégats. Ces emprises sont ainsi redécoupées, par calcul des bissectrices entre axes médian, pour en générer de nouvelles (outil *Create polygons from median axis*). Chaque objet surfacique est alors associé à son homologue linéaire avec transfert des identifiants (Figure 4d).

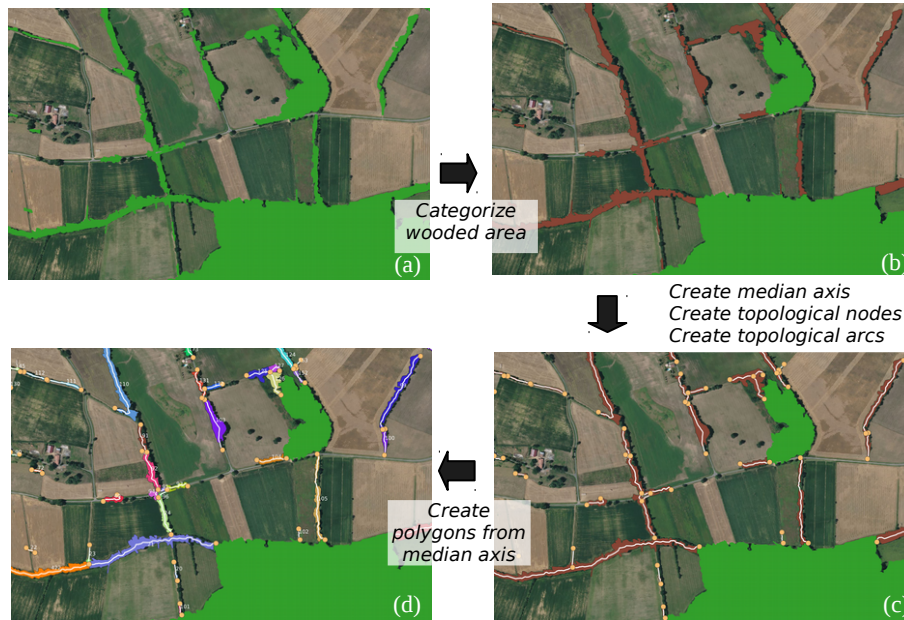


FIGURE 4. Outils de préparation des données de HedgeTools permettant, à partir d'une couche d'éléments arborés (a), d'en extraire les haies (b), d'en dériver les axes médian et le graphe topologique (c) et de redécouper les emprises (d).

5.2. Cas 2 : calcul des caractéristiques morphologiques et physiologiques

Ce n'est qu'à l'issue des différentes étapes de préparation des données que les indicateurs morphologiques et physiologiques peuvent être calculés et enregistrés dans la table attributaire des objets.

Pour la morphologie, l'outil permet actuellement de calculer la longueur, l'orientation, l'élongation, la compacité, la convexité et la largeur des objets haie. La largeur est estimée de deux façons différentes : (1) en créant d'abord des transects perpendiculaires à l'axe median espacés d'une certaine distance puis, en calculant la moyenne ou médiane de leur longueur ou (2), de façon plus approximative, en calculant le rapport entre l'aire du rectangle englobant et la longueur de l'arc median. Pour la physiologie, des statistiques descriptives sur la hauteur des objets sont calculées, en utilisant un CHM raster. La proportion des étages de végétation est obtenue par discrétisation du CHM (Figure 5). Il s'agit ici uniquement des étages vus du dessus mais à terme, les outils seront enrichis d'indicateurs issus du nuage de points 3D LiDAR.

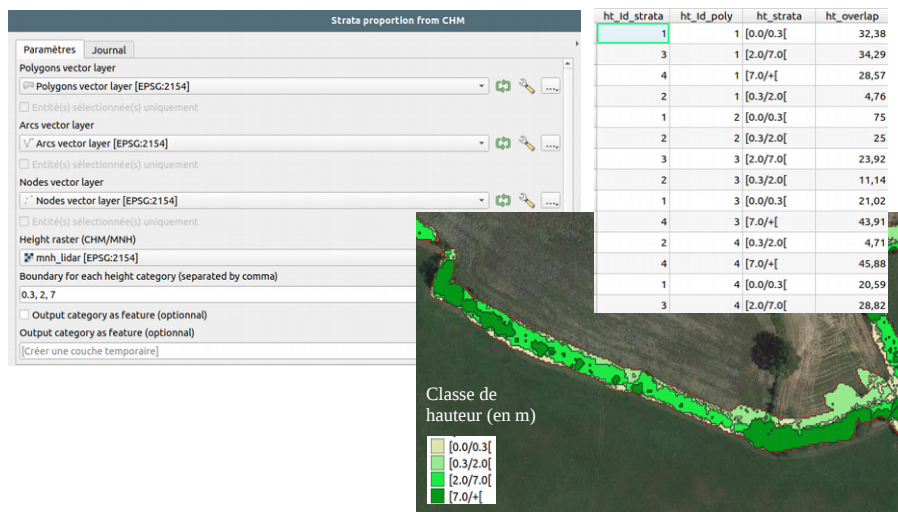


FIGURE 5. Proportion des strates de végétation pour une haie obtenue par discrétisation d'un modèle numérique de canopée (CHM).

6. Conclusion et perspectives

La boîte à outils HedgeTools vise à combler un manque d'outils opérationnels pour caractériser les haies champêtres et évaluer leur multifonctionnalité dans les paysages agricoles. Elle repose sur une structure de données adaptée à la représentation multiple (géométrie et point de vue) nécessaire au calcul de caractéristiques à différentes échelles d'analyse. Ces caractéristiques alimentent la démarche de diagnostic systémique du bocage en dépassant le seul inventaire cartographique. C'est un préalable à la définition d'un plan d'actions pour la gestion des haies.

L'outil sera documenté avant sa diffusion prévue en fin d'année 2023. Des tests de performance sur des jeux de données de taille importante (à l'échelle d'un département par exemple) seront à entreprendre pour s'assurer de la stabilité des calculs et de l'opérationnalité des outils.

Les perspectives d'enrichissement de HedgeTools sont nombreuses. Les développements actuels portent sur l'amélioration de l'extraction de l'axe médian, le calcul d'indices de connectivité dans le graphe, l'évaluation des rôles hydrologiques et brise-vent des haies, et le calcul de variables qualifiant la structure verticale des objets. Le rôle brise-vent implique une évaluation de la porosité que nous devrions pouvoir approcher avec des données LiDAR. L'identification automatique du type de haie selon leur physionomie est également recherchée. Par la suite, l'accent sera mis sur l'évaluation d'indicateurs favorables à la biodiversité.

Bibliographie

- AFAC-Agroforesteries (2021). Guide de réalisation du Plan de Gestion Durable des Haies, version sept. 2021, 66 p.
- Axe M.S., Grange I.D. et Conway J.S. (2017). Carbon storage in hedge biomass—A case study of actively managed hedges in England, *Agriculture, Ecosystems and Environment* vol. 250, p. 81–88.
- Baudry J. et Jouin A., (2003). *De la haie aux bocages : organisation, dynamique, gestion*. INRA éditions.
- Estrada-Carmona N., Sanchez A., Remans R., Jones S. (2022). Complex agricultural landscapes host more biodiversity than simple ones : a global meta-analysis, *Proceedings of the National Academy of Sciences*, 119 (38) e2203385119.
- Fauvel M., Planque C., Sheeren D., Dalla Mura, M (2014). Télédétection des éléments semi-naturels : utilisation des données Pléiades pour la détection des haies. *Revue Française de Photogrammétrie et Télédétection*, vol. 208, p. 111-116.
- Foltête J.-C., Vuidel G., Savary P., Clauzel C., Sahraoui Y., Girardet X. et Bourgeois M. (2021). Graphab: an application for modeling and managing ecological habitat networks. *Software Impacts*, vol. 8: 100065.
- Fortune S.J. (1987). A Sweepline Algorithm for Voronoi Diagrams. *Algorithmica*, vol. 2, p. 153-174.
- Jang B.-K. et Chin R.T (1990). Analysis of thinning algorithms using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, n°6, p. 541-551.
- Jasiewicz J. et Stepinski T. (2013). Geomorphons - a pattern recognition approach to classification and mapping of landforms, *Geomorphology*, vol. 182, p. 147-156.
- Lucas C., Bouten W., Koma Z., Kissling W.D., Seijmonsbergen A.C. (2019). Identification of Linear Vegetation Elements in a Rural Landscape Using LiDAR Point Clouds. *Remote Sens.*, vol. 11, n°292.
- Merciol F., Fauqueur L., Damodaran B.B., Rémy P.-Y., Desclée B., Dazin F., Lefèvre S., Masse A., Sannier C. (2019). GEOBIA at the Terapixel Scale: Toward Efficient Mapping of Small Woody Features from Heterogeneous VHR Scenes. *ISPRS Int. J. Geo-Inf.*, 8, 46.
- Morin S., Commagnac L., Benest F. (2019). Caractériser et suivre qualitativement et quantitativement les haies et le bocage en France, *Sciences Eaux & Territoires*, vol. 4, n°30, p. 16-21.
- Saura S. et Torné J. (2009). Conefor sensinode 2.2 : A software package for quantifying the importance of habitat patches for landscape connectivity. *Environmental Modelling & Software*, vol. 24, n°1, p. 135–139.
- Sheeren D., Masse A., Ducrot D., Fauvel M., Collart F., May S.(2012). La télédétection pour la cartographie de la trame verte en milieu agricole. Evaluation des potentialités d'images multi-angulaires à très haute résolution spatiale. *Revue Internationale de Géomatique*, vol. 22, n°4, p. 539-563.
- Sheeren D., Roué A., Villierme L., et Monteil C. (2015). HedgeTools pour ArcGIS : un outil dédié à la modélisation et la caractérisation des haies dans le paysage. *Conférence francophone ESRI (SIG'2015)*, Versailles, France.
- Touya G., Duchêne C., et Mustière S. (2010). Généralisation et intégration pour un fond vert commun entre l'IFN et l'IGN. *Revue Internationale de Géomatique*, vol. 20, n°1, p. 65-86.
- Villierme L. (2013). *Caractérisation automatique du réseau linéaire de haies dans un paysage agri-forestier*. Stage de fin d'études du Master 2 SIGMA, Toulouse INP / Université Toulouse-Jean Jaurès.

Weninger T., Scheper S., Lackóová L., Kitzler B., Gartner K., King N., Cornelis W., Strauss P., et Michel K. (2021). Ecosystem services of tree windbreaks in rural landscapes—a systematic review. *Environmental Research Letters*, 16, 103002.

Caractérisation spatiale multiscalaire des agricultures contribuant à l’approvisionnement alimentaire de proximité.

Propositions méthodologiques pour
l’accompagnement de l’action publique de
planification alimentaire

Esther Sanz Sanz¹, Claude Napoléone¹, Michel Mouléry¹

*1. Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), UR Ecodéveloppement, Avignon (France)
esther.sanz-sanz@inrae.fr*

RESUME. Cette communication présente nos travaux des dix dernières années en modélisation de données spatiales et temporelles, dont le but est l'analyse et la représentation de la territorialisation des systèmes alimentaires. Concrètement, nous avons caractérisé la forme et la localisation des agricultures contribuant à l'approvisionnement alimentaire de proximité, à deux échelles de l'action publique : la ville et le bassin d'approvisionnement (ou foodshed). En mobilisant les cadres théoriques des relations agriculture/ville/alimentation, ces travaux ont été l'occasion de concevoir et développer des méthodes d'analyses originales (quantitatives, spatialisées et multiscalaires) de la géographie des espaces agricoles fonctionnellement liés à la ville. Notre cheminement s'est déroulé en trois étapes : 1) confirmation de l'existence de signatures spatiales spécifiques aux agricultures liées à la ville ; 2) développement et stabilisation de la méthodologie d'analyse, au travers de démarches de recherche-action ; et 3) mesure et spatialisation du bassin d'approvisionnement alimentaire. Au-delà des innovations méthodologiques, ces démarches ont pour vocation de venir en appui aux actions publiques concernant d'une part, les usages des sols et l'aménagement du territoire, et d'autre part la planification alimentaire ; notamment les initiatives permettant d'augmenter la part des produits de proximité dans l'approvisionnement des cantines scolaires.

ABSTRACT. This paper presents our work over the last ten years in spatial and temporal data modelling to analyse and represent the territorialisation of food systems. Specifically, we have characterised the form and location of agriculture contributing to local food supply, at

two scales of public action: the city and the foodshed. By mobilising the theoretical frameworks of the agriculture/city/food relations, this work was an opportunity to design and develop quantitative, spatialised and multiscale analysis methods of the geography of agricultural spaces functionally linked to the city, in a three-stage process: 1) confirmation of the existence of specific spatial signatures of agricultures linked to the city; 2) development and stabilisation of the analysis methodology, through action-research approaches; and 3) measurement and spatialisation in the concept of the foodshed. In addition to methodological innovations, these approaches are intended to support public actions concerning land use and spatial planning on the one hand, and food planning on the other; in particular, initiatives to increase the share of local products in school canteen supplies.

MOTS-CLES : géographie de l'alimentation, systèmes alimentaires territorialisés, proximité, méthodologie, modélisation, action publique, planification

KEYWORDS: geography of food, territorialized food systems, proximity, methodology, modeling, public action, planning

1. Introduction

L’approvisionnement alimentaire de proximité revêt la caractéristique d’un système complexe confronté à des défis interdépendants pour fournir une alimentation accessible, saine, respectueuse de l’environnement et durable (HLPE, 2017). Notamment, après la crise sanitaire COVID et l’actuelle crise ukrainienne, il y a un enjeu social à expliciter le lien entre les besoins alimentaires et la disponibilité de l’offre de proximité (IPES-Food, 2020). Or, cet enjeu fait écho à une controverse scientifique sur la capacité de l’approvisionnement de proximité à répondre de manière durable à la demande de urbaine de proximité (Kinnunen et al., 2020). En suivant Moustier et al. (2023), nous définissons la proximité dans un mix entre distance et organisation de la chaîne de distribution. Il s’agit d’une offre alimentaire issue de l’agriculture périurbaine et rurale locale, avec un ou deux intermédiaires entre les agriculteurs et les consommateurs.

Dans ce cadre, l’agriculture de proximité tend à se doter d’une valeur stratégique et mobilise des travaux originaux dans l’urbanisme et la recherche. La question est d’identifier quels systèmes agricoles sont les plus aptes à être intégrés dans la planification urbaine et comment les caractériser empiriquement. Or, si les outils d’urbanisme basés sur le zonage d’occupation du sol (PLU et SCoT) peuvent être utilisés pour protéger une superficie agricole, ils tiennent difficilement compte des spécificités de l’agriculture en tant qu’activité économique évolutive. Par ailleurs, au niveau des cadres théoriques et méthodologiques disponibles dans le corpus de connaissances académiques, ils permettent mal de considérer les espaces agricoles comme composante du système urbain (Vidal and Fleury, 2009). En effet, il s’agit de délimiter et caractériser les espaces des agricultures contribuant à l’approvisionnement de proximité. Or, les modèles centre-périphérie canoniques à la Von Thünen (1826) ou issus de la nouvelle économie urbaine (Alonso, 1964; Mills, 1972; Muth, 1969) permettent mal de représenter la complexité spatiale des formes d’agricultures de proximité. Qui de plus est, cette caractérisation est confrontée à un enjeu social d’opérationnalité des formes de représentation pour la pratique de l’urbanisme et de la planification alimentaire (*food planning*).

Pour aborder ce double enjeu académique et social, nous avons développé une démarche « de la fourche à la fourchette » combinant une approche empirique participative à une approche cartographique et modélisatrice. La question de recherche est : quels sont les déterminants socio-spatiaux des géographies du nexus agriculture/alimentation/villes ? Nos travaux s’inscrivent dans trois champs de la géographie quantitative et humaine : 1/ la nouvelle géographie de la sécurité alimentaire focalisée sur la relation entre les dynamiques socio-spatiales des systèmes de production agricole et la durabilité des systèmes alimentaires urbains (Schreiber et al., 2021; Sonnino, 2016); 2/ la planification alimentaire (*food planning*) visant à apporter des connaissances utiles à l’urbanisme opérationnel pour des systèmes alimentaires sains et durables (Brinkley, 2013; Viljoen and Wiskerke, 2012) et 3/ les sciences du changement des usages du sol (*Land Use Change Sciences*) pour caractériser les dynamiques de changement socio-spatiaux et leurs

déterminants, dans une visée modélisatrice spatiale et multiscalaire, notamment en relation avec la sécurité alimentaire (Erb et al., 2016; Verburg et al., 2013). Formellement, nos travaux cherchent à analyser la géographie des espaces agricoles fonctionnellement liés à la ville, dans un cheminement en trois étapes : 1) valider empiriquement l'hypothèse de signatures spatiales des agricultures liées à la ville (Poggi et al., 2021) ; 2) en stabiliser une méthodologie d'analyse (déclaration d'invention) au travers de démarches de recherche-action (observation participante, entretiens, focus groups, ateliers prospectifs) permettant de faciliter des processus collectifs de décision au sein du *living lab* d'Avignon ; et 3) cristalliser les résultats dans le concept de bassin d'approvisionnement alimentaire (*foodshed*).

Nous présenterons, dans cette communication, les cadres d'analyse quantitatifs, spatialisés et multiscalaires permettant de caractériser la forme et localisation des agricultures contribuant à l'approvisionnement de proximité. Dans une visée modélisatrice et d'accompagnement de l'action publique de planification alimentaire, nous le ferons à deux échelles déterminant la dynamique des systèmes alimentaires territorialisées : la ville, en nous focalisant sur l'espace périurbain et le bassin d'approvisionnement alimentaire des villes (ou *foodshed*). Nous illustrerons nos propositions méthodologiques par une analyse empirique des relations dynamiques entre aires de production agricoles et aires de consommation, dans la région d'Avignon/Vaucluse (France).

2. Caractérisation spatiale multi-échelle des agricultures contribuant à l'approvisionnement de proximité

2.1. L'échelle de la ville

2.1.1. Contexte et objectifs

En reprenant Sanz Sanz (2016), il est possible de développer une méthodologie pour caractériser spatialement les agricultures périurbaines en lien fonctionnel avec la ville, c'est-à-dire susceptibles d'approvisionner les systèmes alimentaires territorialisés (i.e. forme émergente alternative au systèmes agro-alimentaire mondialisé, qui a pour but d'inventer/valoriser des modèles de production et de consommation plus respectueux de l'environnement qui valorisent les produits locaux dans des filières de proximité ; Rastoin, 2015). Il s'agit de conceptualiser, dans une approche systémique, les déterminants de la forme et de la nature de l'agriculture périurbaine au sein d'un système agri-urbain (FIGURE 1) dont nous proposons un cadre conceptuel sous l'angle du paradigme post-productiviste de la multifonctionnalité agricole (Véron, 2003). L'hypothèse étant que les forces motrices de l'agriculture périurbaine se traduisent dans le paysage selon une triple dimension (Piorr, 2003) : a) la structure du paysage (i.e. morphologie des zones agricoles et urbaines) ; b) le fonctionnement du paysage (i.e. dimension socio-économique des activités agricoles) ; et c) la gestion et la régulation du paysage (i.e. politiques influant sur l'usage du sol). Il est alors possible de reprendre la démarche du géo-agronome Deffontaines (Deffontaines, 1999 ; Deffontaines et al., 1995 ;

Deffontaines and Thinon, 2006), qui consiste à rechercher dans le territoire des espaces de relative homogénéité d’utilisation agricole, en se basant sur l’hypothèse qu’à chaque type d’agriculture (pratiques et stratégies) correspond une combinaison des formes visibles dont l’arrangement est identifiable dans l’espace.

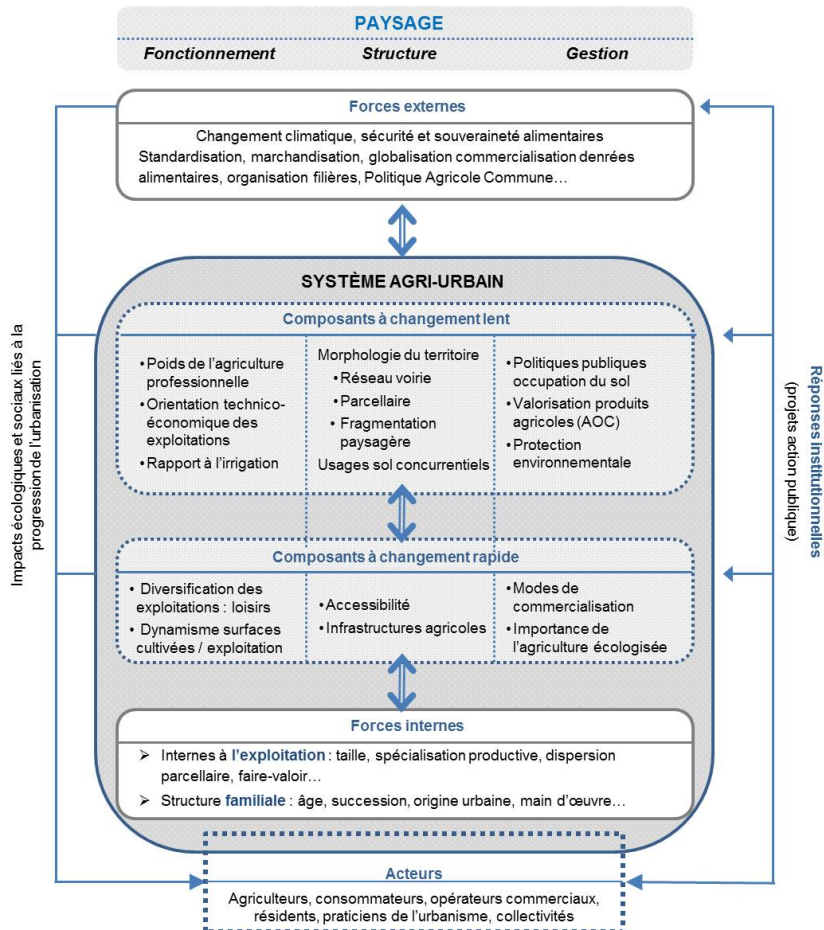


FIGURE 1. Cadre théorique des relations agriculture/ville/alimentation et modèle conceptuel des forces internes et externes façonnant les composants du système agri-urbain. Source : Sanz Sanz, 2016

2.1.2. Méthodologie

A partir de ces deux approches, nous proposons une méthodologie quantifiée permettant de délimiter et caractériser des zones de « relative égale apparence » : les Unités Spatiales d’Agriculture Périurbaine (USAPU). Les USAPU sont des entités spatiales donnant à lire l’agriculture périurbaine, de tailles supérieures à la parcelle, homogènes en termes d’usages et d’aménagements agricoles et urbains et définies

statistiquement (Sanz Sanz et al., 2017a, 2017b, 2016). La finalité est de confirmer empiriquement l'hypothèse d'une **signature spatiale** des agricultures (Poggi et al., 2021), c'est-à-dire des arrangements visibles et identifiables dans l'espace qu'il est possible d'associer aux systèmes productifs. Par exemple, schématiquement, l'horticulture sous abri se développe plutôt sur un parcellaire morcelé aux interstices du tissu périurbain, tandis que les agricultures de filières longues se développent sur un parcellaire régulier et grand, au sein d'un paysage homogène.

La méthodologie de construction des USAPU se développe en quatre phases, à différentes échelles d'analyse emboîtées :

1) Tout d'abord, délimiter et caractériser les USAPU à partir d'une analyse spatialisée, compréhensive et exhaustive d'un site d'étude (dans notre cas l'agglomération d'Avignon) à l'échelle de la parcelle, de l'exploitation et du territoire, et aux moyens d'enquêtes, de la lecture du paysage in situ, de la photo-interprétation et de l'analyse diachronique de l'usage du sol par télédétection.

2) La deuxième étape consiste à traduire les critères définissant les USAPU en variables (n=75) pouvant faire l'objet d'une analyse statistique permettant de caractériser les USAPU selon leur sensibilité à l'influence urbaine (entre autres, analyses factorielles multidimensionnelles de type Analyses en Composantes Multiples, ACM). La robustesse de la méthode est testée sur un cas d'étude contrasté (dans notre cas, Madrid) sans recourir à des enquêtes de terrain approfondies.

3) Troisièmement, pour monter en généralité, nous avons développé un modèle probabiliste multivarié (de type Dirichlet) qui réinterprète les résultats des USAPU dans chaque cas d'étude (7 communes) afin d'estimer leur distribution à une échelle supérieure (le département, 144 communes), sur la base d'un faible nombre de variables (une dizaine) déterminant la majeure part de la variance et simples d'accès (présentes dans les recensements nationaux). Le modèle s'est montré pertinent dans les deux cas d'étude différenciés (le Vaucluse et 4 régions agricoles autour de Madrid).

4) Enfin, nous faisons une classification des communes selon la distribution des USAPU en renvoyant à des éléments actionnables dans l'urbanisme opérationnel et qui ouvrent des perspectives d'action publique complémentaire aux plans locaux d'urbanisme actuels (par exemple, sur la structuration des filières).

2.1.3. Résultats

Nous avons caractérisé et cartographié la signature spatiale des agricultures contribuant à l'approvisionnement alimentaire en identifiant les systèmes à vocation alimentaire et en discriminant, en leur sein, ceux orientés vers les marchés de proximité et ceux dédiés à l'exportation (Sanz Sanz et al., 2018). Nos résultats montrent que les agricultures contribuant à l'approvisionnement alimentaire ne se distribuent pas d'une manière régulière dans un gradient de distance à la ville-centre (FIGURE 2). Par exemple, la viticulture (couleur bordeaux) entretient peu de rapports avec la ville tout en étant l'activité la plus proche. Les agricultures contribuant à l'approvisionnement alimentaire se distribuent plutôt selon une combinaison de

caractéristiques qui résultent des conditions agroécologiques et de l’histoire particulière de chaque lieu en termes d’urbanisation, de développement du secteur agricole, d’organisation des activités (dont les filières agricoles) et de protection environnementale. Dix variables en expliquent significativement la variance ; elles ont été groupées en fonction de l’attribut du paysage dont elles relèvent : sa structure (pourcentage des surfaces artificialisées, rapport aux zones urbanisées, forme du parcellaire), son fonctionnement (agrandissement ou réduction des surfaces cultivées, degré de professionnalisation de l’exploitation, diversification des orientations technico-économiques, mode de faire-valoir des surfaces cultivées) et sa gestion (classification de l’usage du sol dans les Plans Locaux d’Urbanisme PLU, inscription dans un zonage d’Appellation d’Origine Contrôlée AOC).

Au-delà de la mesure de l’agencement des agricultures, la signature spatiale est une voie de caractérisation du système agri-urbain, que nous mettons en œuvre dans une approche modélisatrice (modèle multivarié de type Dirichlet). Elle permet d’extrapoler les dynamiques spatiales observées localement, à une zone géographique plus large et plus complexe (le département, par exemple). Elles permettent également une représentation cartographique localisée (échelle communale) des tendances s’exprimant à une échelle supérieure (fonctionnement de marchés, action publique...). D’autre part, du point de vue opérationnel, cette méthodologie est systémique, générique et présente l’avantage d’être facilement applicable dans d’autres cas d’étude, sans avoir besoin de réaliser des lourdes enquêtes de terrain¹. Le résultat est un outil original de caractérisation et de cartographie des signatures spatiales des agricultures connectées à la ville, actuellement disponible à l’échelle municipale (déclaration d’invention DIRV 17 0122).

¹ La méthodologie a été appliquée -avec adaptations- dans deux projets. Dans le projet Aliville (2016-2019, Fondation de France), nous avons estimé la distribution des USAPU dans 170 communes autour d’Avignon, sur la base d’une dizaine de variables présentes dans le Recensement Général Agricole 2010. Ces résultats, couplés à des données existantes sur les circuits courts de commercialisation, nous ont permis d’estimer le potentiel d’approvisionnement de proximité des 11 légumes frais (bio et conventionnels) permettant de composer les repas servis dans les cantines scolaires d’Avignon (Sanz Sanz, 2021). Nous avons ainsi testé l’hypothèse d’un profil type d’exploitation agricole territoriale (agriculture of the middle - Lev and Stevenson, 2011) susceptible de répondre à la demande des collectivités en produits frais, dans le cadre d’un Système Agro-Alimentaire Localisé (CIRAD-SAR, 1996). Par ailleurs, dans le projet Agriville-USAP (2017-19), nous avons développé un prototype générique à des fins d’urbanisme opérationnel (Boussougou-Boussougou et *al.*, 2021) qui a permis de caractériser l’agriculture périurbaine de deux régions françaises (Provence-Alpes-Côte d’Azur et Poitou-Charentes) en développant une approche statistique à l’échelle municipale (moindres carrés partiels PLS-PM à partir de données disponibles, concernant la géographie (BdTopo, BdAlti), la démographie (INSEE 2010), l’occupation du sol (OSO 2016), les activités agricoles (RGA 2010), l’environnement et les zones protégées (Carmen) et les prix du foncier (SAFER 2016).

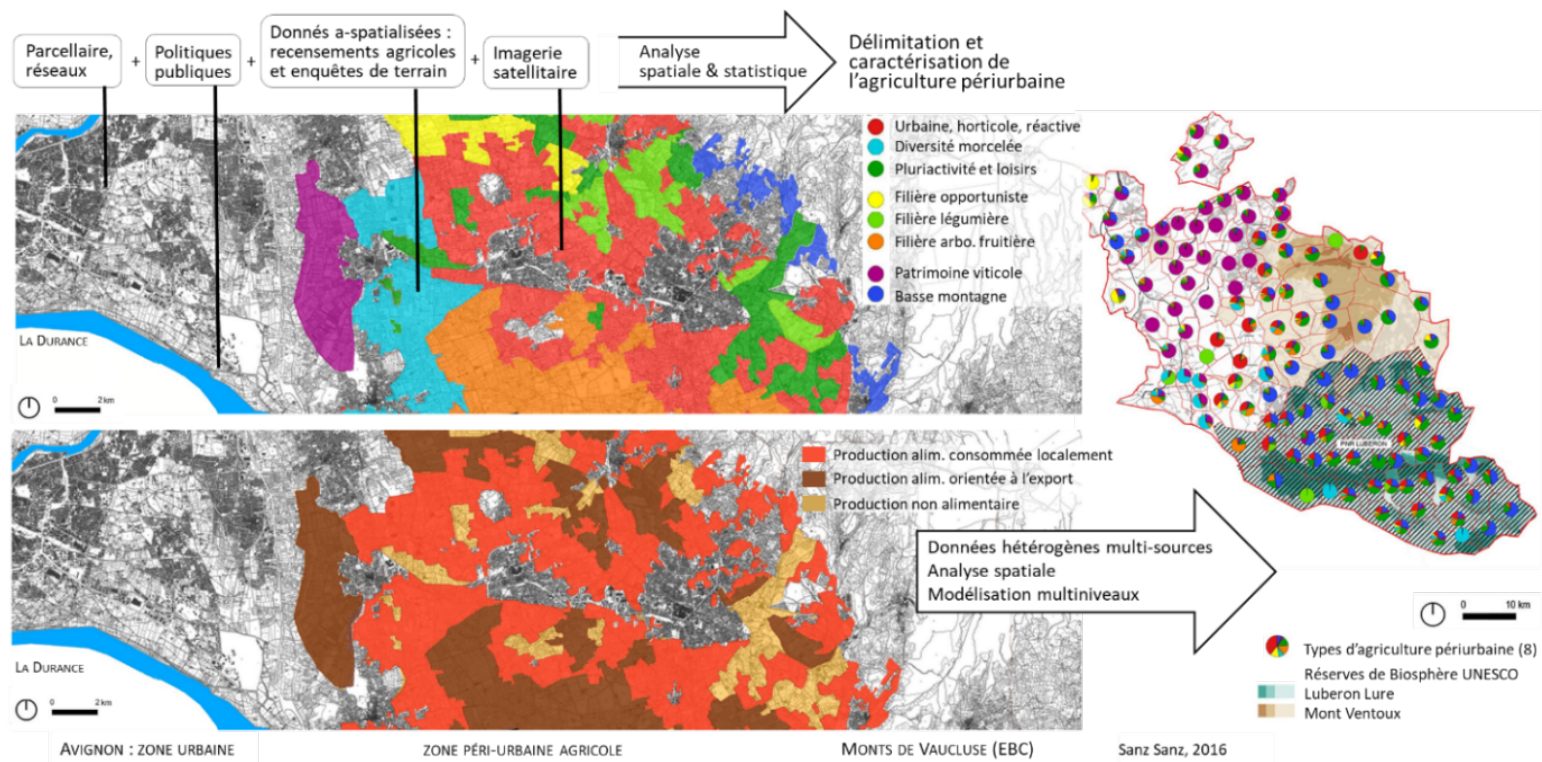


FIGURE 2: Délimitation et caractérisation de l'agriculture périurbaine (USAPU, à gauche) dans la zone d'étude d'Avignon et extrapolation à l'échelle du département par une approche modélisatrice (modèle multivarié de type Dirichlet, à droite).

SAGEO'2023 – Québec (Canada), 5 – 9 Juin 2023

2.2. L'échelle du bassin d'approvisionnement alimentaire ou foodshed

2.2.1. Contexte et objectifs

Afin d'approfondir nos analyses sur la forme et la localisation des relations entre aires de production agricoles et aires de consommation alimentaires, à une échelle pertinente pour l'action publique territoriale, nous avons utilisé nos résultats concernant les agricultures contribuant à l'approvisionnement alimentaire pour estimer et localiser les liens entre diversification agricole et disponibilité alimentaire². Des travaux récents dans le champ de la nouvelle géographie de la sécurité alimentaire mobilisant la notion de bassin alimentaire ou *foodshed*, permettent d'estimer la capacité productive et/ou les flux d'aliments approvisionnant les villes-régions et donc d'estimer leur suffisance alimentaire potentielle (FAO/RUAF/WLU, 2018; Schreiber et al., 2021). Toutefois, les travaux académiques disponibles restent encore mécanistes (reposant sur des distances et des productions allouées normativement à des usages du sol mobilisant des approches géomatiques) et tiennent insuffisamment compte de facteurs pédo-climatiques et socio-économiques localisés, pourtant essentiels au développement des chaînes d'approvisionnement alimentaire de proximité (Willett et al., 2019). Un enjeu méthodologique dans le champ des Land Change Sciences est donc de redessiner les bassins alimentaires à partir de la caractérisation de l'expression locale des liens entre la signature spatiale des systèmes productifs, la diversification agricole et la disponibilité alimentaire de proximité (Sanz Sanz et al., 2021).

2.2.2. Méthodologie

Dans un premier temps (Mouléry et al., 2021, p. 20), nous avons combiné le modèle statistique et spatialement explicite MFSS (*Metropolitan Foodshed and Self-sufficiency Scenario*) conçu par le ZALF (Zasada et al., 2019), avec les rendements agricoles régionaux (Agreste, FranceAgriMer) et avec des données spatialement explicites sur le type de culture (Recensement Parcellaire Graphique), la profondeur du sol (European Soil Database) et la topographie (Digital Elevation Model over Europe). Nous avons ensuite couplé l'approche géomatique avec les données sur les consommations alimentaires au niveau européen (FAO), afin d'estimer la suffisance alimentaire de la ville-région d'Avignon. Nous avons considéré initialement un bassin alimentaire d'un rayon de 30 km autour de la ville-centre (171 communes), que nous avons ensuite étendu à 100 km (1 358 communes). Dans un deuxième temps (Mouléry et al., 2022), nous nous sommes focalisés sur un produit d'origine animale peu présent à proximité de la ville étudiée pour des raisons pédo-climatiques (la viande bovine). En mobilisant les théories de l'écologie du paysage (connectivité, modèle continent-île - Décamps and Décamps, 2007; Taylor et al., 1993), l'approche par la proximité de l'économie géographique (Torre and Wallet,

² Ces travaux ont été développés en collaboration avec le Leibniz Centre for Agricultural Landscape Research (ZALF) dans le cadre programme Horizon 2020/FoodShift20230 de l'Union Européenne (n° d'agrément 862716).

2014) et l'hypothèse de signature spatiale, nous avons caractérisé une forme archétypique des surfaces de production attachées aux exploitations bovines orientées vers les filières de proximité. Pour ce faire, nous avons couplé au moyens d'un SIG la géographie des parcelles pâturées (estives, prairies permanentes et prairies temporaires du RPG) avec des données sur les circuits courts de commercialisation du cheptel bovin (RGA). Nous avons ensuite appliqué la méthode de dilatation/érosion pour regrouper les parcelles en « patchs » (en nous référant à l'écologie du paysage) et nous nous sommes appuyés sur des outils d'analyse spatiale simples afin d'évaluer 1) les effets de la distance à l'abattoir le plus proche ; 2) la rugosité, définie comme la complexité des contours des patchs ; 3) la dominance, en fonction de la densité, du nombre de patchs orientés vers les circuits courts et de la surface totale, et 4) les effets de la localisation dans le périmètre d'un parc naturel régional ou national.

2.2.3. Résultats

Théoriquement (c'est-à-dire dans une approche normative), les surfaces pâturées dans un rayon de 30 km autour d'Avignon permettent une autosuffisance potentielle en viande bovine >80%. Toutefois, dès que l'on considère les conditions pédoclimatiques de la région, l'autosuffisance dans ce rayon n'est plus envisageable. Pour la viande bovine, il faut élargir le bassin alimentaire à un rayon de 100 Km afin d'atteindre une autosuffisance >70%. Nos résultats montrent que les facteurs spatiaux qui caractérisent les chaînes d'approvisionnement de proximité en viande bovine sont le nombre de patchs au sein du *foodshed*, la connectivité entre eux et la rugosité des limites (Mouléry et al., 2022). En substance, sur la viande bovine comme sur les productions végétales, l'hétérogénéité spatiale est un indicateur du lien entre proximité et systèmes productifs – y compris si cette proximité est déportée à une distance importante de la ville du fait des contraintes pédoclimatiques. En outre, le fait d'être dans le périmètre d'un parc naturel accentue l'orientation de la production de viande vers les filières de proximité. Au-delà de l'apport opérationnel, notre principale contribution scientifique réside dans une approche permettant de faire évoluer les analyses de *foodshed*, jusqu'à présent plutôt limitées à un cercle de production isotrope autour de la ville, à une caractérisation en archipel à partir de leurs caractéristiques biophysiques et socio-économiques (FIGURE 3).

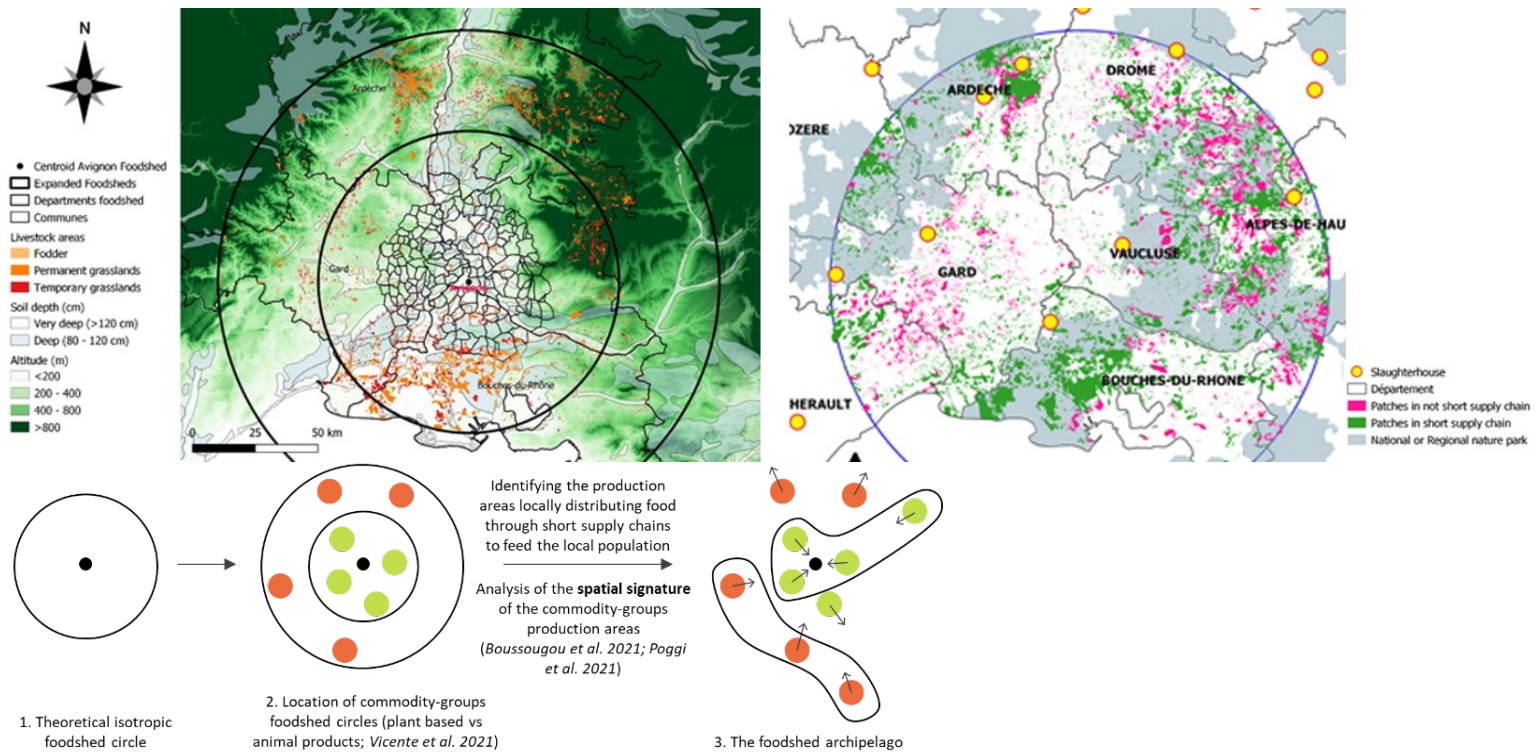


FIGURE 3: en haut à gauche, localisation des parcelles de pâturage (3 catégories du RPG : estives, prairies permanentes et prairies temporaires), profondeur du sol et altitude dans un rayon de 60 Km et 100 Km autour d'Avignon (Vicente-Vicente et al., 2021). A droite, localisation des aires de production orientées vers les circuits de proximité qui peuvent être mobilisables dans une politique alimentaire localisée (Mouléry et al., 2022). En bas, schéma de l'approche méthodologique en trois étapes, utilisée pour passer de l'analyse du foodshed limitée à un cercle de production isotrope autour de la ville à son caractérisation (par groupe de produits) en tant qu'une configuration spatiale d'aires de production (patches) en archipel à partir des caractéristiques biophysiques et socio-économiques (Sanz Sanz et al. 2021).

3. Bilan et perspectives. L'analyse spatiale multi-échelle des agricultures contribuant à l'approvisionnement de proximité, clé d'analyse du nexus agriculture-alimentation-villes

4. Conclusion

En termes de conclusion, les travaux que nous souhaitons présenter mettent en lumière les déterminants des systèmes alimentaires territoriaux en mobilisant les cadres théoriques des relations agriculture/alimentation/villes. Leur originalité réside dans la conception et le développement de cadres d'analyse quantitatifs, spatialisés et multiscalaires combinant des données hétérogènes, et permettant de caractériser la forme et la localisation des relations entre aires de production agricoles et aires de consommation alimentaires. En effet, les outils que nous avons mobilisés se veulent complémentaires, eu égard aux connaissances qu'ils visent à mobiliser : des modèles statistiques aux échelles supérieures (ex. le bassin d'approvisionnement), aux enquêtes de terrain aux échelles locales. Enfin cette approche géomatique permet de livrer des résultats sous forme systémique, mais également la production de cartographies multicritères et la simulation des effets des différentes options de politiques publiques, au moyen d'approches modélisatrices et mobilisables dans une démarche d'accompagnement de la planification alimentaire. Il s'agit, pour nous, de fournir, au-delà de l'innovation académique, des méthodologies systémiques opérationnelles eu égard de la planification alimentaire qui permettent : a) d'identifier des éléments actionnables pour l'action publique et b) de focaliser l'action publique dans les zones où elle a de fortes probabilités d'être efficiente.

Remerciements :

Les travaux présentés sont issus : 1/ de la thèse de doctorat de Sanz Sanz (2016) qui a été financée par l'ANR via une convention industrielle de formation par la recherche (CIFRE n°2012/0608) entre l'Agence Paysages et INRAE (unité Ecodeveloppement, Avignon) ; 2/ des recherches financées par la Fondation de France dans le cadre du projet Aliville (n° d'agrément E 00061222) ; 3/ des recherches financées par le programme de pré-maturation d'INRAE dans le cadre du projet Agriville-USAP en collaboration avec la Société d'aménagement foncier et d'établissement rural (SAFER) et 4/ des recherches financées par le programme Horizon 2020/FoodShift20230 de l'Union Européenne (n° d'agrément 862716).

Bibliographie

- Alonso, W. (1964). *Location and land use. Toward a general theory of land rent*. Harvard University Press.
- Boussougou Boussougou, G., Sanz Sanz, E., Napoléone, C., Martinetti, D. (2021). Identifying agricultural areas with potential for city connections: A regional-scale methodology for

urban planning. *Land Use Policy*, 103, 105321.
<https://doi.org/10.1016/j.landusepol.2021.105321>

Brinkley, C. (2013). Avenues into Food Planning: A Review of Scholarly Food System Research. *International Planning Studies*, 18, 243–266.
<https://doi.org/10.1080/13563475.2013.774150>

CIRAD-SAR (1996). *Systèmes agroalimentaires localisés : organisations, innovations et développement local*. Orientations et perspectives issues de la consultation du CIRAD « Stratégies de recherche dans le domaine de la socio-économie de l’alimentation et des industries agroalimentaires » No. 134/96. Département des systèmes agroalimentaires et ruraux CIRAD-SAR.

Décamps, H., Décamps, O. (2007). Organisation de l’espace et processus écologiques. *Economie rurale*, 41–54. <https://doi.org/10.4000/economierurale.1990>

Deffontaines, J.-P. (1999). Du paysage comme moyen de connaissance de l’activité agricole à l’activité agricole comme moyen de production du paysage. Un point de vue d’agronome. *Compte-rendus de l’Académie d’agriculture*, 57–69.

Deffontaines, J.P., Thenail, C., Baudry, J. (1995). Agricultural systems and landscape patterns: how can we build a relationship? *Landscape and Urban Planning*, 31, 3–10.
[https://doi.org/10.1016/0169-2046\(94\)01031-3](https://doi.org/10.1016/0169-2046(94)01031-3)

Deffontaines, J.-P., Thinon, P. (2006). Emergence d’un concept : Un itinéraire entre agronomie et géographie. Chapitre 5, in: Legay, J.-M. (Ed.), *L’interdisciplinarité Dans Les Sciences de La Vie*, Indisciplines. Éditions Quae, pp. 45–50.

Erb, K.-H., Lauk, C., Kastner, T., Mayer, A., Theurl, M.C., Haberl, H. (2016). Exploring the biophysical option space for feeding the world without deforestation. *Nature Communications*, 7, 11382. <https://doi.org/10.1038/ncomms11382>

FAO/RUAF/WLU (2018). *City Region Food System Toolkit Assessing and planning sustainable city region food systems*. FAO, RUAF Fondation, Wilfrid Laurier University.

HLPE (2017). *Nutrition and food systems*. A report by the High Level Panel of Experts on Food Security and Nutrition (HLPE reports series No. 12). CFS Committee on World Food Security.

IPES-Food (2020). *COVID-19 and the crisis in food systems: Symptoms, causes, and potential solutions*. International Panel of Experts on Sustainable Food Systems.

Kinnunen, P., Guillaume, J.H.A., Taka, M., D’Odorico, P., Siebert, S., Puma, M.J., Jalava, M., Kummu, M. (2020). Local food crop production can fulfil demand for less than one-third of the population. *Nature Food*, 1, 229–237. <https://doi.org/10.1038/s43016-020-0060-7>

Lev, L., Stevenson, G. (2011). Acting Collectively to Develop Mid-Scale Food Value Chains. *Journal of Agriculture, Food Systems, and Community Development*, 1, 119–128.
<https://doi.org/10.5304/jafscd.2011.014.014>

Mills, E.S. (1972). *Studies in the structure of the urban economy*. Johns Hopkins Press, Baltimore.

Mouléry, M., Sanz Sanz, E., Debolini, M., Napoléone, C., Josselin, D., Mabire, L., Vicente-Vicente, J.L. (2022). Self-Sufficiency Assessment: Defining the Foodshed Spatial

- Signature of Supply Chains for Beef in Avignon, France. *Agriculture*, 419.
<https://doi.org/10.3390/agriculture12030419>
- Moustier, P., Holdsworth, M., Anh, D.T., Seck, P.A., Renting, H., Caron, P., Bricas, N. (2023). The diverse and complementary components of urban food systems in the global South: Characterization and policy implications. *Global Food Security* 36, 100663.
<https://doi.org/10.1016/j.gfs.2022.100663>
- Muth, R.F. (1969). *Cities and housing: the spatial pattern of urban residential land use*. University of Chicago Press.
- Pierr, H-P. (2003). Environmental policy, agri-environmental indicators and landscape indicators. *Agriculture, Ecosystems & Environment*, 98, 17–33.
[https://doi.org/10.1016/S0167-8809\(03\)00069-0](https://doi.org/10.1016/S0167-8809(03)00069-0)
- Poggi, S., Vinatier, F., Hannachi, M., Sanz Sanz, E., Rudi, G., Zamberletti, P., Tixier, P., Papaix, J. (2021). How can models foster the transition towards future agricultural landscapes? *Advances in Ecological Research* S0065250420300453.
<https://doi.org/10.1016/bs.aecr.2020.11.004>
- Rastoin, J.-L. (2015). Éditorial. Les systèmes alimentaires territorialisés : considérations théoriques et justifications empiriques. *Économies et Société*, Série Systèmes agroalimentaires AG, Tome XLIX, n°37, 1155–1164.
- Sanz Sanz, E. (2016). *Planification urbaine et agriculture. Méthodologie systématique de caractérisation de l'agriculture périurbaine à partir d'une recherche empirique en France et en Espagne*. Thèse pour l'obtention du titre de docteur en géographie (UAM) et en études urbaines (EHESP).
- Sanz Sanz, E., Napoléone, C., Hubert, B. (2017a). Caractériser l'agriculture périurbaine pour mieux l'intégrer à la planification urbaine: propositions méthodologiques. *L'Espace géographique*, 46, 174–190. <https://doi.org/10.3917/eg.462.0174>.
- Sanz Sanz, E., Napoléone, C., Hubert, B., Mata, R., Giorgis, S. (2017b). Repenser la planification urbaine à partir des espaces agricoles. *Revue d'Économie Régionale & Urbaine*, 511–536. <https://doi.org/10.3917/revu.173.0511>
- Sanz Sanz, E., Martinetti, D., Napoléone, C. (2018). Operational modelling of peri-urban farmland for public action in Mediterranean context. *Land Use Policy*, 75, 757–771.
<https://doi.org/10.1016/j.landusepol.2018.04.003>
- Sanz Sanz, E., Napoléone, C., Mouléry, M. (2021). Agroecological transitions and local food self-sufficiency assessment. From the isotropic circle to the archipelago foodshed. Agroecological transformation for sustainable food systems. Insight on France-CGIAR research., *Les dossiers d'Agropolis internationale*, 26, 98. <https://doi.org/DOI:10.23708>
- Sanz Sanz, E., Walthall, B., Napoléone, C., Vicente-Vicente, J.-L., Hinojosa, L., Pierr, A. (2023). Choosing modelling approaches for participatory food governance in city-regions. Comprehensive guidelines for a system-perspective selection. *Environmental Science & Policy* 145, 139–150. <https://doi.org/10.1016/j.envsci.2023.03.021>
- Schreiber, K., Hickey, G.M., Metson, G.S., Robinson, B.E., MacDonald, G.K. (2021). Quantifying the foodshed: A systematic review of urban food flow and local food self-sufficiency research. *Environmental Research Letters*, 16, 023003.
<https://doi.org/10.1088/1748-9326/abad59>

- Sonnino, R. (2016). The new geography of food security: exploring the potential of urban food strategies. *The Geographical Journal*, 182, 190–200. <https://doi.org/10.1111/geoj.12129>
- Taylor, P.D., Fahrig, L., Henein, K., Merriam, G. (1993). Connectivity Is a Vital Element of Landscape Structure. *Oikos*, 68, 571–573. <https://doi.org/10.2307/3544927>
- Torre, A., Wallet, F. (2014). Introduction: the role of proximity relations in regional and territorial development processes, in: *Regional Development and Proximity Relations*. Edward Elgar Publishing, pp. 1–44. <https://doi.org/10.4337/9781781002896.00006>
- Verburg, P.H., Mertz, O., Erb, K.-H., Haberl, H., Wu, W. (2013). Land system change and food security: towards multi-scale land system solutions. *Current Opinion in Environmental Sustainability*, 5, 494–502. <https://doi.org/10.1016/j.cosust.2013.07.003>
- Véron, F. (Ed.) (2003). *Les cahiers de la multifonctionnalité* n.2. Propositions de recherche soutenues par le dispositif INRA-CEMAGREF-CIRAD sur la multifonctionnalité de l’agriculture et des espaces ruraux. INRA Editions.
- Vicente-Vicente, J.L., Sanz-Sanz, E., Napoléone, C., Moulery, M., Piorr, A. (2021). Foodshed, Agricultural Diversification and Self-Sufficiency Assessment: Beyond the Isotropic Circle Foodshed—A Case Study from Avignon (France). *Agriculture* 11, 143. <https://doi.org/10.3390/agriculture11020143>
- Vidal, R., Fleury, A. (2009). Aménager les relations entre la ville et l’agriculture. De nouveaux enjeux territoriaux et une nouvelle approche «agriurbaniste». *Urbia. Les cahiers du développement urbain durable* 127–142.
- Viljoen, A., Wiskerke, J.S.C. (Eds.) (2012). *Sustainable food planning evolving theory and practice*. Wageningen Academic Publishers, The Netherlands.
- Von Thünen, J.H. (1826). *Von Thunen’s isolated state*. Pergamon Press, Glasgow.
- Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L.J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J.A., De Vries, W., Majele Sibanda, L., Afshin, A., Chaudhary, A., Herrero, M., Agustina, R., Branca, F., Lartey, A., Fan, S., Crona, B., Fox, E., Bignet, V., Troell, M., Lindahl, T., Singh, S., Cornell, S.E., Srinath Reddy, K., Narain, S., Nishtar, S., Murray, C.J.L. (2019). Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *The Lancet*, 393, 447–492. [https://doi.org/10.1016/S0140-6736\(18\)31788-4](https://doi.org/10.1016/S0140-6736(18)31788-4)
- Zasada, I., Schmutz, U., Wascher, D., Kneafsey, M., Corsi, S., Mazzocchi, C., Monaco, F., Boyce, P., Doernberg, A., Sali, G., Piorr, A. (2019). Food beyond the city – Analysing foodsheds and self-sufficiency for different food system scenarios in European metropolitan regions. *City, Culture and Society*, 16, 25–35.

Proposition méthodologique pour identifier la limite spatiale ville-agriculture : caractérisation de la rugosité du contour

AAA ¹, BBB ², CCC ¹, DDD ¹

1. labo caché

anonyme1@email

2. université cachée

anonyme2@email

RÉSUMÉ. Délimiter la forme urbaine est une étude complexe, que tout acteur de l'aménagement du territoire et des différentes disciplines scientifiques (géographie, économie, écologie du paysage) étudie pour de multiples réflexions (exemple : mesure de l'étalement urbain, détection des zones urbaines ou périurbaines, proximité des zones agricoles par rapport à la ville). Dans notre cas, selon la théorie de Brinkley, ce travail part de l'hypothèse d'un effet positif de la rugosité (complexité) de la limite urbaine, qui influencerait positivement les conditions de connexion de l'agriculture à proximité vers les marchés locaux (Brinkley, 2018). Dans ce contexte, nous proposons une méthode en morphologie mathématique (dilatation / érosion) pour détecter la limite urbaine la plus pertinente en fonction de l'échelle. Cette étude se localise sur 13 villes du sud de la France (département du Vaucluse et des bouches du Rhône) parmi les plus productives en fruits et légumes en circuits de proximité.

ABSTRACT. Delineating the urban form is a complex task, involving all actors of land use planning and different scientific disciplines (geography, economics, landscape ecology) for multiple reflections and objectives (measurement of urban sprawl, detection of urban areas, proximity of agricultural areas to the city). In our case, according to Brinkley's theory, this work states on the assumption that the "rugosity" (complexity) of the urban boundary would positively influence the connection conditions of nearby agriculture to local markets (Brinkley, 2018). In this context, we propose a method in mathematical morphology (dilation/erosion) to detect the most relevant urban boundary at different scales. This study will be located on 13 towns in the south of France (Vaucluse and Bouches du Rhône departments), some of the most productive in fruit and vegetable, targeting local short supply chain.

MOTS-CLÉS : Forme urbaine, rugosité, analyse d'image, morphologie mathématique,

SAGEO'2023 - Québec, 5-9 juin 2023

fractale

KEYWORDS: : urban form, rugosity, mathematic morphology, fractal

1. Introduction

70 % de la population européenne vit dans une aire urbaine dont la surface s'est accrue de 80 % en 50 ans (EEA, 2006). L'expansion urbaine a, ainsi, constitué le processus de changement d'usage du sol le plus significatif en Europe depuis la 2^{de} guerre mondiale, et connaît un rythme de croissance environ deux fois plus rapide que l'augmentation de la population (33 % depuis les années 1950). L'urbanisation devrait continuer d'augmenter de 0,5 à 0,7% par an, ce qui représente un rythme 10 fois plus important que n'importe quel autre changement d'occupation du sol (Nilsson *et al.*, 2013). L'espace géographique terrestre étant fini, cette expansion des zones urbaines, en dépit des politiques suivies de concentration urbaine, de ville compacte et de lutte contre l'étalement, se fait au détriment des milieux naturels et des surfaces agricoles.

On voit apparaître en quelque sorte une « zone de combat » due à des intérêts divergents : installer des nouveaux foyers ayant accès aux facilités urbaines (politiques de la ville), conserver une capacité à produire des aliments pour la population (politiques agricoles, avec, par exemple, le rôle de la SAFER¹). Cette question est d'autant plus vive que l'Europe traverse une situation de tension sur les marchés agricoles consécutive à la guerre en Ukraine. Dans ce contexte, la question de l'alimentation des populations urbaines se pose avec d'autant plus d'acuité que les tendances politiques du moment cherchent à accroître le nombre et l'intensité des circuits de proximité (Voglhuber-Slavinsky *et al.*, 2021 ; Praly *et al.*, 2014), les crises du Covid-19 ou d'Ukraine soulignant la fragilité des filières alimentaires globalisées (Ugaglia *et al.*, 2021). Or, des positions alternatives existent, eu égard aux conséquences de la croissance urbaine sur l'agriculture de proximité. On peut en effet s'interroger sur les vertus potentielles des formes urbaines « compactes » dans une optique de développement durable prôné par l'Union européenne, par rapport à des formes plus étalées ou diffuses. Selon Brinkley (Brinkley, 2018), la complexité de la limite du contour de contact entre zones urbaines et agricoles génère une plus grande probabilité de contact et pourrait favoriser la demande urbaine pour une agriculture située à proximité des limites, voire le long des branches urbaines. En partant de cette théorie, il convient alors de mesurer plus concrètement la forme de cette limite, souvent complexe, car elle doit prendre en compte la distribution spatiale des interstices de l'agriculture dans le milieu urbain. Dans ce contexte, notre objectif est de construire une méthode permettant d'identifier la limite

1. Société d'aménagement foncier et d'établissement rural

ville-agriculture en fonction des configurations spatiales observées à différentes échelles. Cette méthode se base sur l'hypothèse de décrite précédemment. Cela pose trois types de questions : 1/ Comment délimiter un contour urbain à différentes échelles 2/ Quelle métrique utiliser pour caractériser sa forme ? 3/ Quelle typologie des contacts ville-campagne déduire de ces mesures ?

2. Matériels et Méthodes

2.1. Zone d'étude

Pour cette étude, 13 villes voisines de deux départements du Vaucluse et des bouches du Rhône (sud de la France) ont été étudiées. Historiquement, ces villes ont toujours été fortement productives en fruits et légumes (Figure 1) grâce aux conditions pédo-climatiques et bioclimatiques très favorables. Ces dernières décennies, ces villes ont connu de forts bouleversements en matière de croissance démographique (cf. INSEE ²).

Ces événements ont accentué une forte demande en logement (Figure 2) causant une réduction des terres agricoles et la perte de nombreuses exploitations agricoles (Figure 3), toutes filières confondues. Ce scénario a ralenti le processus de commercialisation et fragilisé fortement les relations vers la ville, en dépit des besoins croissants. Nous choisissons la production en fruits et légumes (Figure 1) , ces productions étant davantage en contact avec la ville et ses marchés locaux, que d'autres productions vouées à l'exportation, telles que celles de la filière céréalière (Boussougou Boussougou *et al.*, 2021).

2.2. Données utilisées

Pour le volet urbain, nous utilisons la source de données géomatiques d'accès libre pour le bâti, la BD TOPO® 2018 de l'IGN. Comme référencé dans le cahier des charges, cette source contient une description des éléments du paysage sous forme de vecteurs de précision métrique, classés selon une thématique adaptée. Elle couvre de manière cohérente l'ensemble des entités géographiques et administratives du territoire national. Sa résolution spatiale est de 1/25 000ème. Elle possède 3 versions de 1999 jusqu'à 2019. Nous utiliserons les classes « bâti indifférencié et industriel et commercial (sauf les bâtiments agricoles et serres) » pour tracer les contours de la ville. Pour la partie agricole, nous utilisons le Recensement Parcelaire Graphique de 2018 qui provient des demandes des exploitants agricoles pour l'obtention des primes européennes (PAC) ³. Cette source possède une information vectorisée avec une nomenclature agricole très détaillée et précise à l'échelle de la parcelle et de l'exploitation agricole. En complément,

2. <https://www.insee.fr/fr/information/2008354>

3. Politique Agricole Commune

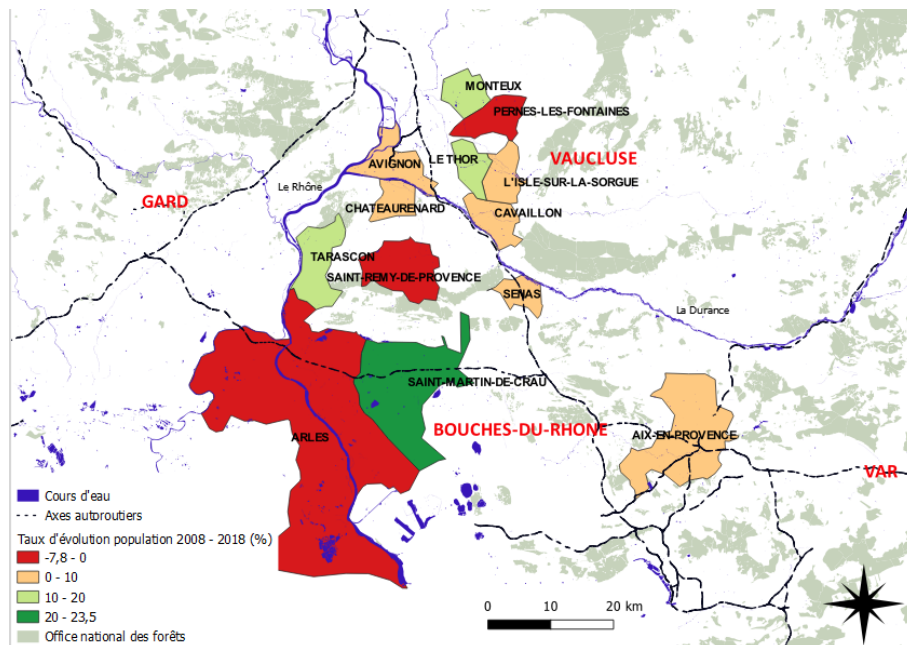


Figure 1. Sélection des villes les plus fortement orientées en fruits et légumes

Villes	Surface Région Urbaine Km2	Nombre Population 2008	Nombre Population 2018	Taux de variation Population 2008 - 2018 (%)	Nombre logement 2008	Nombre logement 2018	Taux de variation Logement 2008 - 2018 (%)	Nombre Emploi 2008	Nombre emploi 2018	Taux de variation Emploi 2008 - 2018 (%)
Aix en Provence	186	142743	143097	+0,2%	77561	84308	+8,7%	85136	94315	+10,78%
Arles	758,9	52729	51031	-3,22%	26371	27883	+5,73%	20317	18849	-7,22%
Avignon	64,78	90109	91729	+1,8%	48682	53970	+10,86%	58629	57824	-1,37%
Cavaillon	45,96	25417	26198	+3,07%	12184	13893	+14,02%	14603	13886	-4,91%
Chateaurnaud	34,95	14817	16012	+8,07%	6844	8057	+17,71%	5243	5703	+8,77%
Isle sur la Sorgue	44,57	18933	19706	+4,08%	9595	11250	+17,25%	6563	6945	+5,83%
Le Thor	35,53	7916	8996	+13,64%	3446	4288	+24,43%	2244	2508	+11,78%
Monteux	39,02	10789	12875	+19,33%	4875	5953	+22,1%	3492	4098	+3,96%
Pernes les Fontaines	51,12	10440	9975	-4,45%	4759	5035	+5,81%	2744	2426	-11,58%
Senas	30,61	6317	6930	+9,7%	2647	3201	+20,95%	1572	1571	-0,06%
Saint Martin de Crau	214,9	10979	13558	+23,49%	4878	6375	+30,69%	4161	5558	+33,57%
Saint Rémy de Provence	89,09	10662	9829	-7,81%	5735	6472	+12,84%	5013	5286	+5,45%
Tarascon	73,97	13514	15195	+12,49%	6162	7211	+17,02%	8872	9638	+8,63%

Figure 2. Fiche d'identité des villes sélectionnées (volet urbain)

nous utiliserons les données développées de l'ODR⁴ de INRAE de Toulouse de Cantelaube et de Lardot qui permet de cartographier un ensemble de parcelles agricoles absentes du RPG⁵ (Cantelaube, Lardot, 2022). Leur démarche se dé-

4. Observatoire de développement rural

5. Registre Parcellaire Graphique

Villes	Surface Région Urbaine Km2	Nombre Emploi Agricole 2008	Nombre Emploi Agricole 2018	Taux de variation Emploi Agricole 2008 2018	Nombre Exploitation Agricole 2010	Nombre Exploitation Agricole 2020	Taux de variation Exploitation Agricole 2008 2018
Aix en Provence	186	386	463	+19,96%	182	144	-20,88%
Arlès	758,9	858	798	-7,04%	450	420	-6,67%
Avignon	64,78	517	427	-17,24%	72	67	-6,94%
Cavaillon	45,96	465	347	-25,32%	107	85	-20,56%
Chateaufort	34,95	354	348	-1,87%	192	121	-36,98%
Isle sur la Sorgue	44,57	187	218	+16,9%	88	76	-13,64%
Le Thor	35,53	174	141	-18,83%	73	62	-15,07%
Monteux	39,02	206	276	+33,9%	117	74	-36,75%
Pernes les Fontaines	51,12	277	233	-15,82%	120	116	-3,33%
Senas	30,61	90	160	+77,09%	83	65	-21,69%
Saint Martin de Crau	214,9	380	320	-15,64%	170	169	-0,59%
Saint Rémy de Provence	89,09	483	387	-20,01%	211	136	-35,55%
Tarascon	73,97	321	272	-15,26%	145	117	-19,31%

Figure 3. Fiche d'identité des villes sélectionnées (volet agricole)

compose en deux étapes : (i) en premier lieu, leur travail consiste à repérer ces parcelles susceptibles d'accueillir des surfaces agricoles qui n'appartiennent pas à des exploitants recevant des aides PAC surfaciques, (ii) ensuite à déterminer les cultures présentes sur ces parcelles identifiées. Cette première sélection avec ces sources de données permet de représenter globalement le paysage agricole autour des villes et ainsi de repérer les zones de contacts entre zones urbaines et zones agricoles (Figure 4).

Sources de données	Mode	Classes sélectionnées	Année sélectionnée
BD TOPO®	vecteur	Bâti résidentiels, industriels et commerciaux	Version 3 – 2018
RPG	vecteur	Classes Fruit et Légumes	2018
RPG complété ODR	vecteur	Classes Fruit et Légumes	2018

Figure 4. Les sources de données utilisées

2.3. Comment délimiter un contour urbain à différentes échelles ?

Notre objectif est de concevoir une enveloppe urbaine en assemblant les éléments bâtis de la BDTOPO (résidentiels, industriels et commerciaux) au fil des échelles. L'objectif principal est de trouver une méthodologie la plus pertinente qui remplit ces conditions :

- souplesse de la méthode (rapidité informatique du traitement pour la création du contour urbain) ;
- recensement des sols agricoles hors des contours urbains, sur l'interface ;
- prise en compte des interstices du contour urbain et ce à différentes échelles.

Plusieurs méthodes existent, citons-en quelques-unes. Par géométrie fractale avec les travaux de l'équipe de Cecile Tannier (logiciel MorphoLim) qui identifient les limites morphologiques des agglomérations urbaines. Une dilatation de Minkowski est appliquée à chaque polygone de bâti de la BDTOPO à l'aide d'une zone tampon. La largeur de la zone augmente automatiquement selon une logique géométrique, qui correspond à la logique fractale. Les résultats sont représentés sous la forme d'un graphique log-log, où l'axe des X représente la largeur du buffer de dilatation et l'axe des Y représente le nombre correspondant de clusters (agrégation des éléments bâti) construits. Le seuil de distance choisi est situé sur la courbe de dilatation, correspondant à un changement majeur dans son comportement (Tannier *et al.*, 2011) Par télédétection, à partir d'images satellites cette approche permet de détecter les objets bâtis, de créer une enveloppe urbaine suivant différentes méthodes de classification, par des critères de distance, de forme, de texture ou de taille (Sidi Youssef *et al.*, 2014).

Par morphologie mathématique, la dilatation/érosion s'appuie sur des notions d'algèbre, de topologie et de probabilité simple. La dilatation est d'abord appliquée aux bâtiments en fonction d'un premier seuil de distance tampon positif qui permet de dessiner une première enveloppe urbaine. Ensuite, l'érosion de cette nouvelle géométrie polygonale, correspondant en quelque sorte à une distance tampon négative, permet d'ajuster la zone bâtie aux entités de référence (Le Berre, Brosset, 2020). Cette double approche permet, à différentes échelles, de regrouper les entités spatiales ou de conserver les connexions entre elles, lorsqu'elles sont proches, denses, et de taille suffisante par rapport aux seuils de distances utilisés en dilatation et érosion. Nous choisissons la méthode de dilatation / érosion dans la suite de notre étude pour créer l'enveloppe urbaine. Un autre avantage de cette méthode est la précision de cet algorithme, mais aussi la rapidité d'exécution du traitement (créé à l'aide d'une fonction Postgis). Ainsi les enveloppes urbaines peuvent être générées à l'échelle d'une région complète, contrairement aux approches de télédétection, où les étapes pour arriver à ce même résultat sont parfois plus conséquentes (récupération des images satellites, assemblage des rasters, post-traitements, phase d'ajustement et d'interprétation). Le second avantage est que l'opérateur peut facilement fixer une gamme de seuils d'échelle par pas variables (dans notre cas, de 20 m jusqu'à 100 m), permettant de visualiser au fil des échelles l'agencement du bâti. Nous n'avons pas retenu le logiciel Morpholim, car ses résultats produisent une synthèse transcalaire (à travers un jeu d'échelles), en s'affranchissant de seuil de distance prédéfini entre les bâtiments (Tannier *et al.*, 2021). Dans notre

cas, nous comparons les résultats de traitements sur notre série d'échelles, avec le paramètre de distance qui modifie les analyses. C'est la raison pour laquelle nous avons opté pour la méthode de dilatation-érosion.

2.4. Les effets de l'échelle

Nous pouvons nous poser la question du choix des valeurs satisfaisantes les paramètres de dilatation (d) et d'érosions (e) pour le calcul des zones tampons. En effet, une enveloppe créée par dilatation-érosion (d/e) entre les bâtis distants de 100 mètres, provoquera des conurbations sur les villes avoisinantes, une enveloppe urbaine trop large, englobera les sols agricoles, éradiquant les interstices. Dans le cas contraire, une résolution fine avec une faible dilatation/érosion fragmentera le tissu urbain, presque à l'échelle du quartier et générera trop de complexité pour notre étude. L'échelle adaptée est peut-être celle présentée en (Figure 5), car elle discrimine correctement les sols agricoles à l'extérieur de la limite et conserve les interstices. Cependant, il est intéressant d'explorer les variations de la limite à travers les échelles, en faisant varier les valeurs de d et e. de ce fait, nous maintenons la méthode du jeu multi-échelle pour analyser et comprendre la géographie de la ville au fil des échelles (agencement du bâti, effet barrière naturelle, comme les rivières ou les forêts) et analyser la variabilité de la rugosité (complexité du contour).



Figure 5. Dilatation et érosion de 40 m, ville du Thor. Les zones agricoles sont en dehors de la limite

2.5. Quelle métrique utiliser pour caractériser la forme et la rugosité urbaines ?

Le degré de complexité (la rugosité) du contour de la limite urbaine doit être mesuré, quantifié par des métriques robustes (Figure 6). En 1822, un des premiers indicateurs a été recherché par Ritter, qui a mesuré la compacité d'une forme en utilisant un simple rapport entre le périmètre et la surface de la forme (Purevtseren *et al.*, 2018). Utilisé dans beaucoup de disciplines géographie, écologie du paysage, climatologie (bassins versants), une multitude d'indicateurs mesurant la complexité ont été pensés. Parmi eux, deux indicateurs de forme ont attiré notre attention :

- L'indice de GRAVELIUS est le rapport entre le périmètre de la tâche et la circonférence d'un cercle de même surface l'entourant (Barké *et al.*, 2017); il est utilisé par exemple pour les études sur les bassins versants (Douvinet *et al.*, 2008)
- FRAC est un indicateur fractal basé sur les logarithmes népériens provenant de la théorie de Mandelbrot, qui reflète la complexité de la forme à travers les gammes d'échelles spatiales (Herold *et al.*, 2002)

Indicateur	Fonction	Formule	Amplitude	Remarque
Gravelius (G)	Indice de Compacité	$\frac{\text{périmètre}}{(2\sqrt{\text{surface}})}$	1 à l'infini	Si G augmente (tend vers infini), la complexité de la forme augmente
FRAC (F)	Index de dimension fractale	$\frac{2 \cdot \ln(\text{périmètre})}{\ln(\text{surface})}$	$1 \leq \text{FRAC} < 2$	1 : forme simple du périmètre, 2 : forme très irrégulière

Figure 6. Indicateur de rugosité

Nous avons testé les indicateurs Gravelius et FRAC en les appliquant sur trois sources de données de résolution différente (environ 400 m², 25 ha et environ 1 km²) sur des éléments bâtis. Pour chaque polygone, nous avons calculé la moyenne des indicateurs Gravelius et FRAC (Figure 7) sur le département du Vaucluse. Nous constatons que plus la résolution est fine, plus les valeurs de ces deux indicateurs augmentent, montrant leur consistance. Par exemple, l'indicateur de Gravelius varie de 1,4 pour 1 km² de résolution, à 3,96 pour 400 m². Finalement, nous choisissons l'indicateur de Gravelius, car il discrimine mieux le niveau de rugosité, l'indice FRAC, quant à lui, montrant des valeurs (trop) proches de 1 (Figure 7) .

2.6. Étude de la rugosité à travers les échelles

Dans la suite de notre étude, nous avons réfléchi à une catégorisation des signatures de rugosité de toutes nos villes par la méthode du delta (R). La

Sources de données	Résolution	Classe	FRAC	GRAVELIUS
OSO 2018	20 m	Bâti dense et bâti diffus	1,17	3,96
Corrine Land Cover 2018	25 HA	Tissu urbain continu et discontinu	1,07	2,07
GHLS 2015	1 km	suburban or periurban et dense urban	1,02	1,4

Figure 7. Test des indicateurs de rugosité sur le département du Vaucluse avec des surfaces > 10 ha

variation d'amplitude de R est calculée par les valeurs de l'érosion de la valeur max rapportée à la valeur min pour un même seuil de dilatation. Cette méthode nous permet de mesurer les écarts relatifs (avec l'indice de Gravelius) entre les seuils d'érosion (e) minimal et maximal, pour une même valeur de dilatation (d). Par exemple, pour une dilatation de 80 m, on calcule l'écart entre les valeurs de l'indice de Gravelius avec une érosion (maximale) de 80 m et minimale de 20 m.

Ce faisant, nous obtenons deux cas de figures opposées : une zone très dilatée et peu érodée (avec peu de circonvolutions) et une zone ayant subi les deux processus opposés (dilatation/érosion) de même force, présentant davantage de rugosité. L'écart entre les deux indique le changement de structure de la limite en fonction des échelles, variant selon les paramètres de la méthode. Il est réutilisé pour classer les villes étudiées. Pour une dilatation donnée d, on obtient la formule suivante :

$$Delta_d = R = g_{\{d;max(e)\}} - g_{\{d;min(e)\}}$$

avec :

- R = l'indice Delta
- G = indice de Gravelius
- d = valeur de la dilatation
- e = valeur de l'érosion

Après étude des résultats et des courbes Delta, nous classifions nos villes en trois configurations (Figure 8), selon leur ressemblance (cette classification est manuelle et non automatisée, car l'échantillon est de petite taille) :

- configuration n° 1 : Les villes possèdent des courbes Delta de type continu, qui tendent vers zéro (vers l'axe X) ;
- configuration n°2 : Les villes possèdent des courbes Delta montrant une rugosité moyenne (à partir de 60 m en dilatation) ;

– configuration n°3 : Les villes possèdent des courbes Delta globalement élevées.

Notons qu'il faut au moins 2 valeurs (minimale et maximale) pour calculer la courbe Delta : la dilatation de 20 m n'est pas prise en compte ; ainsi les valeurs en abscisse commencent à 40 m.

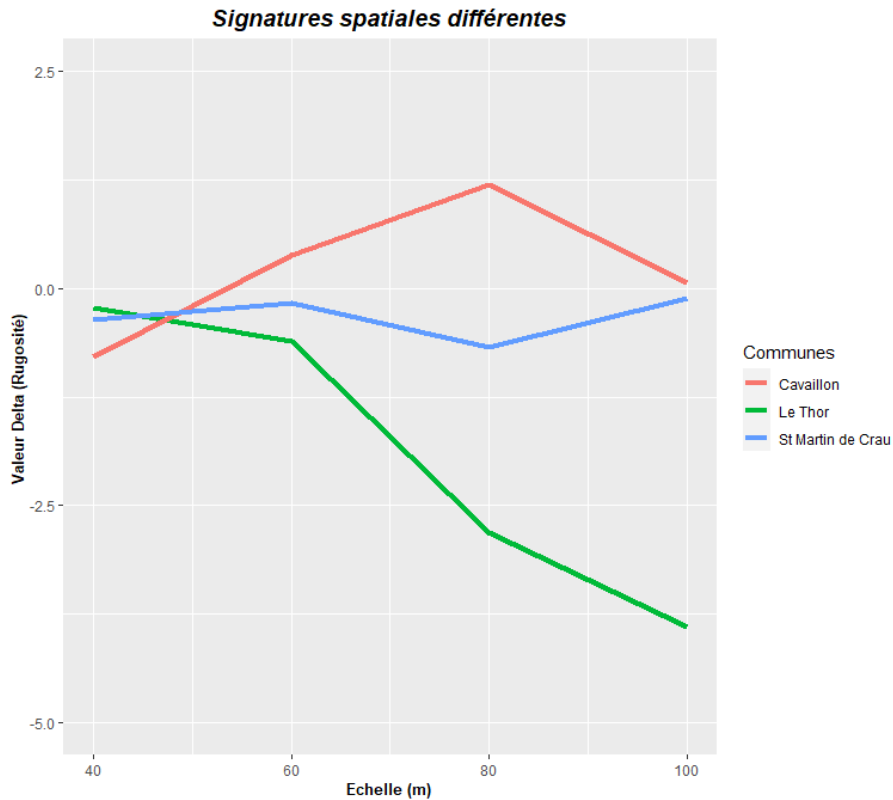


Figure 8. Exemple de signatures de rugosité R (courbe Delta) pour nos villes classifiées en trois groupes

La courbe d'amplitude de l'indice de la rugosité calculée avec l'indicateur de Gravelius en fonction du delta R (Figure 8) illustre les « signatures » morphologiques typiques des limites des zones urbaines étudiées, intégrant une dimension scalaire. Les calculs du delta de ces 3 villes sont référencés pour Cavaillon en Figure 10, Le Thor en Figure 11 et Saint Martin de Crau en Figure 12.

3. Premiers Résultats

3.1. Résultats : Typologie des contacts ville-campagne déduction des mesures

À l'aide de la méthodologie basée sur les courbes d'amplitude delta, le paysage géographique de nos 13 villes se dessine, que ce soit par l'agencement des éléments bâti ou des effets barrières (fleuve, forêt) qui réduit cet étalement urbain. Ainsi nous avons classifié nos villes en trois catégories suivant un code couleur (Figure 9):

- En bleue : la courbe d'amplitude de rugosité (delta) tend vers 0 , le bâti est accolé est proche de la limite de la ville (Saint Remy de Provence, Saint Martin de Crau, Avignon, Tarascon).
- En vert : la courbe d'amplitude est de rugosité moyenne, le bâti est de type tache d'huile, des branches urbaines apparaissent (Pernes, Cavailon, Isle Sur Sorgues, Monteux, Senas, Arles)
- En jaune : la courbe d'amplitude est de rugosité forte, le bâti est constitué en tache léopard, avec des branches urbaines plus étendues et plus complexes (Le Thor, Aix, Chateaurnaud)

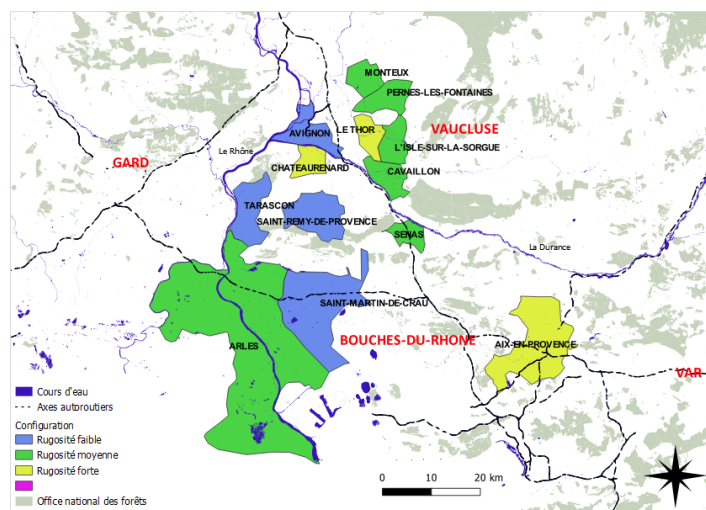


Figure 9. Classement des villes par courbe d'amplitude (Couleur : bleue=faible, verte=moyenne, jaune=forte)

4. Conclusion, discussion, perspectives

Nous avons proposé une méthode originale de classification multiscalaire des formes urbaines. Nous constatons que les villes de différentes tailles et/ou

Ville	Dilat_erosion	Gravelius	Delta
Cavaillon	40_20	5,86	
Cavaillon	40_40	5,08	-0,78
Cavaillon	60_20	4,18	
Cavaillon	60_40	3,75	
Cavaillon	60_60	4,57	0,39
Cavaillon	80_20	3,97	
Cavaillon	80_40	3,69	
Cavaillon	80_60	4,00	
Cavaillon	80_80	5,17	1,20
Cavaillon	100_20	3,23	
Cavaillon	100_40	3,22	
Cavaillon	100_60	4,13	
Cavaillon	100_80	4,00	
Cavaillon	100_100	3,30	0,07

Figure 10. Distribution Cavaillon (Min=3.22 Max=5.86 Moy=4.15 Med=4)

Ville	Valeur Dilat-erosion	Gravelius	Delta
Le Thor	40_20	3,56	
Le Thor	40_40	3,33	-0,23
Le Thor	60_20	3,63	
Le Thor	60_40	3,77	
Le Thor	60_60	3,02	-0,61
Le Thor	80_20	5,96	
Le Thor	80_40	5,73	
Le Thor	80_60	3,70	
Le Thor	80_80	3,15	-2,81
Le Thor	100_20	7,01	
Le Thor	100_40	6,94	
Le Thor	100_60	5,99	
Le Thor	100_80	4,24	
Le Thor	100_100	3,11	-3,90

Figure 11. Distribution Le Thor (Min=3.02 Max=5.99 Moy=4.51 Med=3.73)

Ville	Dilat_erosion	Gravelius	Delta
St-Martin-De-Crau	40_20	2,64	
St-Martin-De-Crau	40_40	2,28	-0,35
St-Martin-De-Crau	60_20	2,46	
St-Martin-De-Crau	60_40	2,27	
St-Martin-De-Crau	60_60	2,29	-0,17
St-Martin-De-Crau	80_20	2,83	
St-Martin-De-Crau	80_40	3,16	
St-Martin-De-Crau	80_60	2,99	
St-Martin-De-Crau	80_80	2,15	-0,68
St-Martin-De-Crau	100_20	2,88	
St-Martin-De-Crau	100_40	2,98	
St-Martin-De-Crau	100_60	2,98	
St-Martin-De-Crau	100_80	2,90	
St-Martin-De-Crau	100_100	2,76	-0,12

Figure 12. Distribution ST Martin de Crau (Min=2.27 Max=3.16 Moy=2.68 Med=2.80)

de forme (compacte vs étalée) et de géographie différentes ont parfois la même signature de rugosité, indiquant des ressemblances dans l'interpénétration des milieux (urbain vs agricole/naturel). Notre méthode, via les données utilisées pour appliquer la méthode de dilatation érosion, prend en compte la géographie du territoire et ses éléments barrières freinant l'étalement urbain (forêts, parcs protégés, rivières), montrant l'impact des politiques publiques de la ville (agencement des éléments bâtis dans l'espace géographique).

Dans nos prochains travaux, il pourra s'agir de trouver l'échelle pertinente pour préserver au maximum l'espace agricole (en termes de capacité de production de densité, mais aussi de proximité aux marchés de consommateurs) hors de la limite urbaine. Par ailleurs, dans la suite de notre étude et pour plus de précisions, un découpage géographique par maille sera nécessaire, afin de confronter si une relation statistique existe entre les éléments agricoles (Figure 4) et la rugosité de la limite de la ville.

En perspective finale, ce travail de thèse posera la question suivante : la rugosité de la limite de la ville a-t-elle finalement un effet positif sur la production agricole vers les marchés locaux ?.

5. annexe

Bibliographie

- Barké M. K., Oussein I., Biolders C., Ambouta K. J., Tychon B. (2017, janvier). Caractérisation morphologique des cuvettes oasiennes du Centre-Est du Niger. *Physio-Géo*, n° Volume 11, p. 255–276. Consulté sur <http://journals.openedition.org/physio-geo/5607>
- Boussougou Boussougou G., Sanz Sanz E., Napoléone C., Martinetti D. (2021, avril). Identifying agricultural areas with potential for city connections: A regional-scale methodology for urban planning. *Land Use Policy*, vol. 103, p. 105321. Consulté sur <https://linkinghub.elsevier.com/retrieve/pii/S0264837721000442>
- Brinkley C. (2018). Fringe benefits: Adding rugosity to the urban interface in theory and practice. *Journal of Planning Literature*, vol. 33, p. 143-144.
- Cantelaube P., Lardot B. (2022). *RPG complété 2018 département 84*. Recherche Data Gouv. Consulté sur <https://doi.org/10.15454/AEPLXS>
- Douvinet J., Delahaye D., Langlois P. (2008, février). Modélisation de la dynamique potentielle d'un bassin versant et mesure de son efficacité structurelle. *Cybergeo*. Consulté sur <http://journals.openedition.org/cybergeo/16103>
- Herold M., Scepan J., Clarke K. C. (2002, août). The Use of Remote Sensing and Landscape Metrics to Describe Structures and Changes in Urban Land Uses. *Environment and Planning A: Economy and Space*, vol. 34, n° 8, p. 1443–1458. Consulté sur <http://journals.sagepub.com/doi/10.1068/a3496>

- Le Berre I., Brosset D. (2020). Aide à la décision par l'identification de formes d'urbanisation : application au littoral breton. *Cybergeog: European Journal of Geography*, vol. 43.
- Nilsson K., Pauleit S., Bell S., Aalbers C., Nielsen T. S. (2013). Peri-urban futures: Scenarios and models for land use change in Europe. *Peri-Urban Futures: Scenarios and Models for Land use Change in Europe*, n° December, p. 1–453. (ISBN: 9783642305290)
- Praly C., Chazoule C., Delfosse C., Mindler P. (2014, décembre). Les circuits de proximité, cadre d'analyse de la relocalisation des circuits alimentaires. *Géographie, économie, société*, vol. 16, n° 4, p. 455–478. Consulté sur <http://ges.revuesonline.com/article.jsp?articleId=19888>
- Purevtseren M., Tsegmid B., Indra M., Sugar M. (2018). The fractal geometry of urban land use: The case of Ulaanbaatar City, Mongolia. *Land*, vol. 7, n° 2, p. 1–14.
- Sidi Youssef M. M., Mallet C., Chehata N., Le Bris A., Gressin A. (2014, juin). Détection de bâtiments à partir d'une image satellitaire par combinaison d'approches ascendante et descendante. In *Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014*. France. Consulté sur <https://hal.science/hal-00989026>
- Tannier C., Thomas I., Vuidel G., Frankhauser P. (2011). A Fractal Approach to Identifying Urban Boundaries. *Geographical Analysis*, vol. 43, n° 2, p. 211–227.
- Ugaglia A. A., Boutry O., Ferru M., Mathé J., Prévost B., Rivaud A. (2021, février). La crise de la covid-19, un levier de changement pour le système alimentaire français? *Revue de la régulation*, n° 29. Consulté sur <http://journals.openedition.org/regulation/18589>
- Voglhuber-Slavinsky A., Derler H., Moller B., Donitz E., Bahrs E., Berner S. (2021, septembre). Measures to Increase Local Food Supply in the Context of European Framework Scenarios for the Agri-Food Sector. *Sustainability*, vol. 13, n° 18, p. 10019. Consulté sur <https://www.mdpi.com/2071-1050/13/18/10019>

ORL : une Ontologie des Relations de Localisation pour l'aide à la localisation de victimes en montagne lors d'un appel aux secours

Mattia Bunel¹, Cécile Duchêne¹, Ana-Maria Olteanu-Raimond¹

1. LaSTIG, Université Gustave Eiffel, IGN-ENSG
prenom.nom@ign.fr

RESUME. Les relations de localisation, e.g. en face de la boulangerie, sont utilisées pour décrire une position. Cet article présente ORL, une ontologie des relations de localisation utilisées en montagne. ORL assure l'interopérabilité entre un secouriste et des outils et méthodes visant à l'aider, pendant un appel d'urgence, à comprendre où est la victime qu'il a au téléphone. Elle a été construite par analyse d'appels aux secours réels et confrontation avec l'ontologie GUM-3-Space. Ses concepts sont nommés avec un point de vue utilisateur et chacun correspond à une méthode différente de construction de la zone vérifiant la relation de localisation considérée avec un objet de référence donné. Elle contient 62 concepts, dont 52 sont instanciables, 11 assurent l'organisation hiérarchique, et 24 sont labellisés comme proches de concepts de GUM-3-Space.

ABSTRACT. Localisation relations, e.g. in front of the bakery, are used to describe positions. This paper presents ORL, an ontology of localisation relations used in mountainous areas. ORL is dedicated to ensuring interoperability between a rescue crew and tools and methods aiming at helping him understand where the victim on the phone is. It was set up by analysis of real rescue calls and comparison with the GUM-3-Space ontology. Its concepts are named with a user approach and each of them corresponds to a different method to build an area where the considered localisation relation is valid with a given reference object. It contains 62 concepts of which 52 can be instanciated, 11 ensure their hierarchical organisation and 24 are labelled as close to a GUM-3-Space concept.

MOTS-CLES : ontologie, relation spatiale, zone montagneuse, secours en montagne

KEYWORDS: ontology, spatial relation, mountainous area, mountain rescue

1. Introduction

Lorsqu'une personne blessée en montagne appelle les secours, l'un des enjeux de cet appel nommé phase d'alerte par les secouristes est de déterminer où se situe la victime. Les services de secours s'appuient souvent pour cela sur les technologies de type GPS mais il arrive que ce ne soit pas possible (*e.g.* si la personne n'a pas de téléphone portable qui capte le réseau données, ou si son GPS est désactivé et que son état la rend incapable de le réactiver). Dans ce cas, les secouristes se basent sur les informations données par la victime pour comprendre où elle peut être. La victime décrit alors sa position au moyen d'*indices* (Viry et Villanova-Oliver, 2019) tels que « je vois un lac » ou « je suis au-dessus du refuge du Couvercle ». Chaque indice constitue un élément de *géoréférencement indirect* (Hill et Zheng, 1999) : la position d'un objet à localiser (ici « je », i.e. la victime) est indiquée au moyen d'une relation (verbe « voir », « au-dessus ») avec un objet repère dont la position est supposée connue (« un lac », « le refuge du Couvercle »). L'objet à localiser est classiquement appelé *cible* et l'objet repère *site* (Vandeloise, 1986, p. 34). Nous appelons *relation de localisation* la relation qui décrit la position de la *cible* par rapport au *site* (Duchêne et al., 2019).

L'objectif du projet ANR CHOUCAS¹ était de fournir des méthodes et outils pour aider les secouristes à localiser efficacement une victime qui décrit ainsi sa position. Parmi les outils et les méthodes figurent : (1) des méthodes dites de *spatialisation* (Bunel, 2021), qui permettent de traduire un *indice* donné par la victime en une région de l'espace dans laquelle elle peut se trouver, qui constitue un géoréférencement direct ; (2) un environnement de géovisualisation permettant à un utilisateur secouriste, au fur et à mesure qu'il interprète ce que lui dit la victime, de rechercher et visualiser les objets repères mentionnés, mais aussi de saisir un indice sous la forme d'un ou plusieurs objets repères et d'une relation de localisation, puis de déclencher la spatialisation de cet indice par appel à la méthode dédiée et de visualiser la région ainsi obtenue (Viry, 2022).

Pour assurer l'interopérabilité entre le secouriste, les méthodes de spatialisation, l'environnement de géovisualisation, un point central a été la constitution de vocabulaires partagés, sous forme d'ontologies, permettant d'identifier, de décrire de manière non ambiguë et d'organiser d'une part les objets repères et d'autre part les relations de localisation utiles à la localisation en montagne. L'objectif de ce papier est de présenter l'Ontologie des Relations de Localisation construite durant le projet, appelée ORL², ainsi que sa méthode de constitution. L'ontologie ORL est destinée à plusieurs usages. D'abord, elle permet d'identifier des méthodes de *spatialisation* à développer. Une méthode de spatialisation est en effet valable pour une relation de localisation donnée (et un ou plusieurs types d'objets donnés). Son deuxième usage est d'indexer les méthodes de spatialisation définies : il s'agit d'étiqueter chaque

¹ ANR CHOUCAS : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne. Financé par l'ANR :ANR-16-CE23-0018.

² <http://purl.org/orl>

méthode de spatialisation développée avec la relation de localisation à laquelle elle est dédiée, de manière à pouvoir déclencher la bonne méthode de spatialisation pour un indice donné. Ensuite, l'ontologie ORL permet d'assurer l'interopérabilité entre l'interface de géovisualisation évoquée ci-dessus, utilisée par le secouriste, et les méthodes de spatialisation développées, en proposant au secouriste via l'interface une liste de relations de localisation pour lesquelles il existe bien une méthode de spatialisation.

La suite de l'article est organisée comme suit. La partie 2 présente un état de l'art sur les ontologies de relations spatiales existantes. Les parties 3 et 4 décrivent respectivement la méthode que nous avons utilisée pour construire l'ontologie ORL, et les choix de modélisation effectués. La partie 5 présente l'ontologie obtenue. Enfin, nous terminons classiquement par une conclusion et des perspectives.

2. Etat de l'art

Dans ce travail, nous utilisons le terme de relation de localisation pour englober des relations qui peuvent parfois sembler ne pas être spatiales, parce que leur expression ne fait pas explicitement appel à une préposition spatiale. C'est le cas par exemple dans les phrases « J'ai marché 2 heures depuis Aussois », « Je vois un lac », qui donnent pourtant des indications sur la localisation de « je » par rapport à Aussois, un lac, une route. En réalité, sémantiquement ces relations sont bien considérées comme des relations spatiales par les linguistes (Vandeloise, 1986) car elles peuvent en effet se ramener à des notions spatiales. A l'inverse, certaines relations spatiales ne permettent pas directement de se localiser (cf. section 4 et Duchêne *et al.*, 2019) et ne sont donc pas des *relations de localisation*.

La modélisation formelle d'une relation est étudiée dans la littérature aussi bien pour modéliser l'existence de relations spatiales entre objets telles qu'elles sont décrites dans un texte (Bateman *et al.*, 2010, Vasardani *et al.*, 2013, Pustejovsky, 2017), que pour modéliser des relations spatiales constatées entre des objets d'une base de données géographique (Touya *et al.*, 2012, Jaara *et al.*, 2012; Miron *et al.*, 2007). Un des formalismes classiquement choisi pour la modélisation des relations repose une ontologie exprimée en OWL (Ontology Web Language).

Sans viser l'exhaustivité, nous identifions ci-dessous quelques ontologies existantes qui portent une attention particulière à la modélisation des relations spatiales ou même explicitement des relations de localisation.

L'extension spatiale de l'ontologie GUM-3-Space proposée par (Bateman *et al.*, 2010), détaille à l'aide d'une hiérarchie assez complexe et complète de concepts, les valeurs que peut prendre une relation de localisation (appelée dans GUM-3-Space modalité spatiale), et définit chacun de ces concepts. On y retrouve des types de relations de localisation classiques telles que les relations cardinales, de distance, projectives, mais par exemple pour les directions cardinales une distinction est faite entre « être au nord d'un objet tout en étant situé dans cet objet » (i.e. dans la partie nord de l'objet), et « être au nord d'un objet sans être situé dans cet objet ». Des relations décrivant des configurations particulières figurent également, par exemple

« surrounding », classé dans une catégorie « relations en rapport avec la forme » (shape committing).

En géomatique, différentes ontologies d'application sont proposées. Par exemple, Touya et al., (2012) proposent un modèle formel destiné à la généralisation cartographique qui permet de représenter des relations spatiales entre objets, mais aussi les conditions ou les méthodes permettant d'assigner une valeur mesurée à une occurrence de relation (une distance), ou d'évaluer si une occurrence de relation est vérifiée pour les relations à modalités booléennes. Il est également possible d'explicitier une méthode qui permet d'évaluer s'il est pertinent de considérer un certain type de relation entre deux objets de types donnés (par exemple en généralisation on ne s'intéressera au parallélisme de deux objets que s'ils sont proches). Miron et al. (2007) proposent une ontologie de relations spatiales qualitatives, nommée ONTOAST pour inférer des nouvelles connaissances dans un jeu de données géographique. Les relations formalisées sont les relations de distance (proche, loin,), de direction (positions cardinales et au centre de) et topologiques.

Parmi ces ontologies, l'ontologie GUM-3-Space nous semble la plus adaptée à notre besoin. En effet, elle est proche de notre problème puisque dédiée à la formalisation d'un énoncé décrivant une situation de référencement indirect. Par ailleurs, elle est assez complète. Par exemple, elle distingue « dans la partie nord » et « au nord, à l'extérieur », ce qui est important dans notre cas d'application. Notre ontologie, ORL, s'inspire de l'ontologie GUM-3-Space. Selon la classification proposée par Guarino (1998), c'est une ontologie d'application.

Enfin, on trouve dans littérature la notion de niveau ou point de vue auquel on se place pour formaliser une relation spatiale. Ainsi, Clementini et Laurini (2008) identifient trois niveaux de formalisation possible : mathématique, informatique et utilisateur. Bucher et al. (2012) distinguent plus simplement deux points de vue : un point de vue computationnel (« strongly connected to computation »), basé sur des considérations géométriques sur les objets (*e.g.* les deux objets sont adjacents), et un niveau utilisateur (« strongly connected to our experience of reality »), qui met l'accent sur le sens dont ils sont porteurs (*e.g.* les bâtiments sont mitoyens). Duchêne et al. (2019) décrivent des relations de localisation selon ces deux points de vue. Le point de vue utilisateur nous semble adapté pour notre travail, puisque la sémantique des relations de localisations est notre préoccupation principale.

3. Méthode de création de l'Ontologie ORL

Deux sources principales ont été utilisées pour constituer l'ontologie ORL. La première est un corpus de 45 appels d'alerte téléphoniques réels reçus par les Pelotons de Gendarmerie de Haute-Montagne (PGHM), qui ont été identifiés, par le PGHM de Grenoble, comme les plus pertinents du point de vue de la localisation des victimes (*i.e.* des objets de repères et relations de localisations employés) et nous ont été fournis sous forme de fichiers audio. Ces alertes datent de la période 2015-2016 et sont pour la plupart issus de la période estivale, qui est la plus chargée pour le PGHM. Il est important de noter que ce corpus bien que représentatif, reste

relativement petit et ne peut être considéré comme exhaustif, c'est pourquoi la construction d'ORL repose sur une seconde source, l'ontologie de relations spatiales GUM-3-Space proposée par Bateman et al. (2010). Cette dernière est dédiée à l'interprétation d'un texte en langue naturelle décrivant des situations spatiales et présentée ci-dessus. Dans un travail préalable, les indices sur la localisation de la victime présents dans le corpus de 45 appels téléphoniques ont été extraits par écoute des fichiers audio et transcrits dans un tableau dédié permettant de les représenter de manière structurée (Bunel et al., 2019b). Dans ce tableau, une ligne correspond à un indice et des colonnes sont notamment présentes pour recueillir la phrase indice en langage naturel telle qu'elle a été prononcée puis, séparément, différents éléments qui peuvent en être extraits, et notamment : sujet (correspondant le plus souvent à la cible ou objet à localiser), verbe, relation de localisation, objet(s) de repères (correspondant le plus souvent au(x) site(s) ou objets dont la position est connue) (Figure 2, sauf la dernière colonne du tableau). Le tableau issu de la transcription des 45 alertes contient 377 lignes (correspondant donc à 377 indices).

Extrait	Sujet	Verbe	Relation Spatiale	Objet de référence 1	Objet de référence 2	Concept relation ontologie CHOUCAS
je suis sur la commune de St-Christophe sur Guiers en Isère,	victime	être	sur (dans)	Saint Christophe Sur Guiers		DansPlanimetrique
je suis sur la commune de St-Christophe sur Guiers en Isère,	victime	être	en (dans)	Isère		DansPlanimetrique
exactement au col de léchaud, sous la croix du petit Som	victime	être	à	Col de Léchaud		Proximal
exactement au col de léchaud, sous la croix du petit Som	victime	être	sous	Sommet du Petit Som		Sous_ProcheDe
C'est pas col de la Glière hein, par rapport au col de la Glière on est de l'autre côté de la Combe de la Glière	victime	être	de l'autre côté de... par rapport à...	Combe de la Glière	Col de la Glière	DeLAutreCoteDeParRapportA

Figure 2. Extrait du tableau utilisé pour recenser de manière structurée les indices contenus dans le corpus d'alertes utilisées pour la constitution de l'ontologie ORL. Cf. Bunel et al. (2019) pour une description de la structure du tableau.

La construction de l'ontologie ORL a alors débuté par un passage en revue de chaque ligne de ce tableau, consistant : (1) à examiner la relation spatiale utilisée dans l'indice modélisé dans cette ligne, (2) à chercher si un concept de relation spatiale équivalent ou proche existe dans l'ontologie GUM-3-Space de Bateman et al. (2010), (3) à tagguer la ligne dans une colonne dédiée avec un nom de relation provisoire qui paraissant approprié, proposer une première définition de cette relation, et noter en commentaire l'existence d'une relation équivalente ou d'une relation proche dans l'ontologie GUM-3-Space.

Ce travail a été réalisé séparément par les auteurs de cet article, puis les résultats ont été confrontés et harmonisés. La dernière colonne du tableau de la Figure 2 contient les noms de relations retenus à l'issue de la phase d'harmonisation. Les colonnes dédiées à l'identification d'une équivalence ou proximité de concept avec l'ontologie GUM-3-Space de Bateman et al. (2010) n'apparaissent plus dans le tableau à ce stade, car les équivalences et proximités de concepts ont été factorisés par relation plutôt que d'être répétés pour chaque ligne du tableau, ce qui aurait été redondant et source d'incohérences potentielles. A l'issue de cette phase d'harmonisation, nous disposons d'une taxonomie de concepts non hiérarchisés. Ensuite, les concepts ainsi identifiés ont été complétés avec des concepts naturellement associés. Par exemple, puisqu'une relation « dans la partie Nord de » avait été définie, des concepts équivalents pour les directions Sud, Est, Ouest ont été ajoutés ; même chose pour les opposés (*e.g.* « à une altitude supérieure vs inférieure à »). Puis les noms des concepts ont été harmonisés et leurs définitions ont été peaufinées. Enfin, les concepts ont été hiérarchisés en introduisant des niveaux supérieurs. Les concepts de ces niveaux supérieurs sont inspirés de ceux de l'ontologie GUM-3-Space de Batman *et al.* (2010), et complétés pour les concepts ajoutés dont la sémantique ne permettait pas une bonne insertion dans la hiérarchie de GUM-3-SPACE.

4. Choix de modélisation effectués

Comme expliqué en introduction, l'ontologie ORL a pour rôle principal de définir un vocabulaire dédié aux relations de localisation qui sont partagé entre les différents modèles et outils définis dans le cadre du projet CHOUCAS, pour permettre leur interopérabilité. Les concepts de l'ontologie ont donc vocation à être manipulés par différents acteurs, concepteurs ou utilisateurs de ces modèles et outils, qui n'ont pas nécessairement la même expérience des relations de localisation. Notamment, tout ou partie des concepts de l'ontologie ORL ont vocation à être exposés dans l'interface de l'environnement de géovisualisation dédié aux secouristes (Viry, 2022), où ils seront utilisés par un opérateur humain pour choisir la bonne relation de localisation en jeu dans un indice donné. *In fine*, il est prévu qu'un label *ad-hoc* soit associé dans l'interface à chaque concept de l'ontologie qui y sera exposé, pour l'expliquer à l'utilisateur. Ce label pourra être un mot, une expression, un schéma, etc., et sera potentiellement différent du label affecté au concept au sein de l'ontologie. Néanmoins, il est nécessaire que les concepts définis dans l'ontologie ORL y soient nommés et décrits le plus clairement possible. Par ailleurs, il n'est pas prévu *a priori* d'utiliser l'ontologie ORL pour calculer des similarités entre ses concepts, si bien que nous n'avons pas identifié de contraintes particulières quant au nombre de niveaux intermédiaires à définir pour organiser les concepts. Nous avons choisi ces niveaux intermédiaires de manière à assurer la meilleure clarté possible (de notre point de vue). La suite de cette partie décrit les choix de modélisation qui ont été faits compte tenu de ce contexte d'utilisation.

4.1. Choix, identification et nommage des concepts

Seuls certains **types de relations spatiales** utilisées dans les indices ont donné lieu à une relation de localisation (donc à un concept dans ORL). En effet, en analysant les indices issus du corpus d'appels d'alerte, nous avons constaté qu'il existe globalement deux sortes d'indices : (1) des indices qui dénotent un positionnement (spatial) d'une cible par rapport à un site, c'est-à-dire qu'il est possible au vu de l'indice de construire une zone de localisation compatible de la cible (par exemple à partir de l'indice « je vois un lac », il est possible de construire une zone depuis laquelle on peut voir le lac) ; (2) des indices qui dénotent une configuration (spatiale), le plus souvent de la trajectoire de la cible par rapport à un site, par exemple « je suis repartie du sommet par la droite » ou « nous marchons dans le sens sud-nord ». Ce deuxième type d'indices ne permet pas de déduire directement une zone de localisation compatible de la cible. Nous n'avons considéré comme des relations de localisation, devant être intégrées dans l'ontologie ORL, que celles mises en jeu dans les indices dénotant un positionnement. Il s'agit de relations spatiales dites classiquement « statiques » dans la littérature, alors que celles utilisées dans les indices dénotant une configuration sont plutôt des relations spatiales dites « dynamiques », i.e. dénotant un mouvement (Pustejovsky *et al.*, 2017). Une conséquence de ce choix concerne l'utilisation de l'ontologie ORL au sein des outils, mentionnés en introduction, d'enrichissement d'une base de données d'objets repères par analyse de textes décrivant des randonnées. L'ontologie ORL permettra d'annoter, afin d'en déduire de nouveaux objets repères et leur position, des phrases contenant des indices de localisation de type statique, telles que « Il y a une chapelle à droite du chemin à l'endroit où il tourne à droite » (mais pas « Tourner à droite après la chapelle »).

En ce qui concerne **l'identification et le nommage des concepts**, comme annoncé en section 2 nous avons adopté un **point de vue « proche de l'utilisateur »** tel que proposé par Bucher *et al.* (2012), puisque la sémantique des relations est notre préoccupation principale. Par ailleurs, contrairement au modèle proposé dans (Bunel *et al.*, 2019a) qui distinguait un verbe et une relation de localisation utilisés dans un indice de localisation, nous avons considéré qu'une relation de localisation peut intégrer *à la fois* un verbe et une relation de localisation. Ainsi, nous avons défini dans l'ontologie ORL un concept *CibleVoitSite* (qui repose sur le verbe voir), ou une relation *AvoirASaGauche* (qui intègre le verbe avoir et la locution souvent considérée comme une relation spatiale « à gauche »). Toutefois, dans la plupart des cas, le verbe est « être » ou « être situé ». Pour ne pas alourdir, nous avons alors décidé de l'omettre, sauf cas particulier. Ainsi, l'indice « Je suis aux alentours du Mont-Blanc » donne naissance au concept *AuxAlentoursDe* dans l'ontologie ORL. Mais l'indice « Je suis sur le GR 5 » donne naissance au concept *SituéSurItinéraireOuRéseauSupport*, que nous avons nommé « SituéSur... » et pas seulement « Sur... » par souci de clarté.

Enfin, nous avons été confrontés à la **question du bon niveau de détail** dans l'identification des concepts de l'ontologie ORL : qu'est-ce qui fait que deux relations de sémantique proche doivent être considérées comme identiques et modélisées par un seul et même concept, ou au contraire, considérées comme

différentes et modélisées par deux concepts distincts ? Nous avons fait le choix du critère suivant : deux relations de localisation sont différentes dès lors que nous estimons que les méthodes de spatialisation qui leur sont associées doivent être différentes. Ainsi, certainement parce que nous avons particulièrement étudié la notion « sous », (Bunel *et al.*, 2019a), nous avons défini six relations de localisation de type « sous » différentes, parce qu'elles nécessitent des méthodes de spatialisation associées différentes. Leurs sens sont respectivement « sous tout en étant proche de » (« [juste] sous le sommet »), « en-dessous et à l'aplomb de » (« sous tel sommet [plutôt que tel autre] »), « à une altitude inférieure sur le même versant », simplement « à une altitude inférieure à », etc. Une conséquence de ce choix est qu'il peut y avoir des concepts qui paraissent à première vue assez proches, et entre lesquels l'utilisateur pourra avoir du mal à choisir. Si cela s'avère nécessaire après expérimentation, certains de ces concepts pourront alors être groupés dans l'interface de l'environnement de géovisualisation destinée aux secouristes. La même méthode de spatialisation, la plus générique (ici par exemple correspondant simplement à une altitude inférieure) sera alors associée. Cela rendra les choses plus lisibles, même si on perdra en précision de localisation.

4.2. Organisation des concepts

Lors de la phase d'organisation des concepts identifiés, des **concepts de niveau intermédiaire** ont été définis pour regrouper certaines relations de localisation. Parmi ceux-ci, certains correspondent à des relations de localisation qui peuvent être mises en jeu dans un indice (et donc pour lesquelles une méthode de spatialisation doit être définie). C'est le cas par exemple du concept *AuNordDe*, qui a pour sous classes *DansLaPartieNordDe* et *AuNordDeExterne*, mais peut tout de même être utilisé dans un indice.

Certaines relations de localisations telles que *RelationSpatialeDeProximité* ne sont pas assez spécifiées pour constituer la relation de localisation associée à un indice. Par contre, pour faciliter l'usage et l'organisation de l'ontologie, nous avons besoin de les grouper au sein d'un nouvel concept intermédiaire, les **classes abstraites**. Cette notion est similaire au concept homonyme en programmation orienté objet, c'est-à-dire qu'il s'agit de classes ne sont instanciables. Cependant, nous n'avons pas trouvé de concept équivalent en OWL. Pour indiquer le statut de tels concepts dans OOR, nous avons défini une propriété d'annotation booléenne *atAbstractClass*. Deux concepts intermédiaires abstraits ont été définis pour grouper les relations de localisation. Le concept intermédiaire *RelationSpatialeTernaire* permet de distinguer les relations de localisation qui possèdent deux sites, parce qu'elles se traitent différemment des autres (ex : situé entre X et Y). Le concept intermédiaire *RelationSpatialeItineraire*, quant à lui, regroupe les relations qui sont valables dans un espace contraint constitué par un itinéraire ou un réseau support sur lequel se situent à la fois la cible et le site. Ce choix se justifie par le fait que cette configuration une situation fréquente dans notre cas d'application et que cela induit un cadre de référence et des méthodes de spatialisation spécifiques tenant compte de la dimension linéaire (ou réseau) de l'espace support (e.g. être situé après un jalon le

long d'un itinéraire). Nous n'avons pas cherché à avoir une hiérarchie simple (arbre de concepts) : **certaines classes ont deux classes-mères**, mais ce cas est peu fréquent. L'ontologie proposée est assez éloignée d'une structure « en treillis ».

A ce stade, aucune **propriété** de données n'a été définie sur les concepts de l'ontologie. Cela serait nécessaire pour certaines relations paramétrables. Par exemple, pour la relation de localisation à telle distance, le concept *DistanceQuantitativePlanimétrique* doit pouvoir être caractérisé par une valeur de distance en mètres. Cette évolution est prévue pour les quelques concepts concernés. Cela serait également nécessaire pour pouvoir modéliser un modifieur associé à une relation, selon le terme introduit dans (Bunel *et al.*, 2019a) (e.g. « Je suis juste sous le sommet », modifieur = « juste »).

5. Résultats

Chaque relation de localisation est formalisée par un identifiant unique (URI), un nom unique en français (propriété « rdfs :label »), une définition (propriété « rdfs :comment », une relation hiérarchique (« rdfs : subClassOf »), 0 ou 1 exemple issu d'alertes réelles (propriété « vann :example »), une source faisant référence au nom de la bande audio correspondant à l'alerte (« dcterms :source »), 0 ou 1 propriétés « seeAlso » ou « equivalentClass ». L'ontologie ORL contient au final 62 concepts dont 11 sont des classes abstraites seulement destinées à l'organisation des concepts (cf. section 4), et les 51 autres peuvent être instanciées. Elle a une profondeur de 4 niveaux au maximum. Par rapport à l'ontologie GUM-3-Space, 24 concepts sont étiquetés (propriété seeAlso) comme proches d'un concept de GUM-3-Space et 8 comme identiques à un concept de GUM-3-Space (propriété equivalentClass). Ces huit concepts sont ceux liés aux relations cardinales : « Au nord de » (la cible est située globalement au nord du site), « Au nord de externe » (la cible est située au nord du site, et disjointe de celui-ci), et les relations équivalentes pour les trois autres directions cardinales. La faible proportion de concepts équivalents s'explique par l'incompatibilité de la plupart de nos définitions avec celles de GUM-3-Space. Ces définitions sont généralement très similaires, mais non équivalentes, d'où le recours à la propriété *seeAlso*.

Le premier niveau de l'ontologie ORL est constitué de cinq classes. Les trois premières sont similaires aux classes de plus haut niveau de relations spatiales (*SpatialModality*) de l'ontologie GUM-3-Space. La classe abstraite *RelationSpatialeProximité* (similaire à la classe *SpatialDistanceModality* de GUM-3-SPACE) regroupe les relations qui ont à voir avec la notion de proximité spatiale (distance). La classe abstraite *RelationSpatialeFonctionnelle* (similaire à la classe *FunctionalSpatialModality* de GUM-3-SPACE) regroupe les relations qui dénotent que la cible et le site ont un lien non seulement spatial, mais aussi fonctionnel. La classe abstraite *RelationSpatialePositionRelative* (similaire à la classe *RelativeSpatialModality* de GUM-3-SPACE) regroupe les relations décrivant une position relative de la cible par rapport au site en se basant sur autre chose qu'une distance. C'est cette classe qui regroupe le plus de relations de localisation. Les deux classes suivantes ont été ajoutées à ce niveau pour regrouper de relations liées à la

visibilité du site depuis la cible ou inversement. Il s'agit des classes *CibleVoitSite* (le site est visible depuis la cible) et *SiteVoitCible* (la cible est visible depuis le site).

La Figure 5 montre les deux premiers niveaux de la hiérarchie de l'ontologie ORL. Les 8 classes de l'ontologie GUM-3-SPACE qui sont déclarées comme équivalentes à une classe de l'ontologie ORL apparaissent dans la hiérarchie mais ne font pas partie de l'ontologie ORL.

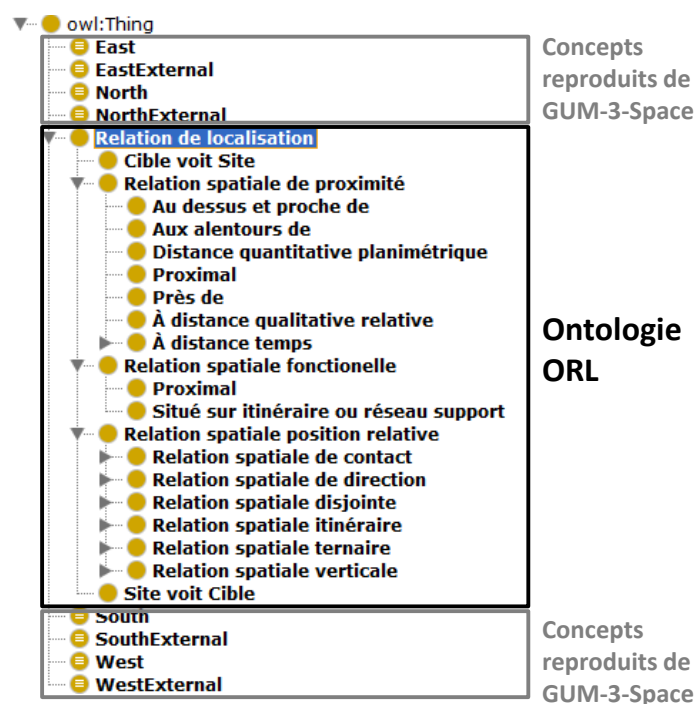


Figure 5. Les deux premiers niveaux de la hiérarchie de l'ontologie ORL.

6. Conclusion et perspectives

Ce papier décrit une ontologie, nommée ORL, des relations de localisation utilisées pour décrire une position par rapport à un objet repère en zone montagneuse. Il s'agit d'une ontologie d'application destinée au secours en montagne, qui a été construite par analyse à partir d'un corpus d'appels téléphoniques aux secours et confrontation avec l'ontologie existante GUM-3-Space. L'ontologie ORL contient 62 concepts, dont 52 sont instanciables, 11 assurent l'organisation hiérarchique, et 24 sont labellisés comme proches de concepts de GUM-3-Space. Pour faciliter l'usage de l'ontologie par les secouristes, un soin particulier a été accordé au nommage des concepts, en choisissant un point

de vue utilisateur. Ainsi, en général les noms des concepts sont une concaténation d'un verbe et d'une relation de localisation, sauf pour le verbe être (« je suis... »).

Chaque concept de l'ontologie ORL correspond à une méthode différente de *spatialisation*, i.e. de construction de la zone vérifiant la relation de localisation considérée avec un objet de référence donné (cf. partie 1). Un prototype mis en œuvre sur une alerte réelle a montré que l'exposition dans une interface utilisateur et la manipulation programmatique des concepts d'ORL permettait d'assurer effectivement l'interopérabilité des méthodes de spatialisation avec l'environnement de géovisualisation mentionné en section 1 (Viry *et al.*, 2023). L'ontologie ORL est publiée dans une version 1, et a vocation à évoluer. Les perspectives à court terme sont d'implémenter la connexion de l'environnement de géovisualisation avec la totalité de l'ontologie ORL, et d'évaluer l'utilisabilité de ses concepts auprès des secouristes ainsi que son usage pour l'annotation des relations de localisation dans le texte, et de la mettre à jour en fonction des besoins identifiés dans ces travaux.

8. Références

- Abadie, N., 2012. Formalisation, acquisition et mise en oeuvre de connaissances pour l'intégration virtuelle de bases de données géographiques. Les spécifications au coeur du processus d'intégration (Thèse de doctorat). Université Paris-Est.
- Bateman, J.A., Hois, J., Ross, R., Tenbrink, T., 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence* 174, 1027–1071.
- Bunel M., Olteanu-Raimond A-M., Duchêne C. 2019a. Objets et relations spatiales composites et prise en compte du vague pour interpréter un référencement spatial indirect. *Revue Internationale de Géomatique, Lavoisier*, vol.29, no.1, pp.81-106.
- Bunel M., Duchêne C., Olteanu-Raimond A-M., Villanova-Oliver M., Bonhoure G., Jouan T. 2019b. An approach for identifying and analysing reference features and spatial relations used in mountain emergency calls . ICC'2019, 15-20 Juillet 2019, Tokyo, Japon.
- Bunel M. 2021. Modélisation et raisonnement spatial flou pour l'aide à la localisation des victimes en montagne. Université Gustave Eiffel, Mémoire de thèse.
- Clementini, E., Laurini, R., 2008. Un cadre conceptuel pour modéliser les relations spatiales. *Revue des Nouvelles Technologies de l'Information RNTI-E-14*, 1–17.
- Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V.-K., Srintzis, M.G., 2005. Knowledge-assisted semantic video object detection. *IEEE Transactions on Circuits and Systems for Video Technology* 15, 1210–1224.
- Duchêne C., Olteanu-Raimond A-M., Bunel M., Brasebin M. 2019. État de l'art sur les relations de Localisation. Livrable L.3.2.1 du projet CHOUCAS (ANR-16-CE23-0018).
- Guarino, N., 1998. *Formal Ontology and Information Systems*. IOS Press, pp. 3–15.
- Hill, L., Zheng, Q., 1999. Indirect Geospatial Referencing through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers, in: *Proceedings of the 62nd Annual Meeting of the American Society for Information Science*. Washington, DC, pp. 57–69.

- Hudelot, C., Atif, J., Bloch, I., 2008a. FSRO : une ontologie de relations spatiales floues pour l'interprétation d'images. *Revue des Nouvelles Technologies de l'Information-les Relations Spatiales : de la Modélisation à la mise en oeuvre* 53–84.
- Jaara, K., Duchêne, C., Ruas, A., 2012. A model for preserving the consistency between topographic and thematic layers throughout data migration. Presented at the 15th International Symposium on Spatial Data Handling, Bonn, Germany.
- Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., 2003. WonderWeb Deliverable D18 Ontology Library. IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web.
- Medad, A., Gaio, M., Moncla, L., Mustière, S., Le Nir, Y., 2020. Comparing supervised learning algorithms for Spatial Nominal Entity recognition. *AGILE: GIScience Series 1*, p. 1-18. <https://doi.org/10.5194/agile-giss-1-15-2020>.
- Miron, A., Gensel, J., Villanova-Oliver, M., Martin, H., 2007. Relations spatiales qualitatives dans les ontologies géographiques avec ONTOAST. Presented at the Spatial Analysis and Geomatics (SAGEO), Clermont-Ferrand.
- Pustejovsky, J., 2017. ISO-Space: Annotating Static and Dynamic Spatial Information, in: Ide, N., Pustejovsky, J. (Eds.), *Handbook of Linguistic Annotation*. Springer Netherlands, Dordrecht, pp. 989–1024. https://doi.org/10.1007/978-94-024-0881-2_37.
- Touya, G., Balley, S., Duchêne, C., Jaara, K., Regnaud, N., Gould, N., 2012. Towards an Ontology of Generalisation Constraints and Spatial Relations, in: 15th ICA Workshop on Generalisation and Multiple Representation - Jointly Organised with EuroSDR Commission on Data Specifications. Istanbul.
- Van Damme M-D., Olteanu-Raimond A-M., Méneroux Y. 2019. Potential of crowdsourced data for integrating landmarks and routes for rescue in mountain areas. *International journal of cartography*, vol 5, n° 2-3, p.195-213.
- Vandeloise, C., 1986. *L'Espace en français : Sémantique des prépositions spatiales*, Seuil. ed, Travaux linguistiques. Paris.
- Vasardani, M., Timpf, S., Winter, S., Tomko, M., 2013. From Descriptions to Depictions: A Conceptual Framework, in: Tenbrink, T., Stell, J., Galton, A., Wood, Z. (Eds.), *Spatial Information Theory*. Presented at the COSIT 2013, Springer International Publishing, Scarborough, UK, pp. 299–319
- Viry, M., 2022. GASPAR —Geoviz Application for Searching And Rescuing People. *Mappemonde* [Online], 134 | 2022, Online since 27 October 2022, connection on 05 January 2023. DOI: <https://doi.org/10.4000/mappemonde.8063>.
- Viry M., Villanova-Oliver M. 2019. L'Ontologie d'Alerte Choucas : de la modélisation des connaissances à un outil support d'un raisonnement géovisuel. *Sageo*, Nov 2019, Clermont-Ferrand, France. ([hal-02279330](https://hal.archives-ouvertes.fr/hal-02279330)).
- Viry M., Bunel M., Villanova M., Olteanu-Raimond A-M., Duchêne C., Davoine P-A. 2023. Prise en compte de l'incertitude dans un processus de localisation de victime en montagne. Une approche méthodologique, logicielle et cognitive. *L'imperfection des données géographiques Tome 2 Applications*. A paraître.

Qaujikairit, un outil d'information et d'alerte précoce sur les événements météorologiques pour les communautés nordiques du Nunavik

Charles Gignac¹, Thierry Badard¹, Stéphane Voyer-Boutet¹,
Jacynthe Pouliot¹ et Richard Fortier²

1. Centre de Recherche en Données et Intelligence Géospaciales (CRDIG), Université Laval, 1055, avenue du Séminaire, Pavillon Louis-Jacques Casault, Ville, de Québec, Québec, Canada, G1V 0A6, charles.gignac.3@ulaval.ca

2. Centre d'études nordiques (CEN), Université Laval, 2405, rue de la Terrasse, Pavillon Abitibi-Price, Ville, de Québec, Québec, Canada, G1V 0A6

RÉSUMÉ. Les changements climatiques affectent de manière tout particulièrement marquée les régions nordiques. C'est le cas du Nunavik (Québec), où l'on dénombre au total quatorze communautés inuites qui occupent le territoire au nord du 49^{ème} parallèle. Les enjeux liés aux événements météorologiques extrêmes y sont de plus en plus présents et les besoins en information criants. Afin d'épauler les décideurs au sein des communautés, un outil d'information sur les conditions météorologiques et d'alerte précoce sur les risques liés est proposé. Basé sur le web, l'application Qaujikairit (alerte) récupère, stocke, traite et présente sous forme synthétique un contenu informatif sur la météo actuelle et à venir; une cartographie des conditions de glace ainsi que des indicateurs de risques liés à des événements météorologiques, comme par exemple les blizzards, froid extrêmes, précipitations majeures pour chacune des communautés nordiques sur le territoire du Nunavik.

ABSTRACT. Climate change is having a particularly strong impact on northern regions. This is the case in Nunavik (Quebec), where there are a total of fourteen Inuit communities occupying the territory north of the 49th parallel. The issues related to extreme weather events are increasingly present there and the need for information is critical. In order to support decision makers in the communities, a weather information and early warning tool is proposed. The web-based Qaujikairit (alert) app retrieves, stores, processes and presents in synthetic form information on current and upcoming weather; ice conditions and risk indicators related to weather events such as blizzards, extreme cold and major precipitations for each of the northern communities on the Nunavik territory.

MOTS-CLÉS : Climat, Météo, Alerte précoce, Vulnérabilité, Risques, Nord, Communautés

KEYWORDS: Climate, Weather, Early warning, Vulnerability, Risks, North, Communities

1. Introduction

Selon le dernier rapport du Groupe d'experts Intergouvernemental sur l'Évolution du Climat (GIEC, 2022), les plus grandes fréquences, intensités et durées d'événements météorologiques comme, pour n'en citer que quelques-uns, les blizzards, les tempêtes, les épisodes de fortes précipitations, la hausse des températures et les surcotes participent à accroître la vulnérabilité des infrastructures et ont des impacts directs et palpables sur la vie des communautés, entre autres celles qui occupent le territoire nord-québécois. D'un point de vue climatique (moyen-long terme), ces événements entraînent une accélération du dégel du pergélisol, une augmentation de la fréquence des glissements de terrain et résultent en des changements dans la dynamique de la végétation et du couvert de glace sur les lacs et en mer (Ford et al. 2006). Les informations sur le temps et le climat sont essentielles pour réduire, comprendre et anticiper les risques liés au réchauffement de la planète, ces dernières offrant un aperçu des tendances climatiques et des phénomènes météorologiques extrêmes récents. C'est considérant cela que, lors de la 3e Conférence des Nations Unies sur la réduction des risques de catastrophes, tenue en mars 2015 à Sendai, au Japon, les nations membres se sont engagées à « *accroître significativement la disponibilité et l'accès à des systèmes d'alerte précoce multirisques, à l'information et aux évaluations sur les risques, pour les populations d'ici 2030* » (France Diplomatie, 2021). C'est d'ailleurs ce contexte qui a motivé la mise sur pied de l'initiative CREWS¹, un système d'alerte précoce aux catastrophes du climat, ciblé principalement sur les pays les moins avancés et les petits États insulaires en développement. La transposition de tels efforts envers les communautés nordiques du Québec, fortement touchées par les changements climatiques, apparaît ainsi comme logique et pertinente.

De tels systèmes d'alertes peuvent aider les organisations à anticiper les effets potentiels sur leurs opérations et à créer des plans de résilience. La donnée et l'information, tant géomatique que climatique ou météorologique se placent donc au centre des actions de mitigation. Or, des contraintes spécifiques aux communautés nordiques s'imposent. Ces communautés doivent jongler avec des défis posés par le manque de données, le fait de reposer sur l'utilisation d'appareils cellulaires pour l'accès au web et sur l'accès à des connexions à faible débit. Ces contraintes posent des défis importants en ce qui a trait à l'accès à l'information.

Le développement de l'application s'inscrit dans un plus large projet financé par le réseau Sentinelle Nord et portant sur l'étude des aléas naturels au Nunavik. C'est considérant ce contexte particulier et afin d'assister les décideurs au sein des communautés, qu'une application sur les conditions météorologiques et les risques liés, nommé Qaujikairit est proposée. Léger, compatible avec les téléphones cellulaires et basé sur le web, cet outil récupère, stocke, traite et présente sous forme synthétique un contenu informatif sur la météo actuelle et à venir, une cartographie des conditions de glace ainsi que sur des indicateurs de risques liés à des événements météorologiques pour chacune des communautés du Nunavik. L'application est

¹ Lien: <https://www.crews-initiative.org/fr>

conçue avec une philosophie modulaire lui permettant d'évoluer au gré des besoins exprimés par les communautés. De nouveaux modules, répondant à des besoins évolutifs déterminés par nos contacts dans les communautés, pourront être aisément annexés dans le futur.

La structure, telle que présentée dans la figure 1, présente rapidement les modules de l'application Qaujikairit. Jusqu'ici, cinq modules ont été mis sur pied dans l'application Qaujikairit, soit les modules sur les données météo, le module sur les données de télédétection, le module ETL (*Extract-Transform-Load*), le module stockage vers la base de données (BD), puis, finalement, le module Utilisateur. Ces divers modules vous seront présentés dans les sections suivantes.



FIGURE 1. Présentation du schéma de l'architecture de l'outil Qaujikairit.

2. Territoire à l'étude : La région du Nunavik

Le Nunavik est un territoire de plus de 500 000 kilomètres carrés, couvrant une grande partie de la région nordique du Québec (Figure 2). Il est délimité par le

territoire du Labrador à l'est, la Baie James et la Baie d'Hudson à l'ouest, le Québec méridional au sud et le Détroit d'Hudson au nord. Le gouvernement local, l'Administration Régionale Kativik (ARK), est un organisme créé en 1978 pour gérer les affaires municipales et les services publics dans les 14 communautés inuites. Ce territoire est peuplé par environ 13 000 Inuits, répartis dans ces communautés (Duhaim et al., 2020). Elles sont situées le long de la côte, principalement le long du Détroit d'Hudson, de la Baie d'Ungava et de la Baie d'Hudson. La région est caractérisée par un climat froid et arctique, avec des températures moyennes de -20°C en hiver et de 5°C en été. Le territoire est riche en ressources naturelles, notamment en poissons, en gibier (phoque, morse, caribou, etc.) et en minéraux (or, cuivre, nickel, etc.). La pêche et la chasse sont des moyens de subsistance importants pour les communautés Inuit, tandis que l'exploitation minière engendre des revenus supplémentaires.



FIGURE 2. Le territoire du Nunavik et ses 14 communautés côtières (en rouge) et 2 sites d'intérêt économique (en bleu). L'échelle cartographique est de 1 : 3 250 000 et le fond de carte est tiré d'Open Street Map.

3. Données utilisées et processus liés dans Qaujikairit

Les choix de données et de technologies ont été réalisés suite à des inventaires exhaustifs et afin de répondre de manière optimale aux objectifs du développement de l'application. Les sections ci-dessous présentent les sources de données en question ainsi que les traitements qui leur sont appliqués amenant au bon fonctionnement de l'application Qaujikairit. L'évolution de ces traitements et de leur déroulement est suivie grâce à l'outil Cronitor (Annexe 1).

Le diagramme de classes présenté ci-dessous (Figure 3) formalise les interactions entre les diverses tables dans la base de données et permet du même coup, d'asseoir la logique menant aux divers calculs effectués en arrière-plan de l'application. C'est sur cette logique et ces relations entre les tables que les requêtes de calcul des indicateurs météorologiques et les requêtes d'affichage sont effectuées.

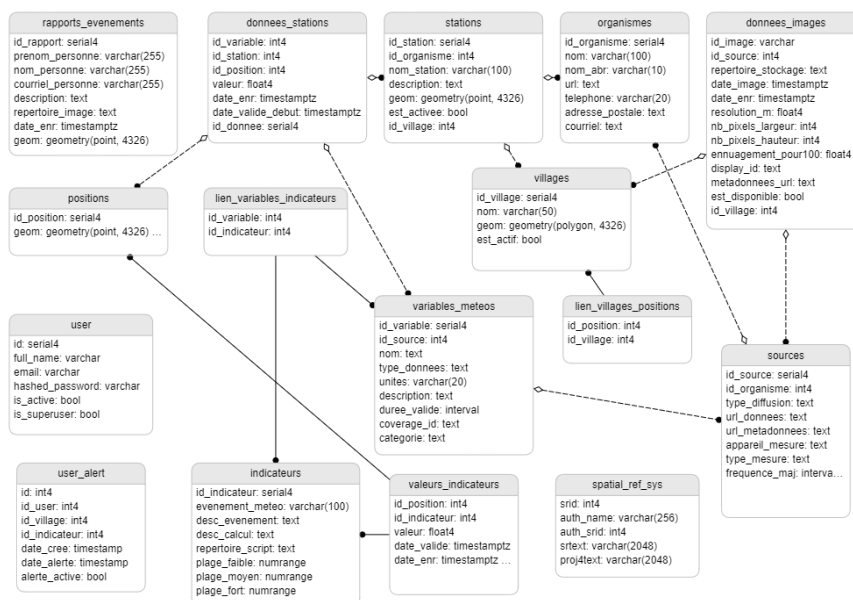


FIGURE 3. Diagramme de classes de la base de données Qaujikairit. Ligne pleine = relation 1:1, Ligne pointillée = relation 1:plusieurs.

3.1. Les données de stations météorologiques d'Environnement et Changement Climatique Canada (ECCC)

Les données d'Environnement et Changement Climatique Canada (ECCC) pour les stations météorologiques sont des données ponctuelles, acquises à une résolution

temporelle horaire, disponibles en temps quasi-réel via le site web d'ECCC et que l'on trouve une station à chacune des 14 communautés. Malgré le fait que cette source de données ne couvre pas en totalité le territoire d'intérêt (ex. Baie Déception et Mine Raglan), il n'en demeure pas moins que ce jeu de données de stations est l'un des jeux de données de stations météorologiques les plus fiables pour la région compte tenu qu'il s'agit de données provenant du gouvernement fédéral. De plus, elles fournissent toute une gamme d'informations pertinentes pour informer sur la situation météorologique actuelle telles les températures, vitesses et direction des vents, précipitations, taux d'humidité et point de rosée.

Chacune des stations est, comme indiqué dans le schéma de processus (Figure 4), liée à un organisme émetteur dans la table - *organisme* -, ainsi qu'à une communauté de la table - *villages* -. Un lien avec la table - *variables_meteos* - est également créé afin d'associer un identifiant unique à chaque catégorie de variable répertoriée (température, vitesse du vent, etc...). Toutes les données acquises sont ensuite stockées avec ces informations auxiliaires dans la table - *donnees_stations* -. Une fois stockées, les données de stations ECCC sont utilisées dans Qaujikairit afin d'informer les utilisateurs et utilisatrices sur les conditions météorologiques actuelles (Figure 5) et les prévisions journalières à court-terme.

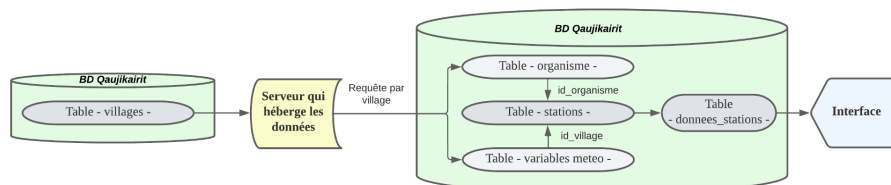


FIGURE 4. Schéma de processus de récupération des données de stations météorologiques d'ECCC

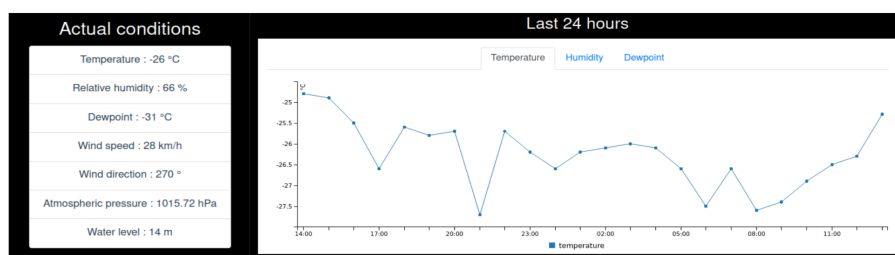


FIGURE 5. Exemple d'informations sur la météo actuelle ainsi que sur la température (graphique) relevées pour les dernières 24h pour une station

3.2. Les données de stations météorologiques du réseau SILA du Centre d'études nordiques (CEN)

Les données du réseau SILA (climat en Inuktitut) proviennent de stations météorologiques installées et opérées par le Centre d'études nordiques (CEN), partenaire du projet, et sont, comme les données de stations ECCC, ponctuelles et acquises à une résolution temporelle horaire. Suivant la même logique que pour les données ECCC, elles sont stockées en BD dans la table - *donnees stations* - en suivant des critères similaires (voir Figure 4). Elles sont disponibles en temps quasi-réel via un serveur FTP du Ministère de l'Environnement, de la Lutte contre les Changements Climatiques, de la Faune et des Parcs du Québec à la suite d'un transfert de données rendu possible par les satellites GOES. Comparativement aux stations d'ECCC, leur couverture spatiale est moins complète mais, contient, en contrepartie, des données sur des sites d'intérêts qui ne sont pas couverts par les données fédérales comme *Deception Bay* et la Mine Raglan (Table 1).

Ces données sont utilisées afin de compléter l'offre d'informations sur la météo actuelle des stations d'ECCC et sont utilisées en priorité lorsque les deux sources s'avèrent disponibles pour une même communauté. Fait important, aucune prévision journalière à court-terme n'est issue de l'information disponible de ces stations.

TABLEAU 1. Stations du réseau SILA du CEN utilisées dans Qaujikairit²

Nom de station	Communauté	Acronyme
Kuujjuarapik	Kuujjuarapik (dans la communauté)	CMK1
Salluit SILA	Salluit (dans la communauté)	CMS1
Baie Déception	Baie Déception (à l'infrastructure)	CMBD
Kangiqsualujjuaq SILA	Kangiqsualujjuaq (dans la communauté)	CQKS
Mine Raglan	Mine Raglan (à la mine)	CQM1
Tasiujaq HT-157	Tasiujaq (dans la communauté)	CQT1

² Source : [Atlas climatiques du Québec - Observations quotidiennes](#)

3.3. Les données de prévision météorologique du Système à haute résolution de prévision déterministe (SHRPD) d'ECCC

Le [Système à haute résolution de prévision déterministe d'ECCC](#) est un système de prévisions météorologiques déterministes à haute résolution qui simule le développement des conditions météorologiques à l'aide de modèles numériques. Les données des modèles numériques, les données radar et les observations météorologiques sont utilisées pour établir les prévisions du SHRPD. Les données des prévisions horaires sur 48h de SHRPD portent sur des variables telles que la température, la pression atmosphérique, la direction et la vitesse du vent, les précipitations, les chutes de neige et l'humidité relative (ECCC, 2023). Ces données sont disponibles via un service web sous forme de cartes de prévisions avec un maillage de 2,5 km et peuvent être visualisées à différentes altitudes (par exemple, le niveau de la mer, 1 km d'altitude, etc.) et à différentes périodes futures (par exemple, 1 heure, 6 heures, 12 heures, etc.). Les données SHRPD constituent le jeu de simulation météorologique court-terme le plus précis spatialement et le plus complet en termes de variables disponibles et ce, tout en couvrant entièrement le territoire à l'étude.

Afin d'éviter d'extraire un trop grand nombre de données, la requête sur les données maillées est effectuée en prenant en compte seulement certaines positions, stockées dans la table - *positions* - sous la forme de géométries GeoJSON en WGS84. Les coordonnées sont passées dans une requête structurée et passée à l'API du *DataMart* d'ECCC. Les coordonnées de chacune de ces positions sont associées à chaque communauté via la table - *lien_villages_positions* - ce qui permet de n'extraire que quelques valeurs avoisinantes à chacune des communautés et de maintenir la taille sur disque de la BD à une taille raisonnable. Un voisinage restreint est pris en compte autour de la communauté (1 km) ce qui englobe généralement un nombre de pixels SHRPD maximal de 4. Advenant une alerte levée considérant les variables simulées dans l'un de ces pixels, l'alerte est considérée levée pour la communauté, intégrant du même coup une composante de proximité dans le processus d'alerte.

Ces valeurs sont stockées dans la table - *donnees_stations* - et par la suite utilisées pour le calcul des indicateurs météorologiques. Les valeurs des différents indicateurs sont calculés en utilisant des variables tirées de la table - *donnees_stations* - et transmises à des scripts Python permettant de calculer leur valeurs respectives, alors retransmises en BD dans la table - *valeurs_indicateurs* - et classées en alerte/pré-alerte/régulier selon les critères listés dans la table - *indicateurs* - (Figure 6). Ces indicateurs, tels que présentés dans la tableau 2, sont inspirés par les définitions des alertes météorologiques d'ECCC (2020). Les indicateurs peuvent, selon leur définition, porter uniquement sur une heure ou être intégrés sur plusieurs heures (Table 2).

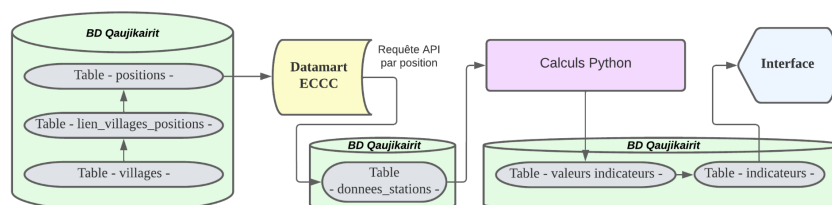


FIGURE 6. Processus de création des indicateurs météorologiques actuellement intégrés dans l'application Qaujikairit

TABLE 2. Indicateurs d'alertes climatiques dans l'application Qaujikairit basés sur les données prévisionnelles tirées des simulations de SHRPD
 [A: Alerte, PA: Pré-Alerte, T: Température, V: Vitesse, P: Précipitations, I: Intégré]

Indicateur	Conditions
Sol gelé	A : $T_{\text{SOL}} < -1^{\circ}\text{C}$
Blizzard (I)	A : $V_{\text{Vent10m}} > 30 \text{ km/h} + P_{\text{Neige}} > 0 \text{ mm/h}$ sur 2 h
Vents forts	PA : $V_{\text{Vent10m}} > 30 \text{ km/h}$ et A : $V_{\text{Vent10m}} > 70 \text{ km/h}$
Neige abondante (I)	A : $P_{\text{Neige}} > 10 \text{ cm}$ sur 12 h
Verglas (I)	A : $P_{\text{Verglas}} > 0 \text{ mm / h}$ sur 2 h
Pluie abondante	A : $P_{\text{Pluie}} > 25 \text{ mm / h}$
Froid extrême	PA : $T_{\text{Éolien}} < -38^{\circ}\text{C}$ et A : $T_{\text{Éolien}} < -48^{\circ}\text{C}$
Brouillard	A : $V_{\text{Vent10m}} < 14 \text{ km/h} + T_{\text{Air2m}} > -3^{\circ}\text{C} + \text{abs}(T_{\text{Air2m}} - T_{\text{Rosée}}) < 2^{\circ}\text{C}$

La température avec le facteur vent ($T_{\text{Éolien}}$) est calculée selon la récente convention canado-américaine de 2001 (équation 1), telle que décrite dans Osczevski et Bluestein (2005).

$$T_{\text{Éolien}} = 13.12 + 0.6215T_a - 11.37v^{0.16} + 0.3965T_a v^{0.16} \quad (1)$$

Où T_a est la température de l'air à 2m en °C et v est la vitesse du vent à 10m en km/h.

3.4. Les données de prévision des marées du Ministère des Pêches et Océans Canada (MPO)

Les données de prévision des marées du MPO comprennent des détails sur les hauteurs de marée, les coefficients de marée, les heures de marée haute et basse et les directions des courants. Ces données proviennent de diverses sources, dont des observations actuelles, des modèles mathématiques et des données historiques. Pour garantir une précision optimale, toutes les prévisions sont mises à jour quotidiennement, ce qui en fait un jeu de données de référence. Ces données peuvent être utilisées pour étudier l'environnement et évaluer les risques côtiers ainsi que pour planifier les opérations maritimes.

Dans l'état actuel de l'application, les données de marées sont traitées comme le sont les stations météorologiques (voir Figure 4). Cela signifie que chaque marégraphe (réel ou simulé) est associé, comme le sont les données de stations ECCC ou SILA, à un organisme émetteur et à une communauté. Les valeurs sont stockées dans la table - *donnees_stations* - et détiennent un identifiant unique du type de variable (ici le niveau d'eau) tiré de la table - *variables_meteos* -. Les données de marées sont utilisées de manière statique pour information aux usagers (voir Figure 5) et seront utilisées dans le développement futur de d'autres indicateurs.

3.5. Les images multispectrales des satellites de la série Landsat du programme conjoint United States Geological Survey/National Aeronautics and Space Administration (USGS/NASA) et les images multispectrales des satellites de la série Sentinel-2 de l'Agence Spatiale Européenne (ESA)

Landsat-8 Operational Land Imager (OLI) est conçu pour fournir des images multispectrales pour une variété d'applications environnementales et de surveillance de la Terre. Les images multispectrales capturées par Landsat-8 OLI couvrent 9 bandes spectrales avec une revisite au 16 jours. La résolution spatiale des images multispectrales Landsat-8 est de 30 mètres pour les bandes du rouge, vert et bleu (RVB) utilisées dans Qaujikairit. Les images capturées par Sentinel-2, couvrent quant à elles 13 bandes spectrales, avec une revisite de 5 jours, et les bandes RVB y ont une résolution de 10m. Ce sont uniquement ces trois bandes qui sont utilisées dans l'application compte tenu que l'algorithme de classification des glaces ne requiert que le RVB et qu'elles permettent de restituer un composé en couleurs naturelles.

Le processus de traitement des images de télédétection est présenté en figure 7. La requête permettant d'identifier les images à télécharger est construite en utilisant la localisation de chacune des 14 communautés et 2 sites d'intérêt économique (ces derniers étant stockés comme villages dans la BD) se trouvant dans la table - *villages* - et stockées localement sur le serveur dans des fichiers GeoJSON. À partir des coordonnées lat/long des communautés, un script Python, faisant appel au

paquet Geemap, est utilisé afin de télécharger l'imagerie dans un rayon de 7.5km autour de la coordonnée fournie via Google Earth Engine (Gorelick et al. 2017). Ce fichier image est transformé en image RVB avec les paquets GDAL, Rasterio et Fiona pour ensuite être passée à un algorithme de cartographie des glaces et être inscrit en BD dans la table - *donnees_images* -. L'image et sa carte correspondante sont alors présentées de manière statique en format JPG aux usagers dans l'interface (voir Annexe 2 pour un exemple de carte de glace).

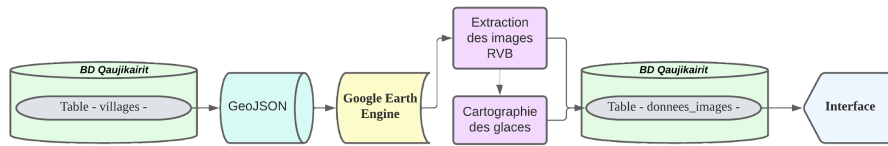


FIGURE 7. Processus de traitement des images multispectrales dans l'application Qaujikairit

4. Discussion, interactions utilisateur et cas d'usage de Qaujikairit

L'application Qaujikkairit est avant tout un exercice d'intégration de données. Les données proviennent d'une multitude de sources, que ce soit les données de stations météorologiques, de modèles climatiques, d'images satellites et sont utilisées pour établir un état de la situation et soulever, le cas échéant, les alertes requises. S'ajoute à ces sources de données externes la capacité des usagers qui, une fois connectés à leurs comptes enregistrés dans la table - *user* -, peuvent transmettre, via un formulaire simple, des informations sur un événement survenu sur le territoire et qui s'enregistre dans la table - *rapports_evenements* -, comme le module Utilisateur l'indique en figure 1. Cette fonctionnalité (voir Annexe 3) permet à un usager de sauvegarder en BD des informations importantes (localisation, descriptif, photographie) sur un événement donné et pertinent en lien avec les risques (glissement de terrain, inondation, tempête, etc.). De plus, les usagers seront en mesure de demander à être informés via courriel dans l'éventualité du déclenchement d'une alerte particulière pour la ou les communautés de leur choix (Figure 8), demande enregistrée dans la table - *user_alert* -. C'est donc non seulement dans la valeur ajoutée par l'intégration des données mais aussi dans l'interaction que les usagers peuvent tirer profit de l'application Qaujikairit.

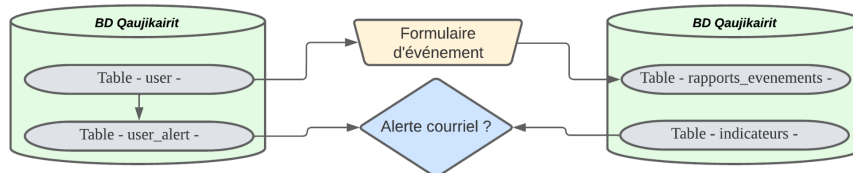


FIGURE 8. Processus de rapport d'événement et d'alerte utilisateur

L'exemple présenté en figure 9 montre l'activation de l'indicateur de blizzard ainsi que de la mise en niveau pré-alerte de l'indicateur de vents forts pour la communauté d'Inukjuak le 31 janvier 2023 en matinée. L'exemple permet de démontrer la pertinence de l'outil et de, tout particulièrement, l'utilisation des données du modèle SHRPD qui donnent une prévision court-moyen terme des conditions météorologiques attendues et du même coup, permettent le calcul d'indicateurs intégrant plusieurs heures d'information, comme c'est le cas avec l'indicateur de blizzard. Dès le lendemain, soit le 1er février 2023, une alerte de froid extrême s'est déclarée pour une large partie du Nunavik, le cas de Quaqtaq est présenté en figure 10.

L'application offre donc une information non seulement actuelle, mais permet d'anticiper les événements à venir grâce aux informations contenues en base de données. Il importe de préciser qu'il ne faut pas considérer les résultats comme des conclusions mais plutôt comme des indications découlant de l'analyse de plusieurs facteurs modélisés. Cette prudence et la parcimonie sont donc de mise considérant que les indicateurs sont établis à partir de données de sources multiples.

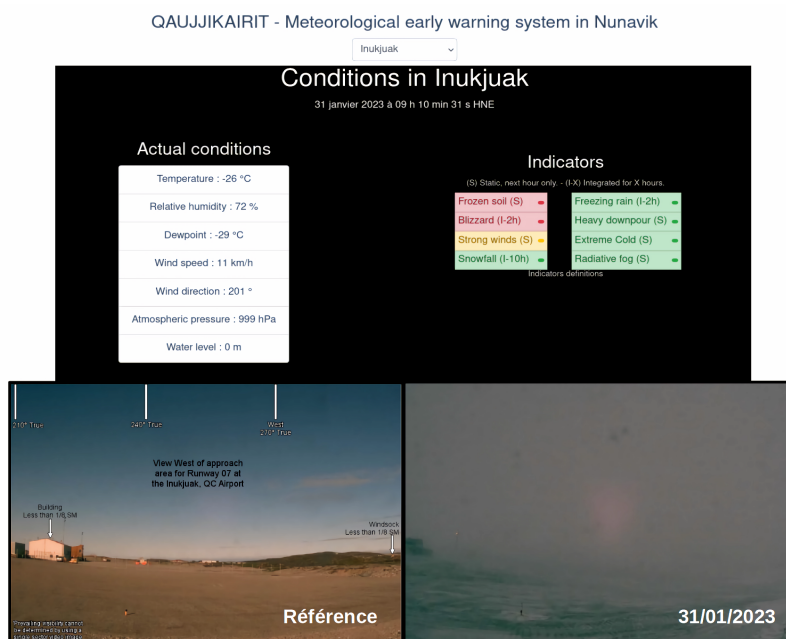


FIGURE 9. Alerte de blizzard et pré-alerte de vents forts pour la communauté d'Inukjuak le mardi 31/01/2023. Photos tirées de NAVCanada³.

³ Webcam de NAVCanada : <https://www.metcam.navcanada.ca/hb/index.jsp>

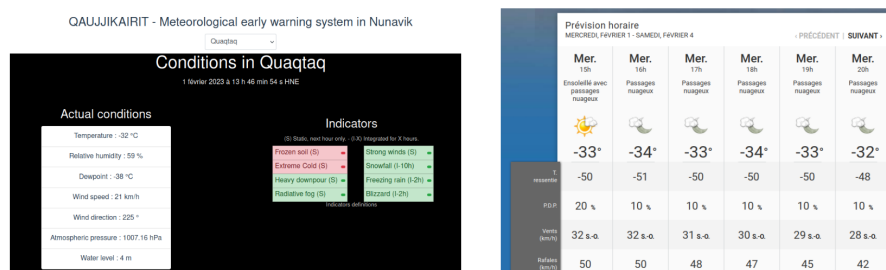


FIGURE 10. Alerte de froid extrême pour la communauté de Quaqtaq le mercredi 1/02/2023. La capture des conditions horaires présentées sur le site de MétéoMédia démontrent que la température, incluant le facteur éolien, atteint le seuil de -50°C .

6. Conclusion

Les développements ayant mené à cette première version de l'application Qaujikairit permettent d'obtenir des résultats prometteurs et laissent présager un excellent potentiel d'application et de développement, tout particulièrement considérant la philosophie modulaire de l'application. L'application constitue en soi un outil simple et convivial afin de se préparer et de se prémunir face aux impacts attendus des conditions météorologiques. Des membres de l'équipe du projet se sont rendus dans les communautés ce qui a permis de présenter l'application et d'obtenir des rétroactions qui seront prises en compte dans l'évolution de l'application. Une rencontre avec un ingénieur de la mine Raglan a également eu lieu et a permis d'identifier des éléments pertinents pour soutenir les travailleurs de la mine dans leurs activités.

Les données utilisées, bien que de sources, d'échelles, de résolutions temporelles et spatiales variées, démontrent une excellente synergie afin d'assurer le fonctionnement de l'application. Certaines lacunes ont été détectées avec l'expérience et le temps, dont la vulnérabilité des stations météorologiques (tant ECCC que SILA) aux conditions difficiles souvent rencontrées au Nunavik, ce qui a pu, au cours de la période d'exploitation de l'application, causer des "blackout" de données pour de courtes périodes. La présence dans la structure de l'application de deux réseaux de stations météorologiques en parallèle assure d'une certaine manière une stabilité de fonctionnement pour l'application Qaujikairit.

Dans un futur proche, il est prévu de développer des modules thématiques portant entre autres sur les glissements de terrain, l'intégration d'informations géospatiales externes par les utilisateurs, l'intégration d'une composante cartographique dynamique pour la navigation et l'interrogation des données, la connexion à des fils RSS d'organisations externes (par exemple : NAV Canada [aviation] ou ECCC [climat]) pouvant compléter l'information disponible et finalement, à raffiner l'algorithme de cartographie des glaces.

Remerciements:

Ce projet a été rendu possible grâce au soutien du programme Sentinelle Nord de l'Université Laval qui est financé en partie par le Fonds d'excellence en recherche Apogée Canada. Les auteurs souhaitent remercier le Centre d'études nordiques, le Ministère de la Lutte contre les Changements Climatiques, de la Faune et des Parcs pour l'accès aux données des stations météorologiques du réseau SILA. Ils remercient également l'équipe d'étudiants de génie géomatique de l'Université Laval qui ont développé le premier prototype de l'application, soit Stéphane Voyer-Boutet, Samuel Hovington, et Florian Philippe.

Bibliographie

- Duhaime G., Caron A. et Everett K. (2020). Le Nunavik en chiffres 2020, Chaire de recherche du Canada sur la condition autochtone comparée, Université Laval, Québec, Français et Inuktitut, 2020, 64 p. <https://bit.ly/3YetPyf>
- Ford J.D., Smit B., Wandel J. (2007) Vulnerability to climate change in the Arctic: A case study from Arctic Bay, Canada, [Global Environmental Change](#), vol. 16, n° 2, p. 145-160.
- Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D. et Moore R. (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone, [Remote Sensing of the Environment](#), vol. 202, p. 18-27
- Groupe d'experts Intergouvernemental sur l'Évolution du Climat.- GIEC (2022) Synthesis Report of the IPCC sixth assessment report (AR6), IPCC. [IPCC](#)
- ECCC (2020). Critères d'alertes météo publiques. <https://bit.ly/3YijPnY>
- ECCC (2023). Données et Produits du Système à haute résolution de prévision déterministe. <https://bit.ly/3YdE85U>
- France Diplomatie (2021). CREWS : alerter précocement les populations des catastrophes climatiques. Gouvernement Français, <https://bit.ly/40ozF0S>
- Osczevski R. et Bluestein M. (2005). The new wind chill equivalent temperature chart, [Bulletin of the American Meteorological Society](#), vol. 86, n° 10, p. 1453-1458.

Annexe 1 : Surveillance des traitements par Cronitor

Le bon fonctionnement des divers modules est supervisé grâce à l'outil Cronitor (Figure 11) qui permet de surveiller les tâches planifiées (ou *cronjob*) sur la machine virtuelle Ubuntu et ce, via le web. Sans devoir être connecté via SSH sur le serveur, il est alors possible de visualiser l'état des diverses tâches planifiées et passées et de demander à recevoir un message d'alerte par courriel dans le cas d'une anomalie.

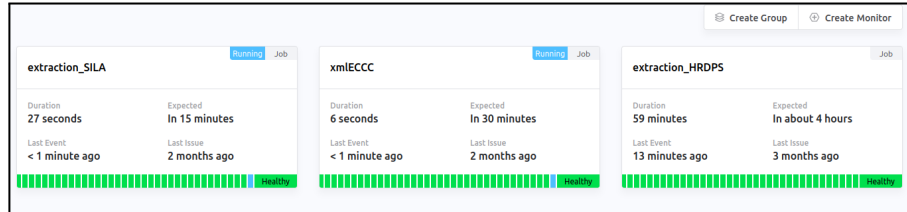


FIGURE 11. Interface de suivi des opérations de Cronitor

Annexe 2 : Cartographie des conditions de glace

Les conditions de glace sont surveillées dans le voisinage des communautés grâce aux données Landsat-8 et Sentinel-2. Un algorithme automatisé permet de restituer une carte des conditions de glace (Figure 12) à partir des images Landsat et Sentinel-2 pendant l'exécution du module -ETL-. Cet algorithme fait appel à l'apprentissage machine, plus précisément à une classification par forêt d'arbres aléatoires, suite à un segmentation de l'image en objets homogènes.

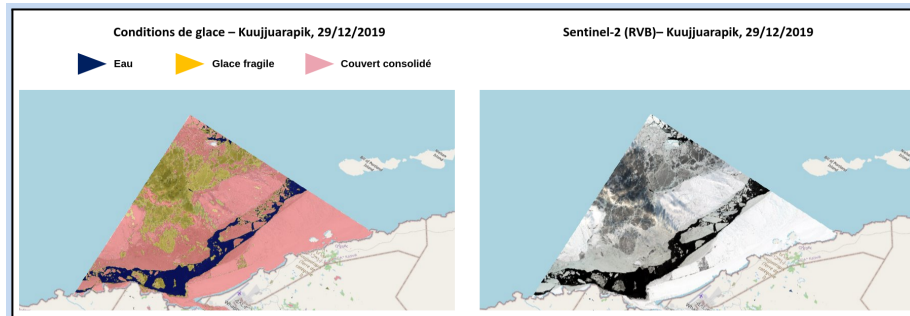


FIGURE 12. Exemple de cartographie des conditions de glace à proximité de la communauté de Kuujjuarapik en décembre 2019

Annexe 3 : Soumission d'événement par l'utilisateur

Afin de construire une archive des événements et de pouvoir peaufiner nos approches de détection d'événements et mieux comprendre les facteurs expliquant leurs occurrences, les utilisateurs de Qaujikairit ont la possibilité de soumettre des rapports d'événements via le formulaire ci-dessous (Figure 13).

Submit an event record

Surname

Name

Email

Content

Longitude (decimal degrees WGS84) <input type="text" value="-77.64"/>	Latitude (decimal degrees WGS84) <input type="text" value="55.28"/>	Image <input type="text" value="Parcourir... northern-quebec-landslide-2021.jpg"/>
---	---	--

FIGURE 13. *Soumission d'un événement de glissement de terrain à Kuujjuarapik en avril 2021*

Developing an Ontology-Based Framework to Select Geospatial Data for Natural Hazards Early Warning Systems (Case study: Nunavik, Québec, Canada)

Amirhossein Vahdat¹, Jacynthe Pouliot¹, Thierry Badard¹,
Richard Fortier²

1. Centre de recherche en données et intelligence géospatiales (CRDIG) and
Département des sciences géomatiques, Université Laval, Québec, Canada

vahdat.amirh@gmail.com

jacynthe.pouliot@scg.ulaval.ca

thierry.badard@scg.ulaval.ca

2. Centre d'études nordiques (CEN) and Département de géologie et de génie
géologique, Université Laval, Québec, Canada

richard.fortier@ggl.ulaval.ca

ABSTRACT. The goal of the Qaujjikairit project ("alert" in Inuktitut) is to establish a key natural hazards early warning system (EWS) for enhancing natural hazard awareness as well as natural hazard mitigation and response strategies. Geospatial and Earth Observation (GEO) data are essential for spatio-temporal analyses in order to accelerate effective early warning and mitigate the damages. Accessing this data quickly and efficiently is an everyday challenge for disaster management experts. The difficulties in accessing in a useful way distributed GEO data all over the Nunavik territory (Québec, Canada) leads this study to design a system to make the most advantage of existing data from GEO portals like NASA in the Nunavik's EWS. This research explores the relevant characteristics of required GEO data in natural hazards EWS. It aims to improve data searching based on aggregating knowledge of natural hazards and geospatial datasets. The main contribution of this research is an ontology-driven conceptual modeling for natural hazards using GEO datasets. This approach facilitates multi-source GEO data discovery and enables the integration of different GEO datasets from multiple heterogeneous sources for early prediction and real-time analysis. It is expected that the system will improve the efficiency of gathering required GEO datasets in whole lifecycle natural hazard EWS by reducing the time of gathering the required information and natural hazard-related data discovery, in addition to streamlining responders' man-machine interaction processes.

KEYWORDS: Natural Hazard Early Warning System, Ontology Modeling, Retrieving Geospatial Data, Integrated Information Systems

1. Introduction

1.1. Context and Problematics

The establishment of early warning systems (EWSs) is an essential part of a comprehensive strategy to risk management of natural hazards (Zschau & Küppers, 2013). Large-scale and quick access to Geospatial and Earth Observation (GEO) data offers opportunities to enhance natural hazard warning timeliness and awareness as well as hazard mitigation and response strategies (Yu *et al.*, 2018). However, the volume and diversity of data related to GEO data, especially real-time observations, are exponentially growing, which presents a substantial challenge for the retrieval, integration, analysis, and presentation of information in EWSs (Gomes *et al.*, 2020). Even knowing where the data is stored and how it can be accessed, there is a need to know about data quality and whether the data is best suited for use in a particular purpose. By utilizing metadata models that employ a standardized paradigm and adhere to establish spatial data quality standards (such as the International Organization for Standardization (ISO) standards), data users are able to assess the suitability of various geospatial datasets for their intended purposes. Fitness for use can be investigated by comparison of information related to different data quality dimensions, including objective attributes such as completeness, accuracy, timeliness, and consistency, as well as subjective attributes including relevancy, credibility, accessibility, data authority (Arruabarrena, 2018). Nevertheless, inconsistent information in metadata records is a significant barrier for users to efficiently and effectively discover and select spatial data (Kalantari *et al.*, 2021). Additionally, the presented information in the metadata is often irrelevant to users' needs, and the user interface presents further difficulties in the discovery and selection process (e.g., irrelevant search results). Therefore, innovative approaches are needed to make the most advantages of GEO data in warning and forecasting systems.

The semantic web has been used to discover GEO data, opening up novel views on the issue of heterogeneity in GEO data and its usage (Li *et al.*, 2017). Ontologies can further provide standards for conceptualizing and referring to domains of interest (Gruber, 1993). Most previous research in disaster management applications using semantics (including ontologies) in GEO data discovery is largely concerned with integrating GEO datasets for specifically defined tasks (Qiu *et al.*, 2017; Phengsuwan *et al.*, 2020). These systems enabled the automated analysis of data from different geospatial sources. However, the process of data discovery is still limited to the existing EWSs in the area or does not cover the geospatial characteristics to improve dataset selection. Qiu *et al.* (2017) proposed an ontology-based approach that automates data discovery for environmental models in flood disaster management systems. They leveraged existing environmental models for the region and defined geospatial datasets' semantic constraints based on the used datasets. This task-driven disaster data link approach successfully discovers appropriate datasets as model input. However, for regions without implemented environmental models, identifying and inventorying available data is the initial step in establishing an EWS. Experts can then decide which model suits a specific region, as data is integral to science, and defining tasks and models is inseparable from available data. Phengsuwan *et al.* (2020)

presented a formal knowledge base pertaining to the domain of landslides. The research emphasized the verification of warning signs utilizing social media and observation data to construct an EWS. While the proposed framework effectively identified potential datasets for validating the occurrence of landslides, its contribution was limited to the monitoring and prediction element of the EWS, with no mention of the other elements. Furthermore, the authors did not specify how they addressed the representation of spatial attributes within their design, such as spatial coverage derived from the metadata. However, some studies have investigated the implementation of existing ontologies, such as Semantic and Sensor Networks (SSN), GeoSPARQL, and DCAT into GEO dataset modeling in order to improve data integration and selection (Wang *et al.*, 2020; Beretta *et al.*, 2021). To the best of our knowledge, no research has previously explored the role of potential GEO information and datasets to develop elements of the natural hazard EWS as a generalized framework.

1.2. Research Project and Objectives

This research is part of a more extensive project called Qaujjikairit project¹, founded by Sentinel North and led by the Centre d'études nordiques (CEN) and Centre de recherche en données et intelligence géospatiales (CRDIG) at Université Laval. The Qaujjikairit project aims to develop a natural hazard EWS and an online advanced foresight tool for extreme meteorological events and natural hazards in the making in Nunavik (Québec), Canada. An integral component of the project is to improve data discovery and the fitness for use of potential GEO data in natural hazards EWSs. Its main goal is to enhance the efficiency of gathering required GEO datasets by increasing the number of relevant GEO datasets, reducing the time of gathering the necessary natural hazard-related data in the discovery process, and streamlining responders' man-machine interaction processes. To do that, the research is divided into two parts. Part one deals with identifying and selecting required GEO information and data, while the second part underlies the recommendation of potential GEO data. We are focusing on the work done in the first part of the research project which aims to facilitate GEO information and data selection. Figure 1 shows the overall research project architecture system, which consists of three primary layers: (i) data preparation layer, (ii) data discovery layer, and (iii) natural hazard EWS applications layer. The data preparation layer collects GEO metadata through data access methods (e.g., RDBMS and APIs like Common Metadata Repository (CMR)). Through text processing, the required dataset characteristics are extracted. Datasets classify into different themes through semantic matching and similarity measurement between data characteristics and GCMD terminologies. The data discovery layer maintains the ontology, which represents knowledge of natural hazards and data sources in a triplestore and makes searching for potential GEO datasets possible. In this layer, a hybrid recommendation system is also designed to improve the dataset's ranking and suggestion. The natural hazard EWS application layer ultimately offers client APIs, enabling users to access the functionalities available by the data discovery layer. This

¹ <https://sentinellenord.ulaval.ca/en/research/qaujikkaut-line-advanced-foresight-tool-extreme-meteorological-events-and-natural-hazards-nunavik>

paper specifically details the knowledge base, and it presents an outlook of the ontology-driven conceptual modeling. The approach is taken to perform three main functions: (1) to construct the links between main concepts in the natural hazard EWS's elements and required GEO information, (2) to design the ontology with required GEO dataset characteristics and to instantiate the ontology with natural hazard-related GEO web datasets, and finally (3) to establish the rules for the reasoning of and to suggest potential GEO information and datasets for use in an integrated natural hazard EWS.

This research introduces an ontology-based framework to select geospatial data and overcome the current GEO search approaches, which rely on manual retrieval based on metadata indices while preparing data for natural hazards EWSs (Qiu et al., 2017). Compared with previous research, the main outcomes of this approach are by emphasizing firstly, the demands of all elements of natural hazards EWS in terms of spatial data and information, secondly, the discoverability of data by incorporating semantics into geospatial metadata and enhancing its usability in terms of consistency and relevant information, and finally the ability to compute data similarity based on thematic and spatial elements.

The rest of this paper is organized as follows. In Section 2, the needs of GEO information and datasets are discussed in the context of natural hazard EWSs. The network of ontologies and the proposed metadata data model are described in Section 3 while their performance is evaluated in Section 4. Section 5 is allocated for discussion, and finally, key findings and future research avenues are provided in Section 6.

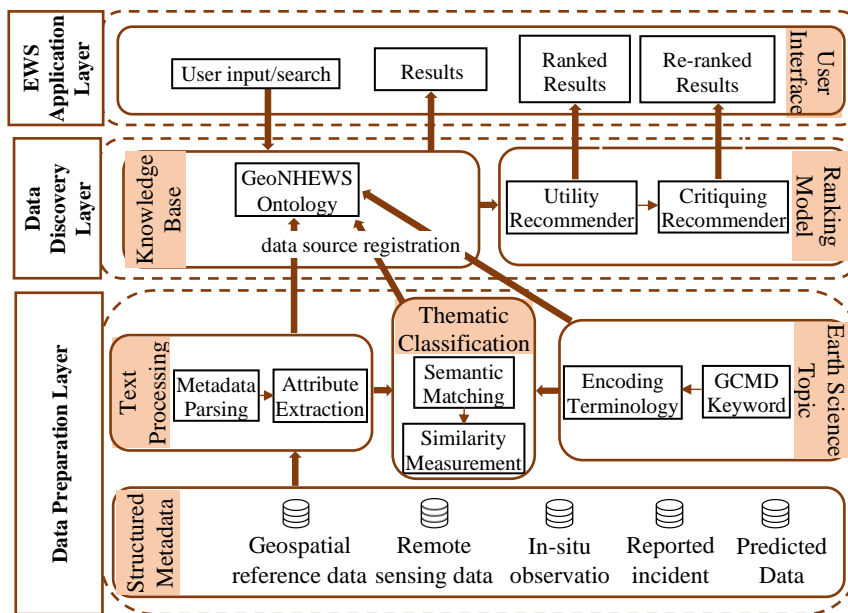


Figure 1. The research project architecture system.

2. Natural Hazard Related-Geospatial Data Identification

In order to understand the characteristics of key natural hazards and identify related GEO data, recognizing the preparatory phase of a natural hazard process is essential. As proposed by Sättele et al. (2012), natural hazard processes can be categorized by the following threefold factors: 1) basic disposition, which is the general and long-term characteristics of an area like topography, land use, and soil texture that increases the potential of occurring natural hazards; 2) variable disposition, which is related to the time-dependent parameters such as rain, temperature, wind; and finally 3) trigger events which are pre-events at the origin of natural hazard processes if the dispositions (basic disposition + variable disposition) are noteworthy. An effective EWS embraces all aspects of emergency management which integrates four main elements: 1) risk knowledge, 2) monitoring and predicting, 3) disseminating information, and 4) response (Wiltshire, 2006). Since the main part of response capability element deals with educating communities to understand the risks and know how to react, this is not covered in this study. The GEO data required to improve natural hazard EWSs can be identified based on the checklist identified during the Third International Conference on Early Warning (Wiltshire, 2006). Based on the natural hazard processes and identified information for each element of EWSs, a conceptual framework of integrated natural hazard EWSs to retrieve static and dynamic geospatial data and information is shown in Figure 2.

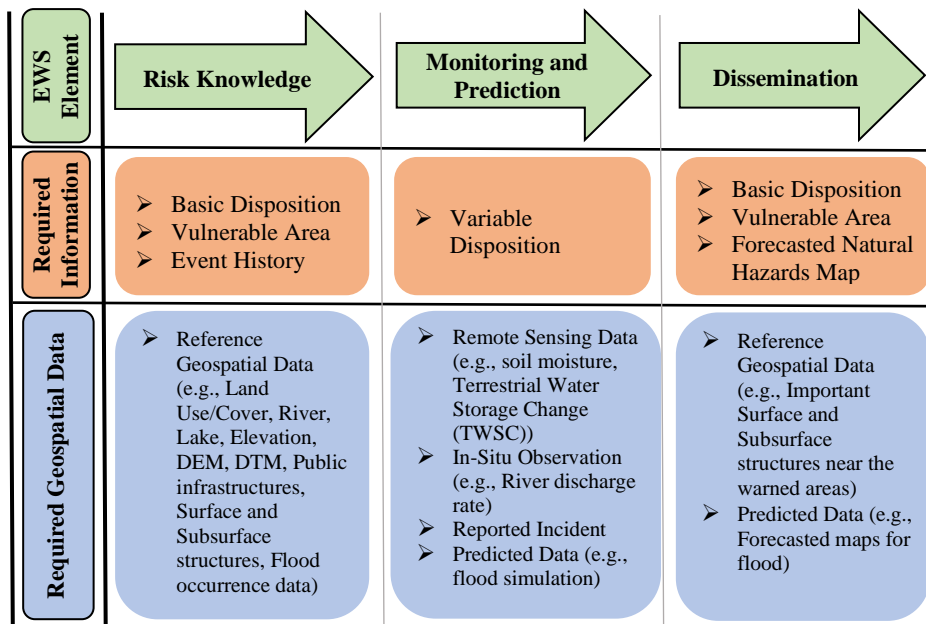


Figure 2. The conceptual framework of integrated natural hazards EWS based on using geospatial data (a showcase for flood-related geospatial data (Acosta-Coll et al., 2018)).

In the following, each EWS's elements requirements used in the conceptual framework will be discussed. In risk knowledge element, a geospatial information system (GIS) is helpful in storing and regularly updating natural hazard risk datasets, as well as for accessing those published into web services. To make maps and data widely available in this element, it is essential to address key questions such as: Are the hazards and vulnerabilities well understood? What are the patterns and trends of hazards in the region? In response to the former question, basic disposition-related geo data such as land use/cover and topography serve to characterize hazards. Additionally, vulnerabilities-related geo data entail information and data regarding surface and subsurface structures (Peng *et al.*, 2022) that are prone to risk. One critical criterion for determining a region's vulnerability is the population at risk, which can vary depending on factors such as season and time of day. Accordingly, it is essential to consider time when assessing the region's overall vulnerability. To answer the latter question, historical data on events (e.g., historical flood data) are necessary to identify regional patterns and trends. This information corresponds with the reference GEO data category. In monitoring and prediction element, continuous receiving and processing data related to the "variable disposition" factor or process of the natural hazard itself in meaningful formats in real-time or near real-time is vital. Also, the knowledge of natural hazard warning signs can be used to identify appropriate data sources for natural hazard precursor verifications. Therefore, reported incidents that people observe for a specific location can be another data source. The mentioned information enables EWS to provide dynamic and timely decision-making in response to natural hazards. The related GEO data for real-time or near real-time information is organized into the categories of remote sensing data, in-situ observation, reported incident, and predicted data. Finally, for the dissemination element, decision-makers and responsible authorities need a usable and simple warning so as to be understood and to feel confident in acting. The GEO data about vulnerable areas and the major factors in the hazard risk (basic disposition) in the warned regions help to understand what and where are most in danger. Also, considering the real-time positioning of people and generally moving objects gives decision-makers a clear understanding of when an event has a significant impact and who is in danger.

3. Natural Hazard Early Warning System Ontology

In order to address the mentioned problematics in improving fitness for use of GEO data in natural hazard EWSs, it is worthwhile to formally delineate each of the previous elements and their relationships to potential geospatial data sources (Wiegand & García, 2007). In this research, an ontology-based approach is proposed to lessen the need for domain knowledge when querying for appropriate GEO data and provide convenient quick access to GEO data to support decision-making. An ontology is a formal and machine-readable description of all the entities of a domain and the relationships existing between these entities (Gruber, 1993). The ontology is implemented in OWL 2 and is built in Protégé as recommended in the Ontology 101 development process (Noy & McGuinness, 2001). This process comprises seven steps which are described in more detail in the following:

1. *Determine the domain and scope of the ontology.* The new ontology called GeoNHEWS (which stands for Geospatial Natural Hazard Early Warning System) is designed and developed to conceptualize a formal representation of potential GEO data to use in natural hazard EWSs. The GeoNHEWS Ontology consists of two main modules: natural hazard EWS and GEO metadata model.

2. *Consider reusing existing ontologies.* As studied in Section 1, no ontologies fully model a natural hazard EWS based on the required GEO information and data. Nevertheless, existing resources were reused when possible, focusing intensely on vocabulary included in modern ontologies and onto-terminologies. The most important ontologies in reusing the terms and relations are elaborated herein:

- *Global Change Master Directory (GCMD) Keyword vocabulary:* The controlled keywords Earth Science, Instrument/Sensors, and Platforms/Sources were chosen to evaluate the fitness for use of the dataset to run a processing chain. This vocabulary transformed into Simple Knowledge Organization System (SKOS) structures and used for discovering and using earth science and environmental data.
- *Data Catalog Vocabulary (DCAT):* The DCAT RDF vocabulary is used in our model to define the characteristics of GEO datasets. It is compliant with structural and administrative metadata formats such as ISO 19115 and helps to facilitate the aggregation and usage of metadata from multiple data catalogs in the web. This allows data publishers to effectively describe their datasets using standard vocabulary for easy access and interoperability.
- *Data Quality Vocabulary:* This ontology was used to define the quality of datasets for more complex descriptions of various aspects, such as spatial resolution.
- *GeoSPARQL, and Time:* They provide more detailed semantic descriptions of space and time.
- *Sensor, Observation, Sample, and Actuator (SOSA):* The ontology is used for bringing observation context like in-situ observations and remote sensing imageries. It defines a set of observations as a collection giving the main properties of the observation context.

3. *Enumerate essential terms in the ontology.* The key terms of the ontology related to the application scope were extracted in the previous phase of the requirements specification (Figure 2). These terms are mainly divided into three parts: (1) phenomena and natural hazard processes, (2) elements of EWSs and the GEO information required for them, and (3) the GEO data characteristics needed to improve their fitness for use in these systems.

4. *Define the classes and the class hierarchy.* A series of ontology classes were selected from the list of essential terms. The core classes of GEO metadata model and natural hazard early warning systems modules are shown in Figures 3 and 4, respectively. A combination of the top-down and bottom-up approaches was used in

developing the class hierarchy. For example, to define Discharge Rate as an Observable Property associated with "Flood", we started from the "Observable Property" class and created the "Ground Water Processes" class. Then, it was related to the intermediate concept of "Discharge".

5. Define the properties of classes. To connect classes and define attributes, object and data properties of classes were used in OWL 2 and defined based on the minimum set of variables required. The class relationships are illustrated in Figures 3 and 4. All reused classes in the GeoNHEWS ontology inherited the related object properties and data properties from their respective source ontologies (e.g., SOSA, DCAT, Time, and GeoSPARQL ontologies). Additionally, further details are provided in Figure 3 regarding the core data properties used for capturing the characteristics of geospatial datasets across different categories.

6. Define the facets of the slots. This step entails defining cardinality and value constraints. Data and object properties of new classes connect individuals with literals. The restrictions defined in reused ontologies were used for the rest of the ontology. For instance, the value of data property `dcterms:title` for `dc:Dataset` class is restricted to be of datatype `rdfs:Literal` and datatype `geo:wktLiteral` for the value of property `geo:asWKT`.

7. Create instances. Individual instances of classes were created in the last step. Utilizing the ontology network presented so far, a dataset containing all the measurements and descriptions described by the GEO datasets was constructed. Also, the GCMD science keyword taxonomy is adopted in `skos:Concept` class as individual instances. There are more than 3,000 keywords in total within taxonomy, making it one of the most comprehensive knowledge bases for earth and space science topics.

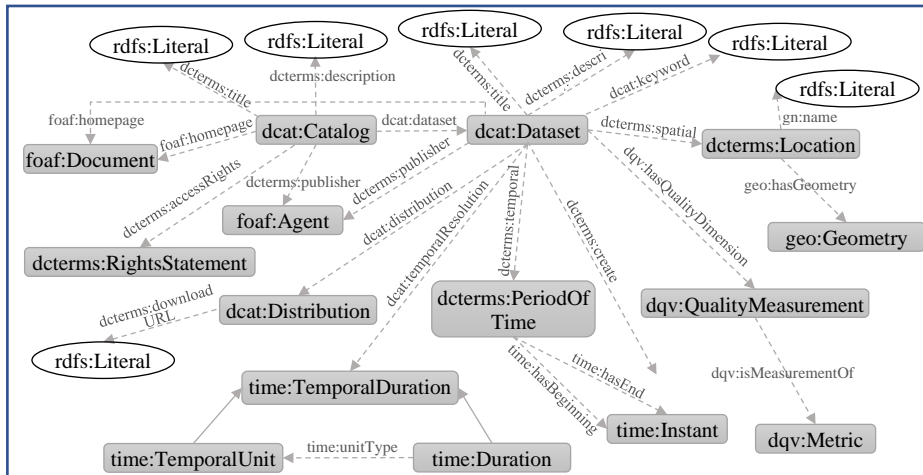


Figure 3. The core object and data properties of the GEO metadata model. The properties connected to ellipsoid shapes are data properties.

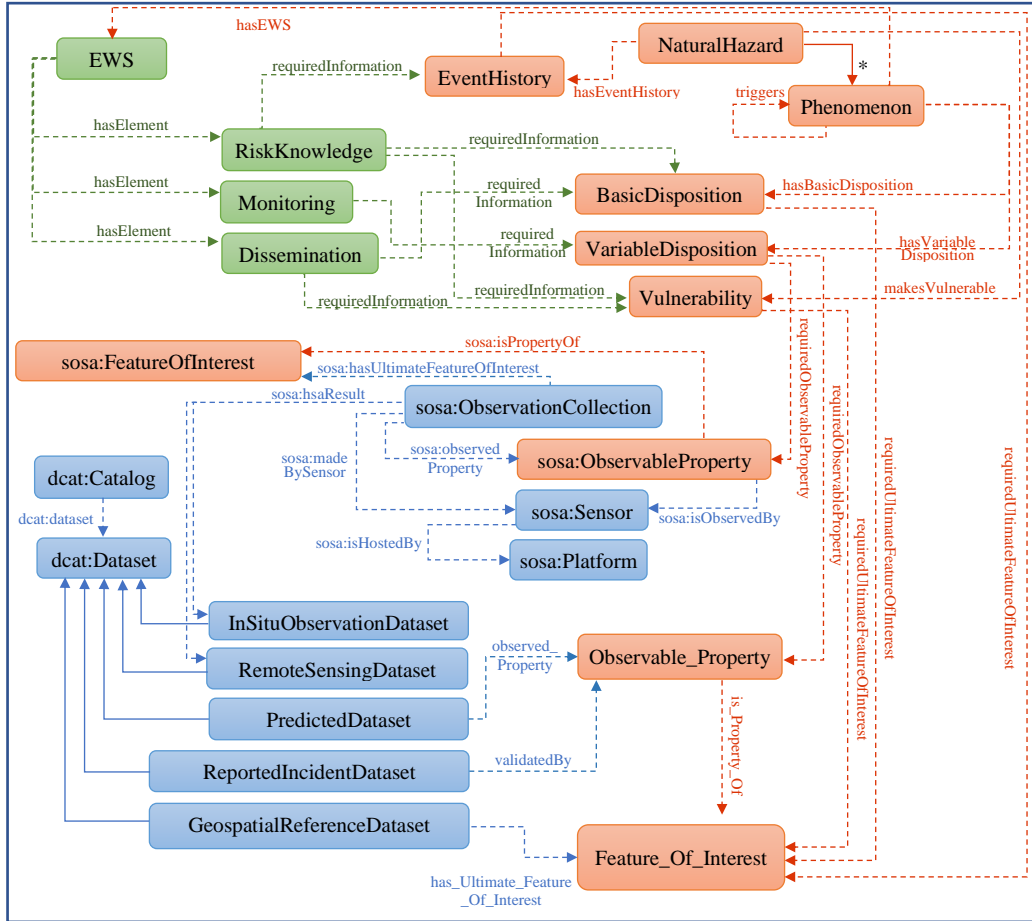


Figure 4. The core classes and object properties in the model of the natural hazard early warning system (EWS). The contents in green are associated with the EWS's elements and their requirements. The contents in red are the natural hazard processes and the associated requirements. The contents in blue refer to the necessary geospatial datasets and their associated interrelationships. * Solid arrows are *rdfs:subClassOf*.

4. GeoNHEWS in Action

To evaluate the usefulness and efficiency of the proposed ontology-based approach, we developed a prototype system, the GeoNHEWS Platform, using Java 8, JavaScript, HTML 5, and CSS. The prototype leverages various web techniques, such as Apache Jena API, GeoNames API, and OpenLayers. Figure 5 illustrates the GeoNHEWS GUI data discovery system, which enables users to input queries and retrieve datasets. Additionally, the page includes interactive controls that guide users to access additional information, such as required information for each EWS's element of various natural hazards.

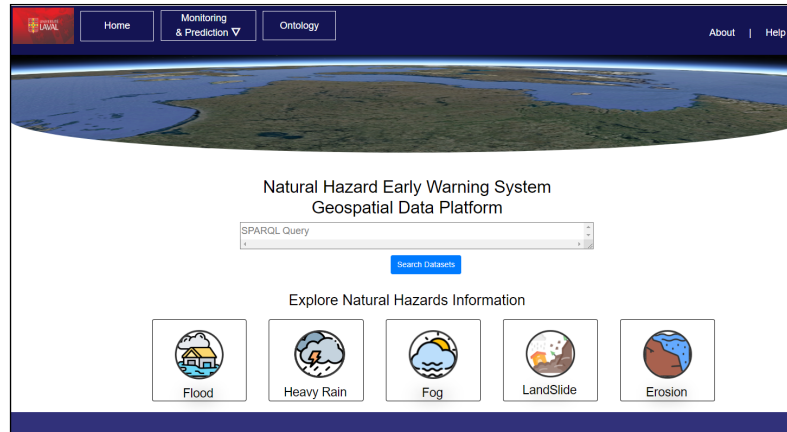




Figure 5. System main page.

The competency questions approach was conducted to verify the coverage of the GeoNHEWS Ontology and its application in natural hazard EWS's elements. Competency questions can be used to assess the validity of an ontology by determining if it captures the intended meaning of a concept or term (Noy & McGuinness, 2001). This evaluation process helps ensure that an ontology is accurate and complete. Also, the HermiT reasoner was employed to validate the ontology consistency. To test the ontology and assess the achievement of the objectives of this study, three competency questions were addressed.

Table 1 shows the first two competency questions which were answered by crafting SPARQL queries using the Protégé software to retrieve information from the knowledge base. The first competency question is designed to confirm that the link between concepts in the EWS and the processes of natural hazard and its associated GEO information is correct. To demonstrate how ontology can support domain experts' understanding and the process of locating appropriate GEO data, the second competency question was designed to do so. Specialists must not only identify the most relevant GEO information and datasets in relation to their needs, but it is beneficial for them to receive recommendations about other GEO information and data to facilitate effective risk-based decision-making. Competency question 3, on the other hand, was evaluated using the developed prototype to retrieve geospatial datasets. It assesses the suitability of potential GEO data for Early Warning Systems (EWSs). Specifically, "What remote sensing datasets are needed to observe the heavy rainfall events between 2019 and 2022 in Nunavik?" Three possible satellite image datasets associated with rainfall events (10.5067/TRMM/TMPA/DAY-E/7, 10.5067/GPM/IMERG/3B-HH/06, and 10.5067/GPM/IMERGDE/DAY/06) from the GES DISC (the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC)) catalog, among other natural hazards-geospatial datasets, were imported to ontology to see if the system could identify them. Figure 6 shows a snapshot of the search results page and the output of competency question 3. The results of these competency questions show that ontology can provide the necessary

GEO information and potential GEO datasets. Moreover, the ontology can suggest potential data sources and accompanying metadata that domain experts can utilize for making timely decisions against risks.

Table 1. Competency questions and their results

Competency Question 1	Competency Question 2										
<p>What are observed properties that can be used in the monitoring element of EWS for flood events?</p>	<p>What may additional observed properties beyond observable properties for rising water levels be needed in the discovery process of datasets for flood event monitoring?</p>										
<p>SPARQL query: </p> <pre>SELECT ?ObservableProperty ?UltimateFeatureOfInterest WHERE { :FloodMonitoring :requiredInformation ?VariableDisposition. ?VariableDisposition :requiredObservableProperty ?ObservableProperty. ?ObservableProperty :isPropertyOf ?UltimateFeatureOfInterest.}</pre> <table border="1"> <thead> <tr> <th>ObservableProperty</th> <th>UltimateFeatureOfInterest</th> </tr> </thead> <tbody> <tr> <td>DISCHARGE RATE</td> <td>GROUND WATER</td> </tr> <tr> <td>FLOW DIRECTION</td> <td>SURFACE WATER</td> </tr> <tr> <td>FLOW VELOCITY</td> <td>GROUND WATER</td> </tr> <tr> <td>FLOW VELOCITY</td> <td>SURFACE WATER</td> </tr> </tbody> </table>	ObservableProperty	UltimateFeatureOfInterest	DISCHARGE RATE	GROUND WATER	FLOW DIRECTION	SURFACE WATER	FLOW VELOCITY	GROUND WATER	FLOW VELOCITY	SURFACE WATER	<p>SPARQL query: </p> <pre>SELECT ?SuggestedObservableProperty WHERE { :Flood :triggeredBy ?Phenomena. ?Phenomena :hasEWS ?EWS. ?EWS :hasElement ?Monitoring. ?Monitoring :requiredInformation ?VariableDisposition. ?VariableDisposition :requiredObservableProperty ?SuggestedObservableProperty.}</pre> <p>SuggestedObservableProperty</p> <ul style="list-style-type: none"> TOTAL SURFACE PRECIPITATION RATE RAIN LIQUID SURFACE PRECIPITATION RATE ACCUMULATIVE CONVECTIVE PRECIPITATION CONVECTIVE PRECIPITATION
ObservableProperty	UltimateFeatureOfInterest										
DISCHARGE RATE	GROUND WATER										
FLOW DIRECTION	SURFACE WATER										
FLOW VELOCITY	GROUND WATER										
FLOW VELOCITY	SURFACE WATER										

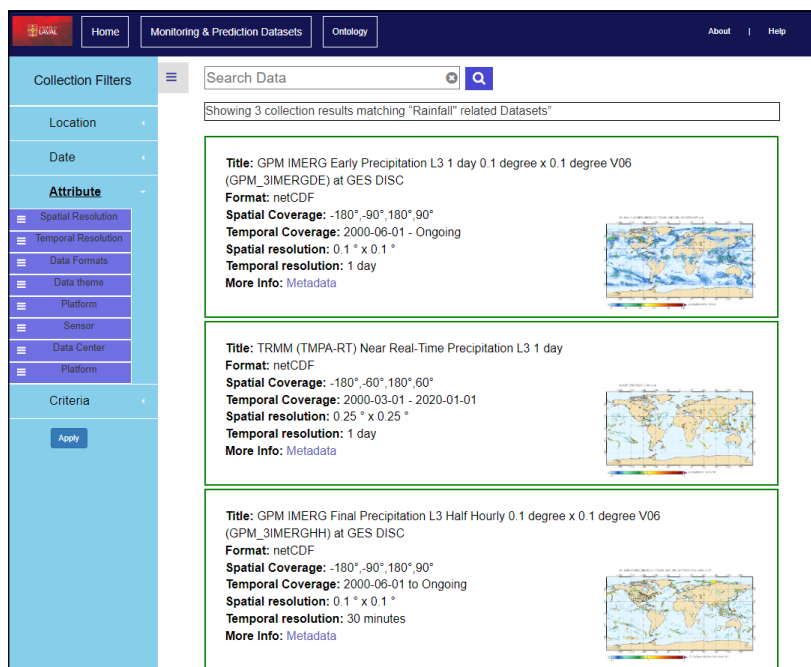


Figure 6. Search results page.

5. GeoNHEWS Advantages and Discussion

The GeoNHEWS ontology employs existing standards to semantically unify the critical information contained in the metadata of geospatial multi-source datasets while defining datasets relationships with phenomena. In addition to continuous access to data and critical information throughout the entire process of natural hazards EWS, an advantageous aspect of the GeoNHEWS ontology is its ability to compute data similarity based on thematic and spatial elements. These thematic and spatial functionalities alongside other semantic attributes enable non-experts in geographic information science to retrieve all relevant geospatial datasets associated with a specific phenomenon to use in natural hazards EWSs.

5.1. Thematic Classification

As previously noted, many users are encountering issues related to the usability of spatial data discovery, including unmatched keywords and search outcomes (Kalantari et al., 2021). In this study, each dataset has been assigned to specific themes, using the GCMD and controlled vocabularies under their observable properties and/or ultimate features of interest. The SOSA model is beneficial for representing collected observations via the *FeatureOfInterest* and *UltimateFeatureOfInterest* concepts, which respectively correspond to the measured entity and the underlying real-world phenomenon. Indeed, for discovery purposes, it is usually the ultimate feature of interest that really matters to the data user. Using a taxonomy that integrates scientific terminology from GCMD ontology and the semantic matching process leads to higher thematic classification accuracy than utilizing no semantics, only pure keyword matching (Li, 2018). Consequently, when a user seeks to retrieve datasets, the ontology is capable of retrieving datasets possessing similar thematic attributes as well as datasets possessing common ancestors. For instance, the observable property of TRMM observation is the ‘Total surface precipitation rate’. Through the SKOS ontology, this concept has a broader concept of ‘Precipitation’, which is the ultimate feature of interest of the observation and is in the scheme of ‘Atmosphere’. Therefore, the relationship between these concepts through GCMD vocabulary is *Atmosphere* > *Precipitation* > *Total surface precipitation rate*. When the ontology requires datasets with the *Total surface precipitation rate* observable property, it first retrieves datasets that have the exact match with this concept. Subsequently, the system computes the shortest path of the aforementioned concept with other themes of datasets, retrieves relevant results and ranks them according to the distance from their common ancestors.

The Q-TREE algorithm (Djidjev et al., 1991) is used as the foundation for computing the shortest distance in the taxonomy. The relative order of subcategories associated with a given node is determined using the SKOS ontology, taking advantage of the fact that the taxonomy is ordered. Given any two nodes v, z , the query algorithm for finding their distance $d(v, z)$ is defined as

$$d(v, z) = d(v, r) + d(z, r) - 2d(LCA(v, z), r) \quad (1)$$

Where $d(v, r) + d(z, r)$ is the distance of the nodes v and z from the root r (in the case that both have the same root, otherwise the nodes are not related to each other), and $d(LCA(v, z), r)$ is the distance from the root to their least common ancestor. In the example above, the semantic distance from "Atmosphere > Precipitation > Total surface precipitation rate" and, for instance, datasets with the "Atmosphere > Precipitation > Liquid precipitation" theme computed using formula one is found to be two.

5.2. Spatial Filter

The metadata model employed in this study incorporates the GeoName database to indicate the semantic aspect of spatial representation through a hierarchical feature. Specifically, when users desire to specify a particular region of interest in their query, this approach enables the retrieval of parent/child relationships between geographic entities. These entities are identified by their unique GeoNames IDs, in conjunction with the Geometry class that specifies each dataset's coordinate systems and geometry. This approach can facilitate the data selection process specifically in three aspects. Firstly, users unfamiliar with the intended area's coordinates can use the area's name for retrieval purposes. Secondly, this process can handle spatial queries even in cases where the coordinate system is missing from the metadata of datasets, but the region description is available. Finally, in a situation when the system retrieves multiple datasets with the same conditions, such as the same percentage of spatial coverage with the user's query, the user may prefer datasets explicitly produced for that region. For instance, when searching for a dataset associated with Nunavik territory, the GeoName hierarchical feature for this region is encoded by "Earth > Canada > Quebec > Nunavik". Many remote sensing observations in NASA provide data covering the whole world, with spatial coverage of "-180, -90; 180, 90" in a latitude and longitude coordinate system. While these data cover a broad range of spatial queries regarding Nunavik, they often do not provide sufficient detail about the region compared to data provided for Canada or, specifically, for Quebec. Therefore, calculating the semantic distance between the spatial query and spatial representation of datasets through their hierarchical features can help rank the datasets. Although the research conducted by Beretta et al. (2021) introduced GeoName features in the ontology design, it does not support the third aspect.

6. Conclusion and Future Work

This paper presented the first phase of a more comprehensive system that aims to identify and rank GEO datasets suitable for supporting EWSs. This phase, which is intended to identify and integrate expert knowledge for data selection, results in two contributions. Firstly, the GeoNHEWS ontology was able to enlarge and enrich the retrieval of suitable datasets thanks to establishing relationships among key concepts in natural hazard EWSs and necessary geospatial information. Secondly, with regard to natural hazard EWS's requirements, the knowledge base was able to suggest GEO information and datasets that are directly or indirectly vital when establishing different parts of EWSs.

Although the topic of ontology modeling is not new, expanding this technique to make it suitable for GEO data discovery is of great importance to advance the natural hazard EWSs. Integrating it with measuring data similarity based on thematic and spatial elements improves the search results and provides much more related GEO datasets. In comparison with other works in the domain of natural hazards EWS, the GeoNHEWS is not limited to a region or existing tasks. Based on our research, in many places, like Nunavik, there are no natural hazards EWS, and to establish one, there should be a road map of what GEO information and datasets are needed to use in each element of an EWS. Also, although some research has been done to respond to gathering required GEO datasets in EWSs, there is no established environment to support investigating how much GEO datasets are fit for the needs of EWS elements. While in our proposal, we identified and modeled the required GEO datasets characteristics and opened up the concept of fitness for use for GEO datasets selection in natural hazards EWSs for future studies.

The initial tests conducted using imported data revealed the ontology's conformity, categorization, and realization. We are currently working on evaluating the GeoNHEWS in applying natural hazards EWS in Nunavik. The future work is to develop a new hybrid recommendation system based on the essential criteria to rank and suggest required geospatial data according to their utilities for different users in the natural hazards EWS domain.

Acknowledgements

The research presented herein was supported by the Sentinel North program of Université Laval (via a research fund to R. Fortier and T. Badard), made possible, in part, thanks to funding from the Canada First Research Excellence Fund. Authors wish to acknowledge the other members of the Working Group 'Qaujjikairit project', especially Charles Gignac and Till Groh.

References

1. Acosta-Coll, M., Ballester-Merelo, F., Martinez-Peiró, M., & De la Hoz-Franco, E. (2018). Real-time early warning system design for pluvial flash floods—A review. *Sensors*, *18*(7), 2255.
2. Arruabarrena, B. (2018). Big data & éthique : la qualité des données en débat. Dans : Laurence Balicco éd., *L'éthique en contexte info-communicationnel numérique : Déontologie, régulation, algorithme, espace public* (pp. 23-38).
3. Beretta, V., Desconnets, J. C., Mougnot, I., Arslan, M., Barde, J., & Chaffard, V. (2021). A user-centric metadata model to foster sharing and reuse of multidisciplinary datasets in environmental and life sciences. *Computers & Geosciences*, *154*, 104807.
4. Djidjev, H. N., Pantziou, G. E., & Zaroliagis, C. D. (1991). Computing shortest paths and distances in planar graphs. In *Automata, Languages and Programming: 18th International Colloquium Madrid, Spain, July 8–12, 1991 Proceedings 18* (pp. 327-338). Springer Berlin Heidelberg.

5. Gomes, V. C., Queiroz, G. R., & Ferreira, K. R. (2020). An overview of platforms for big earth observation data management and analysis. *Remote Sensing*, *12*(8), 1253.
6. Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, *5*(2), 199-220.
7. Kalantari, M., Syahrudin, S., Rajabifard, A., Subagyo, H., & Hubbard, H. (2020). Spatial Metadata Usability Evaluation. *ISPRS International Journal of Geo-Information*, *9*(7), 463.
8. Li, M., Guo, W., Duan, L., & Zhu, X. (2017). A case-based reasoning approach for task-driven spatial-temporally aware geospatial data discovery through geoportals. *International Journal of Digital Earth*, *10*(11), 1146-1165.
9. Li, W. (2018). Lowering the barriers for accessing distributed geospatial big data to advance spatial data science: The PolarHub solution. *Annals of the American Association of Geographers*, *108*(3), 773-793.
10. Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.
11. Peng, Y., Ofterdinger, U. S., Miller, A., Meneely, J., McKinley, J., Laefer, D., & Bertolotto, M. (2022, June). Numerical simulation of the coastal flooding in urban centres with underground spaces. In *39th World Congress of the International Association for Hydro-environment Engineering and Research* (pp. 7103-7109).
12. Phengsuwan, J., Shah, T., James, P., Thakker, D., Barr, S., & Ranjan, R. (2020). Ontology-based discovery of time-series data sources for landslide early warning system. *Computing*, *102*, 745-763.
13. Qiu, L., Du, Z., Zhu, Q., & Fan, Y. (2017). An integrated flood management system based on linking environmental models and disaster-related data. *Environmental modelling & software*, *91*, 111-126.
14. Sättele, M., Bründl, M., & Straub, D. (2012, November). A classification of warning system for natural hazards. In *10th International Probabilistic Workshop*, edited by: Moormann, C., Huber, M., and Proske, D., Stuttgart: Institut für Geotechnik der Universität Stuttgart (pp. 257-270).
15. Wang, C., Zhuo, X., Li, P., Chen, N., Wang, W., & Chen, Z. (2020). An Ontology-Based Framework for Integrating Remote Sensing Imagery, Image Products, and In Situ Observations. *Journal of Sensors*, *2020*, 1-12.
16. Wiegand, N., & García, C. (2007). A task-based ontology approach to automate geospatial data retrieval. *Transactions in GIS*, *11*(3), 355-376.
17. Wiltshire, A. (2006, March). Developing early warning systems: a checklist. In *Proc. 3rd Int. Conf. Early Warning (EWC)* (pp. 27-19).
18. Yu, M., Yang, C., & Li, Y. (2018). Big data in natural disaster management: a review. *Geosciences*, *8*(5), 165.
19. Zschau, J., & Küppers, A. N. (Eds.). (2013). *Early warning systems for natural disaster reduction*. Springer Science & Business Media.

Cartographie lidar des biotopes terrestres, marins et intertidaux de l'atoll de Tetiaroa

Benoît Stoll^{1,2}, Poeiti Tuheiava¹, Moana Badie¹, Lila De Oliveira¹

1. Laboratoire GePaSud, Université de la Polynésie française

BP6570, 98702, Faa'a – Tahiti, Polynésie française

2. benoit.stoll@upf.pf

RÉSUMÉ. Un programme de dératisation de l'atoll de Tetiaroa en Polynésie française a eu lieu en juin 2022. Dans le cadre de ce programme, un état des lieux environnemental exhaustif a été entamé avant l'opération et permettra de comparer l'état de l'environnement en présence et en l'absence de rats. La Tetiaroa Society qui coordonne les actions de recherches sur l'atoll, a émis le besoin d'une carte des biotopes afin d'optimiser les campagnes d'échantillonnages terrain. Cet article propose un protocole simple pour cartographier les principaux biotopes marins, intertidaux et terrestres à partir d'un jeu de données optiques Pléiades et données lidar aéroporté. Les biotopes marins ont été cartographiés selon des critères de profondeur issus de la bathymétrie lidar, les biotopes intertidaux interprétés à partir de données optiques et les biotopes terrestres à partir d'une stratification lidar de la végétation au moyen d'indice lidar classiques. La problématique de la représentation des biotopes terrestres dont certains sont superposés est résolue par la proposition d'une stratification des biotopes sur 4 niveaux.

ABSTRACT. A rat extermination program for the atoll of Tetiaroa in French Polynesia took place in June 2022. As part of this program, an exhaustive environmental inventory was initiated before the operation and will help to compare the state of the environment in the presence and absence of rats. The Tetiaroa Society, which coordinates research activities on the atoll, has expressed the need for a biotope map of the entire atoll in order to optimize the field sampling campaigns. This paper proposes a simple protocol to map the main marine, intertidal and terrestrial biotopes from a Pléiades optical and airborne lidar dataset. Marine biotopes were mapped using to depth criteria from lidar bathymetry, intertidal biotopes interpreted from optical data and terrestrial biotopes from lidar stratification of vegetation using conventional lidar index. The problem of representing terrestrial biotopes, some of which are superimposed, is solved by proposing a stratification of biotopes on 4 levels.

MOTS-CLES : Tetiaroa, atoll corallien, environnement, lidar, biotopes marins, terrestres, et intertidaux

KEYWORDS: Tetiaroa, coral atoll, environment, lidar, marine, terrestrial and intertidal biotopes.

1. Introduction

L'atoll de Tetiaroa est situé à 50km au nord de l'île de Tahiti en Polynésie française, par 17°00'S, 149°34'W dans le Pacifique sud.

Une exploitation intensive de coprah a eu lieu sur les *motu* (îlots coralliens) de Tetiaroa depuis le début du 20^{ème} siècle jusqu'à l'achat de l'atoll par l'acteur Marlon Brando en 1967. Depuis, en l'absence de toute activité anthropique, la nature a repris ses droits sur l'atoll à l'exception du *motu* Onetahi où s'est installé l'hôtel The Brando en 2014.

La Tetiaroa Society effectue la promotion de l'éducation, de la recherche et la préservation de l'environnement sur l'atoll et accueille de nombreux programmes scientifiques dans un but de surveillance, de préservation et de restauration environnementale. L'ensemble des thématiques environnementales est abordé comme la nidification des tortues marines (Stoll et al., 2019), les oiseaux marins, les poissons, le corail, les algues marines, la végétation, les insectes (dont les moustiques (Bossin et al., 2016a, 2016b; Stoll and Bossin, 2019)), les espèces envahissantes ou nuisibles, mais aussi le suivi de l'érosion des plages en lien avec l'augmentation de la fréquence des événements climatiques extrêmes.

Une banque de données multithématique de l'atoll de Tetiaroa (Stoll and Bossin, 2019) a été mise en place pour compiler, conserver et gérer les données géomatiques historiques ainsi que celles produites par les chercheurs. Cette banque de données est accessible en ligne sur le serveur de cartes de la Tetiaroa Society.

En juin 2021, une campagne d'éradication des rats a été effectuée sur la totalité des *motu* dans un but de restauration environnementale (Tetiaroa Society, 2022). Dans ce contexte, la communauté scientifique a été chargée d'effectuer un inventaire de la biodiversité avant et après l'opération et a exprimé le besoin d'une cartographie globale des biotopes marins, intertidaux et terrestres afin d'optimiser l'organisation des campagnes d'échantillonnages scientifiques.

2. Biotopes

Les cartes des biotopes peuvent être obtenues par diverses méthodologies dépendant essentiellement des données disponibles et du niveau de précision attendu. (Corbana et al., 2015) présentent deux pistes pour la réalisation de ce type de carte : (i) De façon indirecte en utilisant des paramètres environnementaux liés à la biodiversité ; (ii) De manière directe par cartographie des assemblages spécifiques ou des communautés écologiques à partir de données aéroportées ou de capteurs satellitaires. Tetiaroa ne présentant quasiment aucun relief (5 mètres maximum), les facteurs abiotiques classiquement utilisés (altitude, pente, ensoleillement, etc.) ne permettent pas d'envisager la première méthode. Il est donc nécessaire de trouver des méthodologies adaptées aux conditions et aux données de télédétection disponibles.

Le choix des biotopes est basé sur une liste de 26 zones naturelles (terrestres, intertidales et marines) proposée par le directeur de la Tetiaroa Society à partir de son

expérience du terrain, de ses échanges avec les rangers et les scientifiques travaillant sur l'atoll. Cette liste a été confrontée à la littérature scientifique et confirmée par des prospections terrain pour aboutir à une liste définitive de 16 biotopes terrestres, 9 biotopes marins et 5 biotopes intertidaux (*Annexe 3*).

3. Données

Les données consistent en une image THR Pléiades du 11/09/2014 (50cm R,V,B,NIR), ainsi que d'un jeu d'ortho-photographies aériennes (5cm R,V,B) et de données lidar à haute densité, toutes financées par le projet IDEA (Davies et al., 2016).

La donnée lidar a été acquise en mai 2017 avec un système lidar vert à retour d'onde complète Riegl LMS-VQ820G (Mallet and Bretar, 2007) monté sur un hélicoptère Eurocopter AS350 à une altitude de vol de 500m. Elle est constituée de 1,42 milliards de points de densité moyenne 25pts/m² et une précision verticale de 4,3 cm. Ce nuage de points couvre aussi bien les milieux terrestres que marins. Pour chaque point lidar, plusieurs informations sont disponibles : sa position (x,y,z), son intensité, son numéro de retour (1^{er}, 2^{ème}, etc.), le nombre de retours total, etc.

Dans un premier temps, cette donnée a été traitée pour obtenir un modèle numérique de terrain et un modèle numérique de surface avec une résolution spatiale de 30cm (Gruen et al., 2017; Serkan Ural et al., 2019). Dans un second temps, la donnée lidar a subi des prétraitements nécessaires à l'étude de la végétation : les valeurs aberrantes, les bâtiments ainsi que d'autres structures anthropiques ont été supprimés. Les points très proches du sol correspondant à des déchets végétaux jonchant le sol ont également été nettoyés. Enfin, les valeurs d'élévation sont normalisées par rapport au sol (modèle numérique de terrain) afin que la hauteur des points corresponde à la hauteur effective de la végétation.

4. Approche multi-strate de végétation

4.1. Problématique des biotopes terrestres

L'expérience acquise lors des études sur la détection de cocotiers dans les atolls de Polynésie française (Desmier et al., 2011; Teina et al., 2009, 2008b, 2008a) a montré que les méthodes classiques de classification d'images satellites optiques sont peu adaptées à notre problématique puisqu'elles ne permettent l'accès qu'à la canopée, excluant de fait les sous-bois des cocoteraies (re)naturalisées.

La complexité structurelle des forêts de Tetiaroa est perceptible lorsque l'on traverse un *motu* du côté océan vers le lagon avec, par exemple, la présence successive (i) d'une frange arbustive, (ii) d'une cocoteraie dense avec présence d'arbres endémiques, (iii) d'une superposition, au-dessus de la lentille d'eau, d'une zone herbacée, d'une zone arbustive avec quelques cocotiers épars (iv) et pour finir une cocoteraie dense avec un sous-bois arboré et présence d'espèces arbustives. Ces différentes combinaisons de types et d'espèces de végétation illustrent bien

l'hétérogénéité mais aussi la superposition des différentes strates de végétation qu'il est difficile de cartographier.

Les échantillonnages terrain (*Figure 3.*) effectués confortent le point de vue des biologistes spécialistes de ce milieu : la structure verticale du couvert végétal des atolls peut être modélisée en 5 strates superposées : (i) strate 1 : herbacées (0-1m), (ii) strate 2 : arbustes (1-3m), (iii) strate 3 : petits arbres (3-10m), (iv) strate 4 : grands arbres (10-20m), (v) strate 5 : arbres supérieurs (>20m).

4.2. Modèle multi-strates et métriques de densités relatives de points

Dans un grand nombre d'études (Jennings et al., 1999; Ma et al., 2017) sur la donnée lidar, la végétation est assimilée à un modèle à deux strates : étage inférieur à 2m, étage supérieur à 2m. Cependant ce modèle n'est pas adapté aux forêts naturelles comportant plusieurs sous-étages (arbustes, petits arbres, grands arbres, etc.) répartis dans différentes strates.

Les indices Overall Relative point Density (*ORD*) et Normalized Relative point Density (*NRD*) sont décrites dans (McCallum, Beaty, et Mitchell 2014) comme des métriques multi-strates. Le principe de ce type de métriques lidar consiste à segmenter le nuage de points lidar en voxels empilés selon le modèle de strates adopté (*Figure 1.*) et à en analyser la répartition des points et de leurs densités.

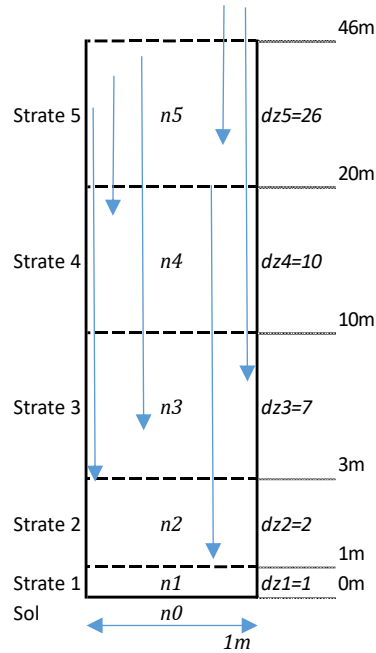
La différence entre ces deux indices réside dans l'utilisation (ou non) des points provenant des strates supérieures en lien avec le phénomène d'occlusion. L'occlusion est le fait que les impulsions lidar ont une forte probabilité d'être bloquées par les étages supérieurs de la végétation, et qu'en conséquence la densité de points lidar sera amoindrie dans les étages inférieurs. Cela est visible au sud du *motu* Rimatuu (*Figure 2*), où la forte densité de végétation dans la strate 4 engendre un trou sur la strate 3.

Nous avons choisi d'utiliser l'indice *ORD* car il prend en compte l'intégralité des points dans le calcul de chaque strate. L'indice *ORD* de la $k^{ième}$ strate est le rapport du nombre de points lidar dans la $k^{ième}$ strate par le nombre total de points dans toute la colonne de végétation et sur le sol.

$$ORD_k = \frac{n_k}{\sum_{i=0}^N n_i} \quad (1)$$

La somme cumulée des indices *ORD* des 5 strates vaut 100%, cela permet une visualisation directe de la distribution verticale de la végétation. De plus, ses valeurs sont cohérentes entre chacune des strates, ce point est crucial pour la cartographie finale : chaque strate est numériquement comparable à une autre. Bien sûr, ces avantages n'évacuent pas le problème de l'occlusion.

La densité de points disponible sur Tetiaroa a permis de calculer l'indice *ORD* sur 5 strates, avec une taille de pixel de 1m. La *Figure 2* en donne un aperçu cartographique.



dz_k est la hauteur de la $k^{ième}$ strate, et n_k est le nombre de points de la $k^{ième}$ strate.
Résolution spatiale : 1m.

Figure 1. Modèle multi-strates de la végétation de Tetiaroa

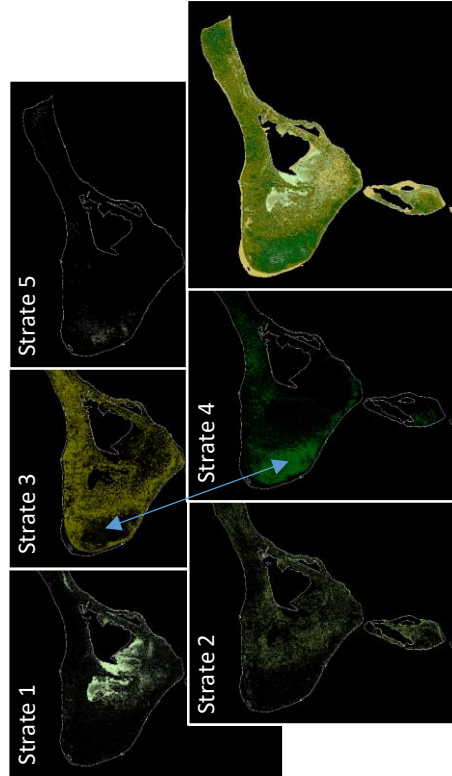


Figure 2. Strates de végétation sur le motu Rimatuu (ORD) et phénomène d'occlusion

4.3. Données terrain

Deux campagnes de vérité terrain ont eu lieu en 2019 et 2021, permettant d'effectuer 1189 échantillonnages de la végétation (Figure 3.), localisés au GPS différentiel (précision entre 2cm et 50cm), et ce, sur tous les *motu* et tous les biotopes.

Le protocole d'échantillonnages consiste à effectuer des transects réguliers traversant les *motu* de part et d'autre. Un point GPS est relevé tous les 50m et les caractéristiques de la végétation dans un rayon de 5m sont relevées pour chacune des 5 strates. Le pourcentage d'abondance-dominance (Delassus, 2015) des espèces principales, secondaires et tertiaires est estimé ainsi que le nom des espèces concernées. Enfin, l'appartenance de l'échantillon à un biotope terrestre est estimée.

5. Vectorisation

La vectorisation des biotopes intertidaux a été effectuée à partir des cartes d'occupation des sols déjà disponibles (Stoll and Longine, 2019). Les vecteurs les plus récents (2014) ont été retravaillés et corrigés par comparaison avec les prises de vues aériennes (ortho-photo à 5cm de 2017) ainsi que les données terrain.



Figure 3. Emplacement des échantillonnages terrain de 2019 et 2021

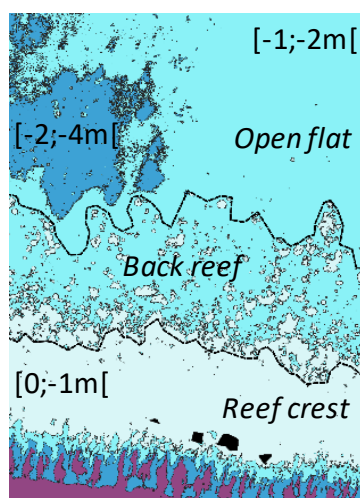


Figure 4. Frontières entre open flat, back reef et reef crest

Les biotopes marins sont essentiellement basés sur la profondeur et la localisation. La bathymétrie a été segmentée selon 6 classes de profondeurs : $[0 ; -1[$, $[-1 ; -2[$, $[-2 ; -4[$, $[-4 ; -10[$, $[-10 ; \text{min}[$. Le raster résultant a été vectorisé, donnant un fichier vecteur de type polygone avec les plages de profondeurs listées dans la table d'attributs.

De plus, un vecteur correspondant à la crête récifale a été extrait de la donnée Pléiades 2014 en repérant le corail vivant à faible profondeur grâce à l'infrarouge. Ce vecteur nous permet de séparer les biotopes situés à l'intérieur du lagon de ceux à l'extérieur.

L'attribution d'une classe à un vecteur s'est faite en analysant les caractéristiques de chaque biotope (Annexe 1.). Certains biotopes sont directement liés à une profondeur comme le *deep lagoon* qui est situé à des profondeurs inférieures à 10m à l'intérieur du lagon. Dans d'autres cas il faut déterminer la frontière séparant deux biotopes contigus (Figure 4.), ainsi le *back reef* est une zone de transition entre le *reef crest* où le corail est prédominant, et l'*open flat* où il n'y a que du sable. La segmentation à 1 mètre de profondeur permet de discriminer les fonds sablonneux des

patates de corail, une opération morphologique d'ouverture (érosion suivi de dilatation) sur le raster segmenté permet de simplifier la géométrie. Déterminer la frontière entre ces trois classes de biotopes est ainsi facilité.

De la même façon, la vectorisation des biotopes terrestres s'appuie sur un seuillage de la valeur de la métrique *ORD* de chaque strate, puis une vectorisation et attribution d'une classe de biotope terrestre. Dans certains cas, un biotope est présent dans une seule strate, et son identification est relativement directe, mais dans d'autres cas, il peut être présent sur plusieurs strates, une identification séparée par strate est opérée puis une fusion des vecteurs correspondants. Dans tous les cas, une vérification de la classe de biotope par rapport aux données de vérité terrain et aux prises de vues aériennes permet de lever toute ambiguïté.

6. Résultats

6.1. Représentation cartographique

La représentation cartographique des biotopes marins et intertidaux est relativement aisée, puisque ceux-ci sont disjoints, sans recouvrements et sans trous. Un jeu de couleur adapté suffit à obtenir une cartographie, claire et lisible.

Concernant les biotopes terrestres, le problème est plus complexe. En effet, la végétation des atolls est caractérisée par une grande diversité d'espèces et une superposition de strates quasi systématique. Un jeu de classes basé sur toutes les combinaisons végétales ne serait pas adapté à une application cartographique.

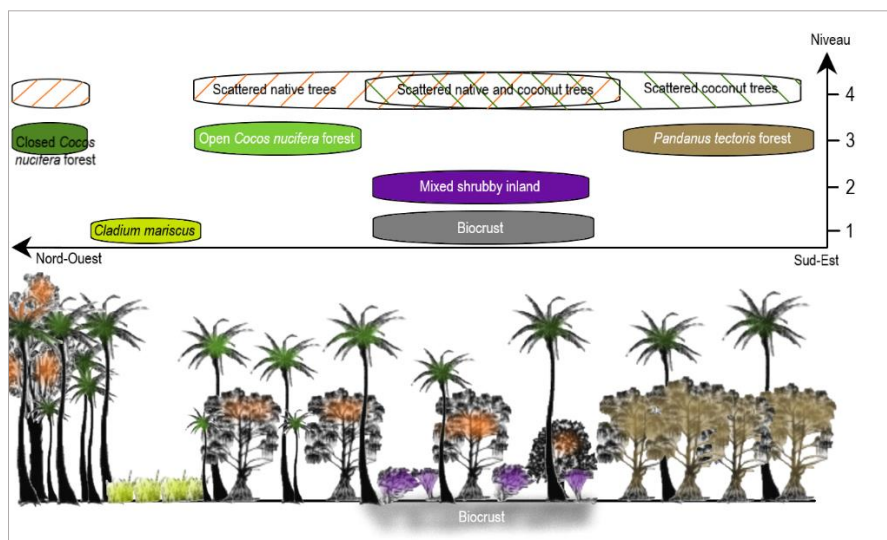


Figure 5. Représentation multi-niveaux des biotopes terrestres

Nous avons donc choisi de représenter les biotopes terrestres selon 4 niveaux (*Figure 5*) : (i) le niveau 1 : le sol et classes associées (ii) le niveau 2 : les arbustes, (iii) le niveau 3 : les couverts arborés principaux et (iv) le niveau 4 : la végétation arborée supérieure clairsemée permettant d'indiquer la présence d'un type d'arbres mais ne constituant pas un biotope forestier particulier. Une transparence appliquée sur les 3 premiers niveaux ainsi qu'un système de hachurages sur le 4^{ème} niveau permet une visualisation claire et réaliste.

6.2. Validation

La validation des biotopes marins n'a pas d'objet puisque ceux-ci sont basés uniquement sur la profondeur, le MNT lidar ayant été validé par ailleurs (Gruen et al., 2017). Par contre, il convient de valider les classes de biotope terrestres en fonction de la donnée relevée sur le terrain, et ceci pour chaque niveau de cartographie.

L'analyse des matrices de confusions et des précisions globales (*Annexe 2.*) pour chaque niveau de biotope terrestre montre que les niveaux 1 et 2 obtiennent une très bonne précision (92.73% et 88.07%), principalement due au faible nombre de classes et leur facilité de détection et d'échantillonnage.

Par contre le niveau 3, avec 5 classes de types de forêts est plus complexe ; la détection comme l'échantillonnage terrain sont moins aisés, ce qui réduit la précision à 59.41%. En particulier, les résultats (*Annexe 2.*) soulignent une confusion importante entre la classe *Cocos nucifera forest* et les classes *Open native forest* et *Closed native forest*, qui partagent un écotone commun ; concrètement il est difficile, sur le terrain, de déterminer la frontière entre une cocoteraie avec présence d'espèces natives d'une forêt native avec présence de cocotiers épars.

Le niveau 4, représentant la végétation éparse, obtient le score le plus faible avec une précision globale de 35.97%. Celle-ci est principalement due à la difficulté d'estimer visuellement les types de végétation et leur taux de recouvrement à plus de 20m de hauteur. L'imprécision relevée proviendrait plus du protocole d'échantillonnage que de la méthode de cartographie par photo-interprétation.

7. Conclusion

Détecter et représenter les biotopes d'un atoll corallien est un problème complexe : il s'agit de cartographier à la fois des biotopes marins, intertidaux et terrestres. Nous proposons pour ce faire d'utiliser les propriétés de la donnée lidar, en association avec de la donnée optique à très haute résolution et des campagnes d'échantillonnages.

La discrimination des biotopes marins est basée sur l'information de profondeur et de localisation dans l'atoll avec un travail vectoriel de nettoyage et de délimitation de la frontière entre certaines classes. Les biotopes intertidaux sont extraits de l'occupation des sols disponible. Les biotopes terrestres, quant à eux, se sont appuyés sur une cartographie lidar *ORD* de la végétation selon un modèle à 5 strates.

Pour obtenir une cartographie claire et concise, nous proposons de représenter les biotopes terrestres sur 4 niveaux (sol, arbustes, couverts forestiers dominants et couverts arborés éparses) associés à un jeu de couleur/transparence/hachurage adapté. Cette cartographie des biotopes, disponible en ligne, est une aide précieuse pour l'évaluation de l'impact de la dératification sur l'environnement de l'atoll de Tetiaroa.

L'objectif d'une cartographie simple des biotopes étant atteint, et conscients que le plein potentiel de ces données n'a pas été atteint, les auteurs envisagent d'explorer plus avant l'utilisation de la donnée lidar conjointement à la donnée optique dans un schéma de classification multi-sources. Cela permettrait de classer les biotopes coralliens plus finement, mais aussi de produire une classification multi-strates de la végétation complexe des atolls, incluant les sous-bois.

Remerciements : à l'ETH de Zurich pour la donnée. Frank Murphy et la Tetiaroa Society pour le soutien logistique et Jean-Yves Meyer pour son aide sur la végétation.

Bibliographie

- Bossin, H., Petit, H., Marie, J., Stoll, B., Cheong Sang, M., 2016a. Releases of Wolbachia mosquito males to sterilize and suppress a population of *Aedes Polynesiensis* on Tetiaroa Atoll, French Polynesia. Presented at the International Congress for Tropical Medicine and Malaria.
- Bossin, H., Petit, H., Marie, J., Stoll, B., Cheong Sang, M., 2016b. Suppression of an isolated population of the mosquito vector *Aedes polynesiensis* on the atoll of Tetiaroa, French Polynesia, by sustained release of Wolbachia-incompatible male mosquitoes.
- Corbanea, C., Langb, S., Pipkins, K., Alleaume, S., Deshayes, M., Garcia Millane, V.E., Strasser, T., Borref, J.V., Toon, J., Michael, F., 2015. Remote sensing for mapping natural habitats and their conservation status – New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation*.
- Davies, N., Field, D., Gavaghan, D., Holbrook, S.J., Planes, S., Troyer, M., Bonsall, M., Claudet, J., Roderick, G., Schmitt, R.J., Amaral Zettler, L., Berteaux, V., Bossin, H., Cabasse, C., Collin, A., Deck, J., Dell, T., Dunne, J., Gates, R., Harfoot, M., Hench, J.L., Hopuare, M., Kirch, P., Kotoulas, G., Kosenkov, A., Leichter, J.J., Lenihan, H., Magoulas, A., Martinez, N., Meyer, C., Stoll, B., Swalla, B., Tartakovsky, D.M., Teavai Murphy, H., Turyshev, S., Valdvinos, F., Williams, R., Wood, S., IDEA Consortium, 2016. Simulating social-ecological systems: the Island Digital Ecosystem Avatars (IDEA) consortium. *Gigascience - Biomed Central* 5. <https://doi.org/10.1186/s13742-016-0118-5>
- Delassus, L., 2015. Guide de terrain pour la réalisation des relevés phytosociologiques. p. 52.
- Desmier, E., Flouvat, F., Stoll, B., Selmaoui-Folcher, N., 2011. Coconut Fields Classification Using Data Mining on a Large Database of High-Resolution Ikonos Images. Presented at the Sixth International Conference on Digital Information Management, SPIE.
- Gruen, A., Tao Guo, Serkan Ural, Matthias Troyer, Sultan Kocaman, 2017. DSM/DTM-related investigations of the Moorea Avatar project. Presented at the Asian Conference on Remote Sensing, New Delhi, India.
- Jennings, S.B., Brown, N.D., Sheil, D., 1999. Assessing forest canopies and understory illumination: canopy closure, canopy cover and other measures. *Forestry: An International Journal of Forest Research* 59–73.

- Ma, Q., Su, Y.J., Guo, Q.H., 2017. Comparison of canopy cover estimations from airborne lidar, aerial imagery, and satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4225–4236.
- Mallet, C., Bretar, F., 2007. Le Lidar Topographique à Retour d'Onde Complète : État de l'art. *Traitement du signal* 24, 441–465.
- Serkan Ural, Armin Gruen, Sultan Kocaman, 2019. Point clouds over Tetiaroa - 3D modeling of a tropical island by topo-bathymetric lidar. Presented at the ACRS (40th asian conference on remote sensing).
- Stoll, B., Bossin, H., 2019. Observatoire de données de Tetiaroa : Contribution à la modélisation de l'environnement des moustiques. Presented at the SAGEO, Clermont-Ferrand.
- Stoll, B., Hart, M., Gaspar, C., 2019. Contribution de la géomatique à l'observation des nids de tortues marines de Tetiaroa. Presented at the SAGEO, Clermont-Ferrand.
- Stoll, B., Longine, M., 2019. Géomorphologie à Tetiaroa : Prémices d'un observatoire de données scientifiques et environnementales. Presented at the SAGEO, Clermont-Ferrand.
- Teina, R., Béréziat, D., Stoll, B., 2008a. A Spatial Poisson Process to Classify Coconut Fields on Ikonos Pansharpened Images, in: *Proceeding of SPIE Asia-Pacific Remote Sensing (Ed.)*. Presented at the SPIE Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications II, p. (10 pages).
- Teina, R., Béréziat, D., Stoll, B., Chabrier, S., 2009. A Comparative Study of Several Supervised Classifiers for Coconut Palm Tree Fields Type Mapping on 80cm RGB Pansharpened Ikonos Images. Presented at the Image Processing Machine Vision Applications II.
- Teina, R., Béréziat, D., Stoll, B., Chabrier, S., 2008b. Toward a Global Tuamotu Archipelago Coconut Trees Sensing using High Resolution Optical Data. Presented at the IEEE International Geoscience and Remote Sensing Symposium, IGARSS'08, IEEE, pp. 797–800.
- Tetiaroa Society, 2022. Restauration écologique de Tetiaroa. URL <https://www.tetiaroasociety.org/fr/programmes/conservation/restoration-ecologique-de-tetiaroa>

Annexe 1. Critères d'identification des biotopes terrestres et marins

Biotopes terrestres	Niv.	Strate 1	Strate 2	Strate 3	Strate 4	Strate 5
Scattered <i>Cocos nucifera</i>	4	Photo-interprétation + terrain				
Scattered native tree	4	Photo-interprétation + terrain				
Scattered introduced tree	4	Photo-interprétation + terrain				
Scattered <i>Pandanus tectoris</i>	4	Photo-interprétation + terrain				
<i>Cocos nucifera</i> forest	3			x	x	x
Closed native forest	3			x	x	x
Open native forest	3			x	x	
<i>Calophyllum inophyllum</i> forest	3			x	x	
<i>Casuarina equisetifolia</i> forest	3			x	x	x
<i>Pemphis acidula</i> / <i>Suriana maritima</i>	2	x	x			
<i>Scaevola taccada</i> / <i>Timonius uniflorus</i>	2	x	x			
Stretch of <i>Cladium mariscus</i>	1	x				
Bio crust	1	x				
Anthropized	1	Photo-interprétation + terrain				

Fresh water	1	Photo-interprétation
Brackish water	1	Photo-interprétation

Biotopes marins	Profondeurs (m)					Position
	[0 ; -1[[-1 ; -2[[-2 ; -4[[-4 ; -10[[-10 ; min[
Pinnacle/ridge top	x	x				Intérieur
Mid lagoon			x	x		Intérieur
Deep lagoon					x	Intérieur
Open flats	x	x				Intérieur
Shifting sand flats	x					Intérieur
Hoa	x					Intérieur
Back reef	x	x				Intérieur
Reef crest	x					Intérieur
Reef crest	x	x	x			Extérieur
Fore reef				x	x	Extérieur

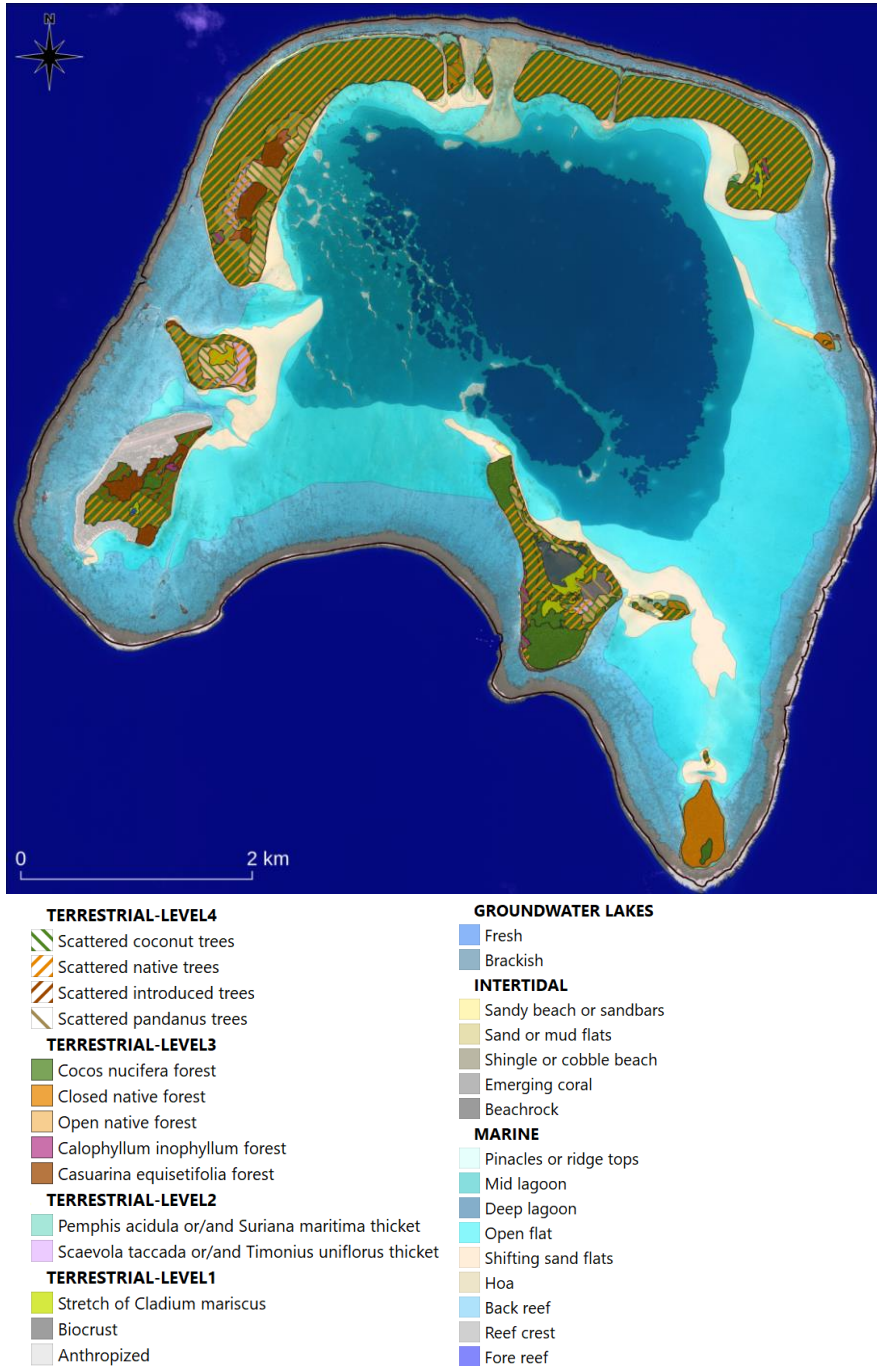
Annexe 2. Matrices de confusions et précision globale des biotopes terrestres.

Terrain niveau 1 Précision globale 92,73%			Terrain niveau 2 Précision globale 98,17%			
Cartographie		Cladium mariscus	Biocrust		Pemphis acidula / Suriana maritima	Scaevola taccada / Timonius
	Cladium mariscus	27	1	Pemphis acidula / Suriana maritima	26	2
	Biocrust	3	24	Scaevola taccada / Timonius	0	81

		Terrain niveau 3 Précision globale 59,41%				
Cartographie		Calophyllum inophyllum	Casuarina equisetifolia	Cocos nucifera	Closed native	Open native
	Calophyllum inophyllum	15	0	1	0	0
	Casuarina equisetifolia	0	11	17	0	6
	Cocos nucifera	5	6	230	70	132
	Closed native	0	0	0	52	11
	Open native	0	0	7	6	74

		Terrain niveau 4 Précision globale 35,97%			
Cartographie		Scattered coconut	Scattered native	Scattered pandanus	Scattered introduced
	Scattered coconut	53	4	8	1
	Scattered native	179	86	28	7
	Scattered pandanus	0	0	0	0
	Scattered introduced	10	14	0	2

Annexe 3. Carte des biotopes terrestres, marins et intertidaux de Tetiaroa



Cartographie de la diversité spécifique forestière des milieux tempérés à partir d'imagerie hyperspectrale

Marc Lang¹, Max Ferriere², Florian de Boissieu², Xavier Briottet³, Sophie Fabre³, David Sheeren¹, Jean-Baptiste Féret²

1. DYNAFOR, Université de Toulouse, INRAe
24 chemin de Borde-Rouge - Auzeville, 31326 Castanet-Tolosane Cedex, France
marc.lang@toulouse-inp.fr

2. TETIS, Université de Montpellier, INRAe, CNRS, CIRAD
500 rue avenue Jean-François Breton, 34090 Montpellier

3. ONERA, Université de Toulouse, 2 avenue Edouard Belin, Toulouse, Cedex, France

RÉSUMÉ. Le développement de méthodes de suivi de la biodiversité est crucial pour endiguer l'érosion de la biodiversité en cours. Nous testons dans cette étude l'applicabilité d'une méthode d'estimation de la biodiversité non supervisée, basée sur l'hétérogénéité spatiale de l'information d'imagerie hyperspectrale. Des indices de diversité spectrale taxonomique α et β et de diversité spectrale fonctionnelle sont calculés et sont comparés à des indices de diversité issus d'inventaires terrain. Les indices de diversité spectrale taxonomique montrent globalement des corrélations élevées avec les indices mesurés sur le terrain pour la diversité α (ρ variant de 0.76 à 0.82) tandis que les corrélations pour les indices de diversité fonctionnelle sont plus variables (ρ variant de 0.45 à 0.63). Les cartes de diversité α montrent que la méthode permet bien de distinguer les peuplements en mélange des peuplements monospécifiques, tout en apportant une information à une échelle plus fine que la BD Forêt V2 de l'IGN. La méthode testée semble donc adaptée au suivi de la diversité taxonomique en milieu tempéré sans que toutefois le lien entre les informations spectrales et taxonomiques soit explicité.

ABSTRACT. The development of biodiversity monitoring methods is crucial to stem the ongoing erosion of biodiversity. In this study, we test the applicability of an unsupervised biodiversity estimation method, based on the spatial heterogeneity of hyperspectral imagery information. α and β taxonomic spectral diversity indices and functional spectral diversity indices are computed and compared to diversity indices from field surveys. Taxonomic spectral diversity indices show overall high correlations with field-measured indices for α diversity (ρ ranging from 0.76 to 0.82) whereas correlations for functional diversity indices are more variable (ρ ranging from 0.45 to 0.63). The α diversity maps successfully allow the discrimination of mixed stands from monospecific stands, while providing information at a finer scale than IGN BD Forêt V2. Hence, the tested method seems to be adapted to the monitoring of taxonomic diversity in temperate environments, without however explaining the link between spectral information and taxonomic diversity.

MOTS-CLÉS : Hyperspectral, Télédétection, classes spectrales, biodiversité, forêt tempérée.

1. Introduction

L'érosion accélérée de la biodiversité est l'une des principales sources de risques environnementaux au même titre que le changement global (Rockström *et al.*, 2009). Il est donc important de mettre en place des outils et systèmes permettant de fournir des informations sur cette biodiversité et son évolution, pour aider à son suivi, mieux comprendre les facteurs impactant cette biodiversité, et proposer des solutions permettant de réduire cette érosion (Niemelä, 2000). La biodiversité est un concept complexe, multidimensionnel, qui nécessite de prendre en compte et intégrer plusieurs niveaux d'information : du fonctionnement des individus à l'étendue et la distribution spatiale des écosystèmes, en passant par la diversité d'une population, et la composition floristique (Pereira *et al.*, 2013).

La télédétection est un outil particulièrement adapté pour le suivi répété et à large échelle de plusieurs composantes de la biodiversité (Vihervaara *et al.*, 2017). Différentes approches de suivi de la biodiversité par télédétection ont en effet été développées ces dernières années (Wang et Gamon, 2019). L'analyse de données d'imagerie hyperspectrales aéroportées a notamment démontré un fort potentiel pour accéder à des informations taxonomiques des milieux forestiers, par l'identification des espèces d'arbres à l'aide de méthodes de classification supervisée (Féret et Asner, 2013; Fassnacht *et al.*, 2016; Laybros *et al.*, 2019), mais aussi par des méthodes non supervisées s'appuyant sur l'hétérogénéité spatiale de l'information spectrale mesurable selon différentes approches, mise en relation avec des indices de diversité taxonomique (Féret et Asner, 2014; Laliberté *et al.*, 2020; Schweiger *et al.*, 2022). Parmi les différentes méthodes qui existent pour relier l'information spectrale à la diversité taxonomique des forêts, certaines s'appuient directement sur une information spectrale continue et un partitionnement de variance (Laliberté *et al.*, 2020), alors que d'autres utilisent une information spectrale discrétisée à l'aide de méthodes de clustering pour en déduire des indicateurs de diversité spectrale analogues aux indicateurs de diversité taxonomique (Féret et Asner, 2014). Si ces dernières méthodes ont plusieurs fois été testées avec succès dans un contexte de forêt tropicale très riche en espèces d'arbre (Féret et Asner, 2014 ; Schäfer *et al.*, 2016), peu d'études à notre connaissance ont été réalisées sur des forêts tempérées moins riches en espèces.

Dans le cadre de ce travail, nous nous intéressons ainsi à l'applicabilité d'une méthode se basant sur la discrétisation de l'espace spectral pour l'estimation de la diversité des forêts tempérées (Féret et Asner, 2014). Bien que non supervisée, cette méthode nécessite de définir le nombre de classes lors de la discrétisation de l'espace spectral. Nous supposons que la valeur de cet hyperparamètre dépend du contexte (tropical ou tempéré) dans lequel est appliquée la méthode et qu'il dépend en particulier du nombre total d'espèces présentes. Nous évaluerons ainsi les performances des indices de diversité spectrale estimés en les comparant à des indices issus d'inventaires forestiers, tout en faisant une analyse de la sensibilité de la méthode au nombre de classes défini par l'utilisateur.

2. Site d'étude et données

2.1. Forêt de Fabas

Le site d'étude est la forêt de Fabas (831 hectares) qui se situe en France dans le département de la Haute-Garonne. Ce massif forestier se caractérise par une diversité importante d'espèces d'arbres. Les relevés de la BD Forêt Version 2 de l'IGN (2007) mentionnent des nombreuses espèces de résineux (douglas, pins divers, épicéa, mélèze et sapin blanc) et la présence de feuillus en mélange, principalement des réserves de chênes et de hêtres. La forêt de Fabas contient ainsi des zones de peuplement très hétérogènes et d'autres plus homogènes.

2.2. Inventaires terrain

Une campagne d'inventaire terrain a été réalisée au cours des mois de mars, avril et juillet 2019. Quarante-quatre placettes de 40 m par 40 m, réparties sur l'ensemble du massif de Fabas, ont été inventoriées. Un relevé sur une placette a consisté à identifier les espèces d'arbre de la strate verticale la plus haute et leur recouvrement dans la placette. Un total de 18 espèces d'arbre a été relevé sur l'ensemble des inventaires avec quatre espèces dominantes représentant plus de 80% de l'abondance cumulée de toutes les placettes : le chêne (*Quercus*), le pin Douglas (*Pseudotsuga menziesii*), le charme (*Carpinus betulus*) et le sapin (*Abies alba*).

2.3. Images hyperspectrales

Les images hyperspectrales aériennes ont été acquises au-dessus de la zone d'étude le 15 septembre 2015 à 8:07 TU (temps universel) à 1287 m d'altitude par ciel clair, à l'aide des instruments HySpex VNIR-1600 et SWIR-320 aéroportés (Norsk Elektro Optikk AS, Lørenskog, Norvège). Les résolutions spatiales des images sont de 1 m et 2 m et les résolutions spectrales sont respectivement de 4,32 et 9,00 nm dans les domaines spectraux Visible/Proche-Infrarouge (VNIR, 400 - 1000 nm) et Infrarouge à courtes longueurs d'ondes (SWIR, 1000 - 2500 nm). Les images ont été produites en luminance et géoréférencées. L'image SWIR a été coregistrée sur la base de l'image VNIR en utilisant l'algorithme Gefolki (Brigot *et al.*, 2016). La méthode de correction atmosphérique Cochise (Poutier *et al.*, 2002) a été appliquée à l'image de luminance résultante pour obtenir une image en réflectance. Les images SWIR et VNIR ont ensuite été ré-échantillonnées à une résolution spatiale de 4 m en utilisant un filtre de voisinage le plus proche, afin de réduire les distorsions géométriques et de préserver les informations spectrales. L'image finale obtenue est une image en réflectance avec 409 bandes spectrales couvrant le domaine de réflexion (400 - 2500 nm) et une résolution spatiale de 4 m.

3. Méthode

3.1. Approche générale

Des indicateurs de diversité ont été estimés directement à partir des données d'imagerie hyperspectrale en appliquant la méthode développée par Féret et Asner (2014) et implémentée dans le package biodivMapR (Féret et de Boissieu, 2020). L'ensemble des traitements et analyses réalisés pour cette étude est présenté en Figure 1.

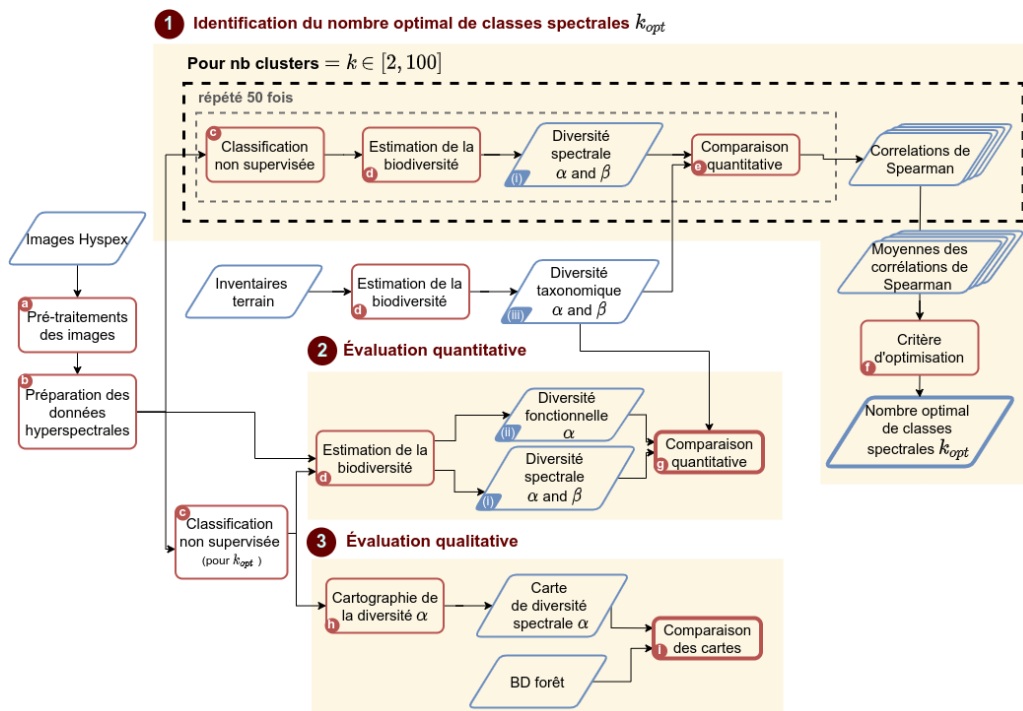


FIGURE 1. Diagramme de flux des traitements et analyses réalisés pour l'évaluation de la précision de l'estimation de la biodiversité forestière à partir d'imagerie hyperspectrale.

Plusieurs étapes de pré-traitement (traitements *a* et *b* dans la Figure 1) sont nécessaires avant le calcul des indicateurs de diversité spectrale (traitements *c* et *d* dans la Figure 1). Deux types d'indicateurs de diversité spectrale peuvent être calculés à partir de biodivMapR: i) les indicateurs associés aux composantes α et β de la diversité taxonomique décrits dans les publications présentant la méthode (annotés (i) dans la Figure 1), et ii) des indicateurs de diversité fonctionnelle décrits par Villeger *et al.* (2008) (annotés (ii) dans la Figure 1). Ces deux types d'indicateurs s'obtiennent à partir d'une séquence de traitement des données hyperspectrales commune, les indicateurs de diversité fonctionnelle se calculant directement à partir d'une information spectrale continue, alors que les indicateurs de diversité taxonomique

nécessitent une étape de clustering afin de déterminer des classes à partir de l'information spectrale continue. Ces classes sont appelées "espèces spectrales" dans l'article original de Féret et Asner (2014) et sont analogues aux "types optiques" définis par Ustin et Gamon (2010). Dans ce travail, nous nous référerons à la dénomination "classes spectrales" afin de ne pas causer une ambiguïté suggérant que la définition taxonomique d'une espèce puisse être associée directement à une classe issue du clustering de l'information spectrale.

Dans le cadre de ce travail, une première étape consiste à estimer le nombre de classes spectrales optimal associé à la classification non supervisée (Figure 1.1). Ensuite, nous comparons la diversité taxonomique α issue des données d'inventaire terrain (annotés (iii) dans la Figure 1) à la diversité spectrale α , et aux indicateurs de diversité fonctionnelle (richesse, équité et divergence fonctionnelles) (Figure 1.2). Enfin, la diversité α est cartographiée et la répartition spatiale de la diversité estimée est comparée avec la BD Forêt V2 afin d'étudier de manière qualitative la qualité des prédictions de diversité (Figure 1.3).

3.2. Estimation de la biodiversité à partir des inventaires terrain

Trois indices sont calculés pour décrire la diversité α et β :

- l'indice de diversité spécifique (S) qui comptabilise le nombre total d'espèces pour une surface considérée (d'un point de vue taxonomique ou spectral) ;
- l'entropie de Shannon (Shannon, 1948) (H) qui prend en compte le nombre d'espèces présentes et leur abondance relative dans une surface considérée ;
- l'indice de dissimilarité de Bray-Curtis (BC) (Bray et Curtis, 1957) qui décrit la dissimilarité de deux relevés terrain en prenant en compte le nombre et l'abondance des espèces en commun. L'indice BC varie de 0 à 1, 0 indiquant deux relevés ayant la même composition spécifique et dans les mêmes proportions et 1 indiquant deux relevés n'ayant aucune espèce en commun.

3.3. Estimation de la biodiversité à partir des données hyperspectrales

3.3.1. Préparation des images

Il est nécessaire d'appliquer plusieurs pré-traitements sur les images hyperspectrales avant le calcul à proprement parler des indices de diversité.

Premièrement, les pixels ne correspondant pas à de la végétation sont identifiés et masqués en utilisant un seuil sur le Normalized Difference Vegetation Index (NDVI) tandis que les pixels correspondant à des ombres sont identifiés et masqués en utilisant un seuil sur la valeur de la réflectance d'une bande Infrarouge centrée autour de 800 nm.

Deuxièmement, l'information spectrale est normalisée afin de réduire la variabilité intracouronne liée aux différences d'exposition au rayonnement solaire incident. Cette normalisation est particulièrement pertinente pour des données à très haute résolution spatiale. La méthode de normalisation consiste ici à appliquer un « *continuum* »

removal » aux données hyperspectrales. Ces traitements sont regroupés dans le traitement *b* dans la Figure 1.

3.3.2. Réduction de dimension

Une méthode de réduction de dimension (i.e. du nombre de variables spectrales) est ensuite appliquée au jeu de données hyperspectrales normalisé. Dans la méthode initialement proposée par Féret et Asner (2014), une analyse en composante principale (ACP) standardisée est appliquée au jeu de données et les composantes permettant de mettre en exergue les couronnes des arbres et l'hétérogénéité du paysage sont sélectionnées visuellement.

Dans cette étude, la méthode de fraction minimale de bruit (« minimum noise fraction » [MNF] en anglais) a été implémentée dans la package *biodivMapR* et est appliquée aux données d'imagerie hyperspectrale. De manière analogue à l'ACP, elle permet de produire un nombre réduit de variables décorréelées entre elles. Plusieurs études soulignent l'efficacité des méthodes MNF sur les données hyperspectrales (Green *et al.*, 1998). Les composantes sont ensuite sélectionnées à l'aide d'une méthode de sélection séquentielle incrémentale, pendant laquelle la corrélation entre H estimé et H calculé à partir des inventaires terrain est maximisée. Ces traitements sont regroupés dans le traitement *b* dans la Figure 1. Les composantes 1 et 6 sont celles qui ont été retenues dans le cadre de cette étude.

3.3.3. Calcul des indices de biodiversité

Deux types d'indice de diversité sont calculés à partir de l'espace spectral réduit précédemment obtenu :

- des indices de diversité analogues aux indices de diversité taxonomique qui requièrent une étape de segmentation de l'espace spectral pour obtenir des classes spectrales discrètes ;
- des indices de diversité fonctionnelle qui peuvent être calculés directement sur l'espace spectral continu.

Les indices de diversité taxonomique doivent être calculés à partir de données de présence ou abondance d'espèces. L'espace spectral issu de la MNF est donc segmenté à l'aide de l'algorithme de clustering k-means pour obtenir un nombre *k* de classes spectrales. Une classe spectrale peut alors être assignée à chaque pixel de l'image. Les indices S et H sont ensuite calculés sur une fenêtre correspondant à l'emprise des placettes de 40 m de côté des inventaires terrain, soit environ 100 pixels sur les données de résolution spatiale de 4 m. La diversité β est enfin calculée en calculant la dissimilarité de BC entre chaque paire de placettes.

Les indices de diversité fonctionnelle sont quant à eux directement calculés sur l'espace spectral continu. Trois indices ont été implémentés dans le package *biodivMapR*, d'après la définition initialement proposée par Villéger *et al.* (2008) pour un jeu de traits fonctionnels de végétation, remplacés ici par des traits spectraux correspondant aux composantes sélectionnées suite à la MNF:

- la richesse fonctionnelle correspondant au volume occupé par la communauté présente dans une surface donnée ;
- l'équitabilité fonctionnelle qui décrit la régularité de la distribution des abondances des espèces présentes dans une surface donnée ;
- la divergence fonctionnelle, définie comme la distance moyenne des espèces d'une surface donnée par rapport à leur centre de gravité dans l'espace fonctionnel.

Ces indices sont également calculés pour une surface correspondant à l'emprise des placette terrains.

3.4. Identification du nombre optimal de classes spectrales

Les indices de diversité taxonomique sont calculés à partir d'une segmentation de l'espace spectral. Le nombre de classes spectrales est un hyperparamètre de la méthode à définir par l'utilisateur. L'hypothèse derrière cette segmentation est que chaque classe spectrale obtenue est liée à une espèce d'arbre ou un groupe d'espèces d'arbre similaires spectralement. Dans le cas idéal d'une séparabilité spectrale parfaite des espèces présentes dans l'image, le nombre de classes spectrales devrait donc être défini en fonction du nombre d'espèces présentes sur le site d'étude. Cette hypothèse est testée ici et l'influence du nombre de classes spectrales sur la corrélation entre les indices de diversité estimés et mesurés est étudiée : la chaîne de traitements du package `biodivMapR` est exécutée pour un nombre de classes spectrales variant de 2 à 100 puis la corrélation entre les indices de diversité estimés et ceux mesurés est évaluée. La corrélation non paramétrique de Spearman (ρ) est utilisée pour les indices diversité α tandis que la corrélation de Pearson (r_p) issue d'un test de Mantel entre les matrices de dissimilarité de Bray-Curtis est utilisée pour la diversité β (traitement e dans la Figure 1.1). Cette opération est répétée 50 fois pour chaque nombre de classes spectrales (étant donné le caractère aléatoire de l'initialisation du `k-means`). Le nombre de classe qui maximise la moyenne des 50 valeurs de ρ est retenu comme nombre de classes optimal k_{opt} (traitement f dans la Figure 1.1).

3.5. Évaluation de la qualité des prédictions de la diversité

La qualité de l'estimation de la biodiversité à partir de données hyperspectrales est évaluée de deux manières. Les corrélations entre les indices de diversité estimés et les indices de diversité issus du terrain sont étudiées dans un premier temps (Figure 1.2), puis les indices de diversité sont spatialisés sous forme de carte de diversité et les patrons de diversité obtenus sont comparés à la BD Forêt V2 dans un deuxième temps (Figure 1.3).

3.5.1. Comparaison des données terrains avec les données prédites

Les indices de diversité taxonomique et fonctionnelle sont calculés pour le nombre optimal de classes spectrales déterminé précédemment. Les corrélations entre ces indices et les données des inventaires terrains sont calculées afin d'évaluer quantitativement la qualité des prédictions de diversité α et β du modèle. De manière analogue à l'étape d'identification du nombre optimal de cluster, la corrélation de

Spearman ρ est utilisée pour les indices diversité α . Chaque type d'indices est comparé avec les données des inventaires terrain dans un premier temps puis les indices taxonomiques et fonctionnels sont comparés entre eux pour déterminer s'ils contiennent une information similaire (traitement g de la figure 1.2).

3.5.2. Comparaison des cartes de diversité avec la BD Forêt

Les indices de diversité α sont ensuite spatialisés sous forme de carte à partir de la distribution des classes spectrales dans et entre les fenêtres individuelles de 10 pixels de côté (traitement h de la figure 1.3). Ces cartes sont particulièrement pertinentes pour l'analyse qualitative des modèles de diversité spatiale obtenus à l'échelle du site ou du paysage et complètent l'évaluation quantitative réalisée pour valider la méthode sur un ensemble limité de placettes inventoriées. Ici, les analyses qualitatives et quantitatives sont combinées et ces cartes d'indices de diversité estimés sont comparées avec les informations fournies par la BD Forêt Version 2 (traitement i de la figure 1.3).

Tout d'abord, les zones forestières correspondant aux espèces pures et aux espèces mixtes dans la BD Forêt Version 2 sont identifiées. Les valeurs H correspondant aux peuplements d'espèces pures et aux peuplements d'espèces mixtes sont ensuite extraites et leurs distributions comparées. Il est attendu que la distribution de H correspondant aux peuplements d'espèces pures soit significativement plus faible que celle correspondant aux peuplements d'espèces mixtes.

4. Résultats

4.1. Nombre optimal de clusters

L'évolution des corrélations entre les indices de diversité estimés et mesurés en fonction du nombre de classes spectrales est présentée sur la figure 2. Pour les deux types diversité α et β , il existe un nombre de classes spectrales pour lequel les corrélations entre diversité spectrale et mesurée sont maximisées.

Les corrélations pour S et H augmentent fortement entre $k = 2$ ($\rho < 0.3$) et $k=6$ ($\rho = 0.68$). Les corrélations augmentent ensuite plus progressivement pour atteindre un maximum pour $k = 26$ ($\rho = 0.76$) pour S et $k = 23$ ($\rho = 0.82$) pour H . Enfin, les corrélations diminuent faiblement de manière quasi-linéaire jusqu'à $k = 100$ pour atteindre 0.63 et 0.68.

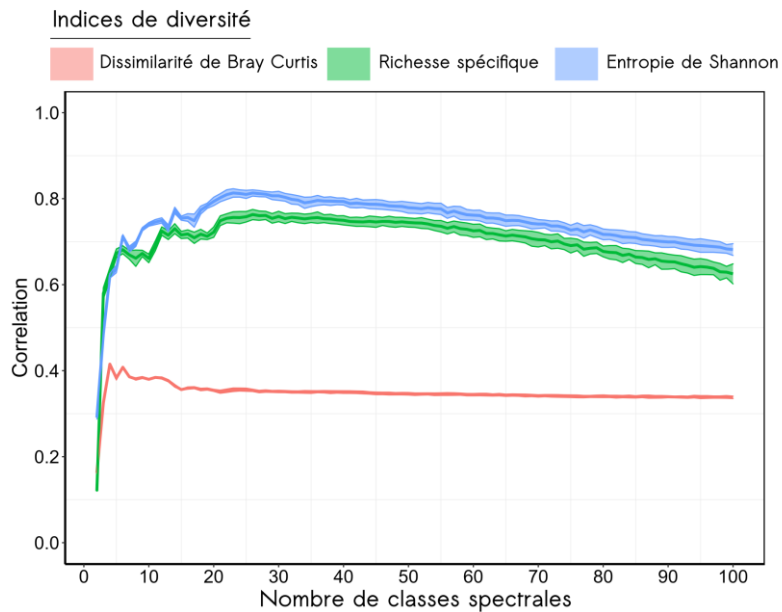


FIGURE 2. Moyenne des corrélations pour 50 répétitions entre les indices de diversité spectraux et les indices de diversité mesurés sur le terrain, en fonction du nombre de classes spectrales. L'enveloppe autour des courbes correspond à \pm l'écart-type des répétitions. La corrélation de Spearman est utilisée pour les indices de diversité α et la corrélation de Pearson de tests de Mantel est utilisée pour la dissimilarité de BC.

Les corrélations entre la diversité β spectrale et la diversité β mesurée sont globalement inférieures à celles obtenues pour la diversité α . La corrélation maximale est obtenue pour $k = 6$ ($r_p = 0.41$) puis diminue jusqu'à $k = 15$ ($r_p = 0.35$). Au-delà, le nombre de classe spectrale a une faible influence sur la corrélation qui diminue très légèrement jusqu'à 0.33 pour $k=100$.

4.2. Relation entre la biodiversité estimée et mesurée pour le nombre optimal de classes spectrales.

Les corrélations de Spearman entre les indices diversité spectrale taxonomique et fonctionnelle, et les indices de diversité issus des inventaires terrain sont présentés sur la figure 3.

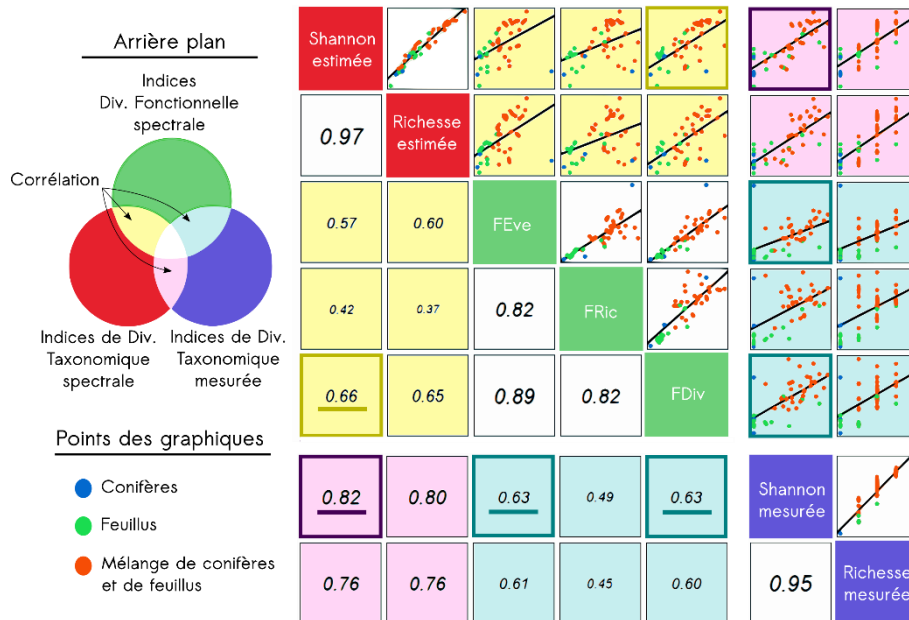


FIGURE 3. Scatter plot et matrice de corrélation entre les indices de diversité spectrale et les indices de diversité mesurée sur le terrain. Les corrélations les plus fortes sont soulignées et encadrées. Les corrélations entre les indices sont indiquées par des arrières plan de couleurs jaune, magenta et cyan.

La plus forte corrélation entre les indices de diversité spectrale et de diversité issue des inventaires terrain est obtenue entre la diversité de Shannon estimée et la diversité de Shannon mesurée ($\rho = 0.82$). Les corrélations des indices de diversité taxonomique (en magenta sur la figure 3) sont toutes supérieures aux valeurs de corrélation des indices de diversité fonctionnelles (en cyan sur la figure 3). Parmi les corrélations entre les indices de diversité fonctionnelle et les indices de diversité mesurée, les corrélations les plus fortes sont celles entre H mesurée (Shannon) et les indices de diversité fonctionnelle FEve et FDiv ($\rho=0.63$).

Parmi les indices de diversité spectrale, les corrélations entre les diversités taxonomiques et fonctionnelles sont faibles à modérées, avec des valeurs allant de 0.37 à 0.66 entre H estimé et la FDiv.

4.3. Cartographie de la biodiversité

La comparaison entre la carte de H estimé avec BiodivMapR et la distribution des peuplements forestiers purs et mélangés identifiés en tant que tel dans la BD Forêt Version 2 est présentée dans la Figure 4a.

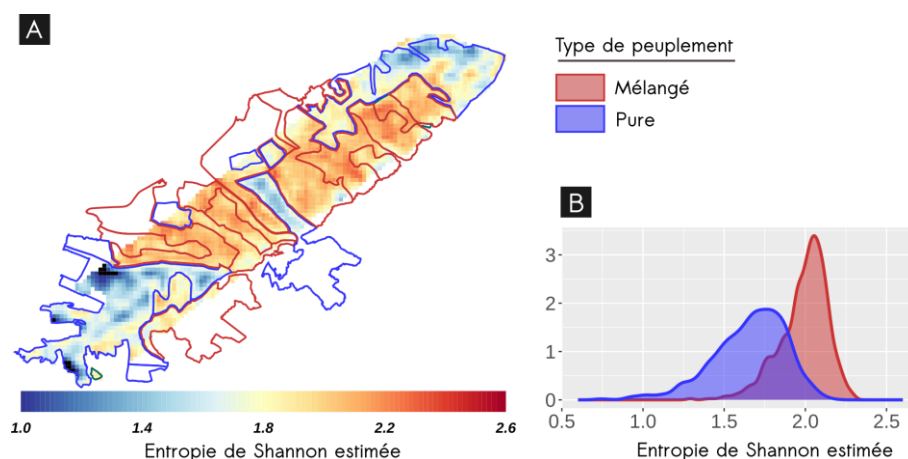


FIGURE 4. (A) Carte de l'entropie de Shannon estimée par BiodivMapR sur le site de Fabas. Un filtre spatial médian a été appliqué sur la carte. Les peuplements mélangés et purs issus de la BD forêt sont superposés en rouge et bleu sur la carte. (B) Distribution des valeurs de l'entropie de Shannon des peuplements purs et mélangés.

Ces cartes montrent que la distribution spatiale des peuplements forestiers purs et mélangés correspond à la distribution de H : les valeurs les plus faibles de H correspondent aux peuplements identifiés comme purs, tandis que les valeurs les plus élevées de H correspondent aux peuplements identifiés comme mélangés. La distribution de H correspondant aux peuplements forestiers purs et mélangés est affichée dans la figure 4b et montre en effet une distribution significativement différente entre les deux types de forêts.

5. Discussion

5.1 Faut-il adapter le nombre de classes spectrales de la méthode au site d'étude ?

Dans cette étude, nous avons émis l'hypothèse que le nombre de classes spectrales avait une influence sur les performances dans l'estimation des indices de diversité en phase de validation. Les résultats montrent que le nombre de classes spectrales a bien une influence sur la précision de l'estimation de la diversité. Cette influence est particulièrement importante pour un faible nombre de classes, que ce soit pour les indices de diversité α ou β . Les valeurs des corrélations entre les indices de diversité spectrale et les indices de diversité taxonomique issus des inventaires terrain augmentent très fortement de $k = 2$ à $k = 5$. Les variations sont ensuite moins importantes voire sont très faibles pour $k > 15$ dans le cas de la diversité β .

Nous avons supposé que le nombre de classes spectrales devait être choisi en lien avec le nombre d'espèces d'arbre présentes sur le site étudié. Les résultats indiquent qu'il existe bien un nombre de classes spectrales pour lesquelles la corrélation entre

les indices de diversité spectrale estimée et les indices issus des inventaires terrain est maximisée. Ce nombre est de $k_{opt}=23$ pour H et de $k_{opt}=26$ pour S. Si ce nombre de classes spectrales est bien inférieur au nombre de classes spectrales utilisé dans les milieux tropicaux par Féret et Asner (2014) ($k=40$) ou Schäfer *et al.* (2016) ($k=55$), il reste supérieur au nombre total d'espèces présentes sur le site de Fabas (18 d'après les relevés). Par ailleurs, les corrélations obtenues pour ces valeurs de k définies dans un contexte tropical, même si non optimales, restent très élevées : pour H, $\rho=0.79$ et $\rho=0.77$ pour ces deux valeurs de k .

Finalement, ce qui semble être déterminant est de discrétiser suffisamment l'espace spectral pour que les classes spectrales puissent rendre compte de la l'hétérogénéité spectrale liée à la diversité taxonomique α . Cette étude suggère un minimum de cinq à six classes. Quant au nombre précis à choisir, s'il existe un nombre optimal de classes spectrales qui permette de maximiser les corrélations des indices spectraux avec les indices issus des inventaires terrain, il n'est pas évident que ce nombre soit lié au nombre total d'espèces présentes sur le site. Le choix d'un nombre de classes similaire aux milieux tropicaux permet en effet d'obtenir des relations similaires.

5.2 L'estimation de la diversité à partir de classes spectrales est possible pour les milieux forestiers tempérés.

5.2.1 Corrélations avec les inventaires terrain

La méthode d'estimation de la biodiversité par imagerie hyperspectrale produit des indices de diversité α ayant une forte corrélation avec les indices issus des inventaires terrain. Après optimisation du nombre de classes spectrales, la corrélation de Spearman obtenue entre mesure et estimation était de 0.76 pour S et de 0.82 pour H. En revanche, la corrélation entre l'indice de diversité β estimé et mesuré à partir des inventaires terrain était plus faible ($r_p = 0.41$). Cela suggère que les classes spectrales ne correspondent pas strictement aux espèces taxonomiques des inventaires terrains. L'estimation de la diversité taxonomique α à l'aide d'indicateurs de diversité produits par `biodivMapR` apparaît donc pertinent dans le contexte de forêts de milieux tempérés, même si l'hypothèse de classes spectrales correspondant à des unités taxonomiques (espèces ou groupes d'espèces) n'est pas vérifiée lors de notre expérience.

5.2.2 Des variations spatiales à l'échelle du massif bien identifiées

La méthode implémentée dans le package `biodivMapR` permet de spatialiser la diversité α sous forme de carte. Au-delà de leur relation quantitative avec les données des inventaires terrain, les cartes produites montrent que la diversité issue de la discrétisation de l'information spectrale correspond spatialement aux peuplements décrits par la BD Forêt. Les zones supposées comme ayant une forte diversité en espèces d'arbre correspondent en effet aux zones avec des fortes valeurs de H. Ces cartes permettent par ailleurs de décrire des variations de la diversité α à une échelle plus fine que celles des peuplements décrits par la BD Forêt.

5.3 Diversité spectrale fonctionnelle et diversité spectrale taxonomique

En plus des indices de diversité α et β initialement calculés dans la méthode développée par Féret et Asner (2014), des indices de diversité fonctionnelle ont également été calculés dans cette étude. Un avantage de ces indices est que leur calcul ne requiert pas de discrétisation de l'espace et qu'ils rendent directement compte de l'hétérogénéité spectrale d'une placette dans l'espace spectral.

Ces indices permettent eux aussi de mettre en évidence un lien entre l'hétérogénéité spectrale et la diversité taxonomique dans les milieux forestiers tempérés. Cependant, même si des corrélations relativement élevées ont été obtenues ($\rho=0.63$ pour FEve et FDiv), elles sont inférieures aux corrélations obtenues avec les indices utilisant les classes spectrales. Par ailleurs, les corrélations entre les deux types d'indices sont modérées (ρ compris entre 0.37 à 0.66). Ces résultats suggèrent ainsi que les deux types d'indices ne caractérisent pas l'hétérogénéité spectrale de la même manière et capturent des composantes complémentaires de la diversité. Il est également possible que les indices de diversité fonctionnelle soient plus sensibles à une hétérogénéité spectrale plutôt liée à des facteurs abiotiques comme c'est le cas dans d'autres études (ex. Wang *et al.*, 2018).

Il est toutefois important de noter que la quantité de données disponibles pour la validation restant relativement modérée (44 placettes), les performances obtenues lors de la comparaison entre les différents indices de diversité spectrale et les indices de diversité taxonomique ont montré une forte sensibilité à l'erreur associée à un nombre limité de parcelles. Une analyse intégrant une plus grande quantité de parcelles inventoriées pour validation permettrait de mieux caractériser la robustesse des performances rapportées sur cet ensemble de validation.

6. Conclusion

Le développement de méthodes opérationnelles pour la caractérisation et le suivi de la biodiversité par télédétection est crucial pour obtenir des informations sur de grandes étendues de manière répétée et standardisée. La méthode initialement développée par Féret et Asner (2014) dans un contexte de forêt tropicale s'avère également pertinente pour caractériser la diversité taxonomique dans un contexte tempéré. De plus, les résultats de cette étude montrent qu'au-delà d'un nombre minimum de classes spectrales, les performances lors de la mise en relation entre indices de diversité spectrale et taxonomique sont modérément impactées, ce qui suggère qu'il n'est pas nécessaire d'adapter précisément le nombre de classes spectrales en fonction du nombre d'espèces présentes sur le site étudié. Cela incite à analyser de manière plus approfondie le lien entre les classes spectrales identifiées par la méthode et les espèces taxonomiques afin de mieux identifier le potentiel et les limites de cette approche. Les corrélations modérées entre les indices de diversité β mesurée et estimée soutiennent par ailleurs une correspondance limitée entre classes spectrales et espèces taxonomiques. Ainsi, si la discrétisation de l'espace spectral et l'utilisation de classes spectrales semble nécessaire pour obtenir les plus fortes relations entre la diversité taxonomique mesurée et la diversité spectrale, la nature de

ce lien reste encore mal décrite. Plus de recherches étudiant spécifiquement cet aspect sont encore nécessaires pour bien comprendre dans quelles conditions le concept de classes spectrales peut être utilisé comme descripteur de la diversité taxonomique.

Bibliographie

- Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 325-349. <https://doi.org/10.2307/1942268>
- Brigot, G., Colin-Koeniguer, E., Plyer, A., & Janez, F. (2016). Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(7), 2923-2939. <https://doi.org/10.1109/JSTARS.2016.2578362>
- Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87. <https://doi.org/10.1016/j.rse.2016.08.013>
- Féret, J., & Asner, G. P. (2013). *Mapping the geographic distribution of canopy species communities in lowland Amazon rainforest with CAO-AToMS (Invited)*. 2013, B52D-01.
- Féret, J., & Boissieu, F. (2020). biodivMapR: An R package for α - and β -diversity mapping using remotely sensed images. *Methods in Ecology and Evolution*, 11(1), 64-70. <https://doi.org/10.1111/2041-210X.13310>
- Féret, J.-B., & Asner, G. P. (2014). Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecological Applications*, 24(6), 1289-1296. <https://doi.org/10.1890/13-1824.1>
- Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65-74. <https://doi.org/10.1109/36.3001>
- Laliberté, E., Schweiger, A. K., & Legendre, P. (2020). Partitioning plant spectral diversity into alpha and beta components. *Ecology Letters*, 23(2), 370-380. <https://doi.org/10.1111/ele.13429>
- Laybros, A., Schläpfer, D., Féret, J.-B., Descroix, L., Bedeau, C., Lefevre, M.-J., & Vincent, G. (2019). Across Date Species Detection Using Airborne Imaging Spectroscopy. *Remote Sensing*, 11(7), 789. <https://doi.org/10.3390/rs11070789>
- Mirkin, B. (2011). Choosing the number of clusters. *WIREs Data Mining and Knowledge Discovery*, 1(3), 252-260. <https://doi.org/10.1002/widm.15>
- Niemelä, J. (2000). Biodiversity monitoring for decision-making. *Annales Zoologici Fennici*, 307-317.
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Hoft, R., Hurtt, G.,

- Jetz, W., ... Wegmann, M. (2013). Essential Biodiversity Variables. *Science*, 339(6117), 277-278. <https://doi.org/10.1126/science.1229931>
- Poutier, L., Miesch, C., Lenot, X., Achard, V., & Boucher, Y. (2002). COMANCHE and COCHISE: two reciprocal atmospheric codes for hyperspectral remote sensing. *2002 AVIRIS Earth Science and Applications Workshop Proceedings*, 1059-0889.
- Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., & Schellnhuber, H. J. (2009). A safe operating space for humanity. *nature*, 461(7263), 472.
- Schäfer, E., Heiskanen, J., Heikinheimo, V., & Pellikka, P. (2016). Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. *Ecological Indicators*, 64, 49-58. <https://doi.org/10.1016/j.ecolind.2015.12.026>
- Schweiger, A. K., & Laliberté, E. (2022). Plant beta-diversity across biomes captured by imaging spectroscopy. *Nature Communications*, 13(1), 2767. <https://doi.org/10.1038/s41467-022-30369-6>
- Shannon, C. E. (1948). (1948), "A Mathematical Theory of Communication", *Bell System Technical Journal*, vol. 27, pp. 379-423 & 623-656, July & October.
- Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*, 186(4), 795-816. <https://doi.org/10.1111/j.1469-8137.2010.03284.x>
- Vihervaara, P., Auvinen, A.-P., Mononen, L., Törmä, M., Ahlroth, P., Anttila, S., Böttcher, K., Forsius, M., Heino, J., Heliölä, J., Koskelainen, M., Kuussaari, M., Meissner, K., Ojala, O., Tuominen, S., Viitasalo, M., & Virkkala, R. (2017). How Essential Biodiversity Variables and remote sensing can help national biodiversity monitoring. *Global Ecology and Conservation*, 10, 43-59. <https://doi.org/10.1016/j.gecco.2017.01.007>
- Villéger, S., Mason, N. W. H., & Mouillot, D. (2008). NEW MULTIDIMENSIONAL FUNCTIONAL DIVERSITY INDICES FOR A MULTIFACETED FRAMEWORK IN FUNCTIONAL ECOLOGY. *Ecology*, 89(8), 2290-2301. <https://doi.org/10.1890/07-1206.1>
- Wang, R., & Gamon, J. A. (2019). Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment*, 231, 111218. <https://doi.org/10.1016/j.rse.2019.111218>
- Wang, R., Gamon, J. A., Schweiger, A. K., Cavender-Bares, J., Townsend, P. A., Zyguelbaum, A. I., & Kothari, S. (2018). Influence of species richness, evenness, and composition on optical diversity : A simulation study. *Remote Sensing of Environment*, 211, 218-228. <https://doi.org/10.1016/j.rse.2018.04.010>

Cartographeur des trajectoires maritimes incertaines du XVIII^{ème} siècle

Christine Plumejeaud-Perreau¹, Bernard Pradines¹

1. UMR 7301 MIGRINTER, CNRS & Université de Poitiers
5 rue des Theodores Lefebvre, 86073 Poitiers, France
prenom.nom@univ-poitiers.fr

RÉSUMÉ. Cet article présente comment des trajectoires de navires ont été reconstruites à partir de sources historiques concernant le commerce maritime du XVIII^{ème} siècle. Il s'intéresse dans un premier temps à la qualification de l'incertitude inhérente à ces sources, en expliquant la méthode de reconstitution des itinéraires. Il détaille ensuite comment la géométrie des tronçons de trajectoire reliant les escales par la mer ont automatiquement été calculés en vue de leur cartographie. L'algorithme programmé en PL/SQL est transposable à d'autres contextes de recherche et disponible sous licence libre. Il expose enfin un outil de recherche et géovisualisation sur le Web de ces trajectoires avec leur niveau d'incertitude associé et son utilité pour les historiens, mais aussi pour l'analyse et la vérification des trajectoires ainsi reconstruites.

ABSTRACT. This article presents how ship trajectories have been built from historical sources dealing with maritime trade in the 18th century. It first summarizes the method for building the routes, and qualifying the uncertainty level linked to each of its segments. Then, it details how the geometries of these segments connecting two successive stopovers were automatically calculated in order to draw a map with maritime paths only. The algorithm, programmed with PL/SQL language, is available under an open-source licence (<https://gitlab.huma-num.fr/portic/porticapi>). Finally, an online tool for querying and mapping these routes with their associated level of uncertainty is exposed (<http://shiproutes.portic.fr>). We show its usefulness for historians, in particular for the control of the validity of built ship trajectories.

MOTS-CLES : Cartographie de trajectoire, calcul de chemin, incertitude, géovisualisation

KEYWORDS: trajectory mapping, pathfinding algorithm, uncertainty, maritime routes, geovisualisation

1. Introduction

Cette recherche s'intéresse à l'incertitude des données historiques portant sur le commerce maritime au XVIIIème siècle afin d'en proposer une visualisation expressive et en particulier cartographique. En effet, nous étudions des sources saisies dans une base de données spatio-temporelle qui nous permettent de reconstituer au moins une année de circuits de navigation sur les côtes françaises, du ponant au levant. L'apport en termes d'analyse est considérable - par exemple pour déterminer les aires de navigation en fonction du tonnage ou encore la durée des rotations - à la condition de lever ou d'explicitier les incertitudes inhérentes à cette reconstitution. Or, dans sa volonté de proposer une visualisation expressive et en particulier cartographique des données sur la navigation, l'équipe interdisciplinaire a fait le choix de donner à voir cette incertitude, ce qui requiert un travail complexe de curation des données et de qualification des degrés d'incertitude des principales variables.

La base de données contient, entre autres, la transcription numérisée de plus de 30 000 congés de navires dans les ports français en 1787. Un congé est une déclaration fiscale d'un capitaine de navire à l'Amirauté d'un port qui l'autorise à partir. En outre, cette base de données contient deux autres sources, à savoir le "Registre sanitaire de Marseille" et les "Cahiers du petit cabotage". Certaines descriptions de navires correspondent entre ces sources et peuvent confirmer une déclaration de l'autre source. Cette vérification est rendue possible par l'identification de navires semblables à travers ces différentes sources.

Cet article élicite d'abord la nature de l'incertitude que nous souhaitons représenter, à partir d'exemples concrets. L'information qui est traitée est de nature très imparfaite (Devillers et Jeansoulin 2005) du fait d'informations inégalement disponibles, de l'imprécision de la formulation des lieux ou des dates, mais également d'informations fausses ou contradictoires. Cette communication résume les règles de l'algorithme qui permet de contrôler la cohérence des informations, (Plumejeaud-Perreau et Marzagalli, 2022), afin de se focaliser sur la cartographie des itinéraires maritimes. Nous proposons un outil de géovisualisation permettant de rechercher et cartographier ces itinéraires, lesquels sont calculés automatiquement par des chemins maritimes suivant une méthode inédite.

Notre article est organisé suivant trois parties. La première résume la méthode de qualification automatique de l'incertitude inhérente à ces hypothétiques itinéraires. La deuxième partie détaille comment les géométries de tronçons reliant deux escales successives par la mer sont automatiquement calculées en vue de leur cartographie. La troisième partie présente un outil de recherche et géovisualisation de ces trajectoires avec leur niveau d'incertitude associé et son utilité pour les historiens.

2. Reconstitution de trajectoires à partir de sources historiques

Cette section décrit comment les trajectoires de navires peuvent être reconstituées à partir d'une base collectant des déclarations d'itinéraires, de sources hétérogènes, dont deux majeures qui sont l'une, les congés fiscaux des Amirautés de France pour 1787 et 1789 qui constituent une partie importante (67 %) de la base de données, et l'autre les recensements sanitaires à l'arrivée à Marseille des navires pour les années 1749, 1759, 1769, 1779, 1787, 1789, 1799.

Dans les congés, les capitaines déclarent au lieu de départ une intention de destination, alors que dans le registre sanitaire de Marseille, ce sont les escales précédant l'arrivée à Marseille qui sont connues avec certitude. La base de données qui enregistre les sources est structurée autour de la notion centrale de point d'escale, qui est un événement se déroulant en un lieu à un instant. En dépit de l'imprécision de la mention du lieu d'escale parfois, celui-ci est rapporté à un des lieux référencés dans notre index géographique, le plus souvent un port mais pas toujours (un naufrage au large des côtes, une destination vague comme « Côte de Bretagne »), en notant son identifiant en plus de la variante toponymique lue dans la source. Cet index a aussi l'avantage de préciser l'entité géographique administrative d'appartenance de 1787, Etat, province, Amirauté ou autre (Plumejeaud et al., 2021). Concernant la date de l'escale, elle est renseignée soit comme une date de sortie ou une date d'entrée, en fonction du mouvement du navire, et elle est précise au jour près sur le lieu d'observation. Les dates sont dans un format spécial : par exemple, 1787=05=10 si exactement à la date du 10 mai 1787, 1787>05>10 si après cette date, ou 1787<05<10 si avant cette date. Ce format spécifie ainsi l'incertitude sur la date d'arrivée à une destination dans les congés, ou de sortie d'une escale précédant l'arrivée à Marseille. Il autorise un ordonnancement lexicographique équivalent à l'ordre temporel, et peut être converti en format temporel normé ISO 8601 en remplaçant les caractères spéciaux (=, < ou >) par des tirets.

Tout document source, qui correspond à une ligne ou un feuillet de déclaration par une même personne dans un port un certain jour, est retranscrit dans une table centrale, nommée « *pointcall* ». Cette table décrit les escales point par point réalisées par un navire, piloté par un capitaine, suivant un document source auquel la base attribue un identifiant (*data_block_local_id*). Par exemple, pour un document référencé « 00188143 » dans les sources décrivant le trajet d'un navire du Havre à Marseille en passant par le Détroit de Gibraltar, la relation *pointcall* contient 3 entrées, décrivant l'action (*pointcall_action*) réalisée par le navire en ces trois ports qui peut être grosso modo soit d'arriver, soit de partir, soit de partir pêcher dans le voisinage côtier, soit de faire un transit. La saisie des informations communes (nom du navire et du capitaine, tonnage, type de navire, pavillon, port d'attache parfois, origine du capitaine, taille de l'équipage, etc.) se fait au point d'observation, celui où est enregistré la déclaration dans le document source, qui est marqué par un attribut *pointcall_function* égal à 'O'. Ailleurs, la description est recopiée automatiquement. Un attribut *pointcall_status* explicite si d'après le document, l'évènement a vraiment eu lieu dans le passé (il a un statut 'PC') ou bien si c'est une intention annoncée (statut 'FC') ou bien si c'est une intention qui n'a pas pu être réalisée (statut 'PU').

Cette information est essentielle puisqu'elle va permettre de rendre compte d'incertitudes historiques. Peut-on en effet confirmer par le croisement de sources qu'une intention (statut 'FC') s'est vu réalisée ultérieurement ? L'ordre de visite de l'escale pour le navire dans ce document est renseigné manuellement dans *pointcall_rank*. Il va permettre d'ordonner les escales dans l'ordre chronologique, plus facilement que les dates indiquées dans l'un (et/ou l'autre) des attributs *pointcall_out_date* et *pointcall_in_date*. Ceci résume ici quelques informations essentielles à la compréhension de la transcription des sources dans la base, décrite par ailleurs (Dedieu et al., 2011).

L'information peut faire l'objet plusieurs interprétations qui vont être ajoutées sur la ligne saisie immédiatement ou *a posteriori* : reconnaître et identifier un même capitaine et lui attribuer un identifiant (*captain_id*), ou un navire déjà repéré dans d'autres documents source et l'identifier par un numéro unique (*ship_id*) : 13 820 navires différents ont été identifiés à ce jour. Notre étude écarte 28 % des 65 030 documents n'ont pas encore fait l'objet de cette identification. Pour la suite, nous allons supposer que cette identification est fiable, et nous concentrer sur les différents types d'incertitude relatifs aux déclarations du mouvement des navires dans le temps (désignés comme chemins ou itinéraires).

2.1. Vérification automatique de l'enchaînement des escales

Toutes les sources ne présentent pas le même statut du point de vue de la réalité des faits énoncés. Ainsi, les sources marseillaises concernent des arrivées, donc des événements passés au moment du récit aux officiers de santé (*pointcall_status* = PC) : le registre mentionne le lieu et la date exacte d'où le navire a appareillé, ainsi que tous les endroits où il a fait escale avant Marseille, certifiés par un visa daté et tamponné. Tous ces lieux sont qualifiés de « observés », *i.e.* certains, dans la variable *pointcall_uncertainty*. Dans quelques cas, les registres de Marseille mentionnent la prochaine destination prévue. Comme pour tous les événements futurs (*pointcall_status* = FC), il est possible que cette intention n'ait pas été réalisée et le pointcall est qualifié de « déclaré ». Dans le cas des congés, le lieu de prise du congé est « observé », donc certain, tandis que les destinations annoncées sont qualifiées *a priori* de « déclarées ». Il arrive parfois qu'un capitaine indique sa précédente escale (*pointcall_status* = PC) en prenant son congé, mais elle n'est pas certaine car c'est une déclaration, qualifiée comme telle : « déclarée ».

La reconstitution d'itinéraire peut se faire en ordonnant chronologiquement les escales d'un même navire. C'est le cas par exemple du navire « la Fidèle Mariane » suivi à travers 6 documents, dont les déclarations sont toutes parfaitement cohérentes : *pointcall_uncertainty* prend la valeur de « confirmé » pour les intentions d'escales en vert dans le tableau 1 qui sont confirmées par l'observation consécutive. Quand les dires sont concordants, il y a un doublon au point d'arrivée et de départ, comme pour la Fidèle Mariane. Pour reconstituer l'itinéraire, il faudrait simplement éliminer un des doublons, en gardant si possible l'escale précisément datée (donc avec *pointcall_function* à 'O'). C'est un des rôles dévolus à la variable « *net_route_marker* » introduite par les historiens, lorsqu'elle prend la valeur 'A' (pour conserver l'escale), ou 'Z' (pour éliminer le doublon).

TABLE 1. Ordonancement des escales du *Fidèle Marianne*, et confirmation de ses déclarations d'intention.

data_block_local_id	ship_id	ship_name	pointcall	out_date	rank	net_route_marker	function	status
00062213	0002931N	Fidèle Mariane	Les Sables-d'Olonne	1787=01=05	1	A	O	PC
00062213	0002931N	Fidèle Mariane	Bayonne		2	Z	T	FC
00140197	0002931N	Fidèle Mariane	Bayonne	1787=03=16	1	A	O	PC
00140197	0002931N	Fidèle Mariane	Dunkerque		2	Z	T	FC
00110541	0002931N	Fidèle Mariane	Dunkerque	1787=05=10	1	A	O	PC
00110541	0002931N	Fidèle Mariane	Les Sables-d'Olonne		2	Z	T	FC
00102845	0002931N	Fidèle Mariane	Les Sables-d'Olonne	1787=06=18	1	A	O	PC
00102845	0002931N	Fidèle Mariane	Bayonne		2	Z	T	FC
00140566	0002931N	Fidèle Mariane	Bayonne	1787=09=04	1	A	O	PC
00140566	0002931N	Fidèle Mariane	Saint-Malo		2	Z	T	FC
00142100	0002931N	Fidèle Mariane	Saint-Malo	1787=10=08	1	A	O	PC
00142100	0002931N	Fidèle Mariane	Saint-Brieuc	1787<10<10!	2	A		FC
00142100	0002931N	Fidèle Mariane	Côtes de Bretagne		3	A	T	FC

TABLE 2. Ordonancement des escales du *Suzanne*, avec croisement des sources de Marseille et des congés de l'Amirauté.

data_block_local_id	ship_id	ship_name	pointcall	out_date	in_date	rank	function	Net_Route_marker	status
00294615	0012925N	Suzanne	La Rochelle	1787=06=01	None	1	O	A	PC
00294615	0012925N	Suzanne	Seudre	None	1787>06>01!	2	T	A	FC
00149798	0012925N	Suzanne	La Tremblade	None	1787=06=05	1	O	A	PC
00151273	0012925N	Suzanne	Marennes	1787=06=19	None	1	O	A	PC
00151273	0012925N	Suzanne	Le Havre	None	1787>06>19!	2	T	Z	FC
00162284	0012925N	Suzanne	Le Havre	1787=08=04	None	1	O	A	PC
00162284	0012925N	Suzanne	Marseille	None	1787>08>04!	2	T	Z	FC
00188143	0012925N	Suzanne	Le Havre	1787=08=06	None	1	A	Z	PC
00188143	0012925N	Suzanne	Détroit de Gibraltar	None	1787=08=24	2	None	historien Z / algorithme A	PC
00188143	0012925N	Suzanne	Marseille	None	1787=09=14	3	O	A	PC

Mais parfois, comme à Marseille, on retrouve des navires qui ont pris congé sur l'atlantique par exemple en annonçant leur intention future d'aller à Marseille. A Marseille, lorsqu'on les retrouve, les étapes détaillées depuis ce port atlantique sont détaillées comme pour le navire Suzanne (tableau 2) : parti du Havre le 6 aout 1787, alors qu'il a pris son congé le 4 aout 1787, le navire a passé le détroit de Gibraltar le 24 aout 1787 avant d'arriver à Marseille le 14 septembre 1787. Ainsi des intentions d'escales s'intercalent entre les vraies escales dans l'ordre chronologique et peuvent tromper la simple logique de l'ordonnement chronologique : le Suzanne n'a pas navigué du Havre à Marseille puis au Havre en deux jours. L'insertion d'un Z devant ces escales « intentionnelles » (déclarées comme FC dans *pointcall_status*) permet d'éviter cela.

Le Z du *net_route_marker* prend également souvent une autre signification, celle de discréditer l'intention annoncée qui n'a pas pu se réaliser, et parfois suivant le point de vue de l'historien. Il y a plusieurs cas qui correspondent à cette situation :

(i) L'intervalle de temps entre les escales ne permet pas la réalisation matérielle des tronçons. C'est le cas par exemple du Turgot, 0021517N, qui quitte Nantes le 4 juillet 1787 pour aller à la Guadeloupe, mais qu'on retrouve à Bordeaux le 12 octobre 1787 déclarant partir pour la Guadeloupe. Sachant qu'à l'époque, un trajet vers les îles antillaises comprend une pause de trois mois aux îles, cela semble impossible : en réalité, depuis Nantes, le navire est passé par Bordeaux avant de filer vers la Guadeloupe. L'analyse intègre ici l'expertise d'historien, car les statistiques que nous avons produites sur la durée moyenne d'un trajet sont extrêmement variables, car dépendantes de la météo et les conditions de navigation de l'époque.

(ii) On démontre que c'est une fausse déclaration (*pointcall_uncertainty* vaut « invalidé ») parce qu'on ne voit pas le navire ressortir de la destination mentionnée, alors que registre des sorties de cette destination annoncée est disponible (Marzagalli et Pfister, 2014), et que le même navire est observé ultérieurement ailleurs.

Quant aux pointcalls succédant au dernier point d'observation d'une suite de documents, ils restent « invérifiables » comme les destinations vers Saint Brieuc et « Côtes de Bretagne » du Fidèle Marianne. Ce qui est également le cas de navires qui n'apparaissent que dans un seul document source, ou de destinations vers l'étranger ou des zones sans registres (comme La Guadeloupe visitée par le Turgot).

Sur la base des règles énoncées, un algorithme (Plumejeaud et Marzagalli, 2022) met à jour la valeur de *net_route_marker* et de *pointcall_uncertain*. avec ses six niveaux, à savoir : observé - confirmé - déclaré - controversé - invérifiable - invalidé. Comme la règle (i) concernant l'intervalle de temps est impossible à évaluer, l'algorithme indique généralement 'A' pour garder le point. Si la valeur de *net_route_marker* diffère de la proposition de l'historien, alors le pointcall est qualifié de « controversé ». Cela signifie qu'il faut se pencher attentivement sur le cas de ce pointcall pour vérifier qui se trompe, de l'algorithme ou l'historien.

2.2. Qualification de l'incertitude des tronçons composant un itinéraire

Afin de construire l'itinéraire, l'algorithme joint les escales de départ aux arrivées ordonnées chronologiquement par navire identifié, en éliminant les points Z sauf ceux controversés. Il reconstruit à la fois les tronçons déclarés dans les sources, reconnus comme tels par la variable « direct » évaluée à vrai, ainsi que les retours implicites déduits (« direct » vaut faux) qui ont dû se faire, comme le trajet de la Tremblade à Marennes de la Suzanne entre le 05 juin 1787 et le 19 juin 1787. Ces tronçons reconstruits représentent 21 % des 70 990 tronçons qualifiés par notre approche. Chaque tronçon est qualifié par une valeur ordonnée d'incertitude (*travel_uncertainty*), déduite à partir de l'incertitude de chaque extrémité (table 3), sachant que les extrémités d'incertitude plus forte sont prioritaires pour décider du niveau d'incertitude d'un tronçon.

TABLE 3. Vue synthétique sur les niveaux de certitude qualifiée de chaque tronçon de navigation.

Travel_ uncertainty	Signification	pointcall_uncertainty sur le départ et/ou l'arrivée	Nombre	Couleur
0	Confirmé	confirmé ou observé, aux deux extrémités	35 355	Vert
-1	Déclaré	déclaré, invérifiable à l'une des extrémités	28 768	Gris
-2	Invalidé	Invalidé à l'une des extrémités	5 544	Rouge
-3	Controversé	controversé à l'une des extrémités	1 323	Orange

Nous pensons avoir qualifié ainsi de façon réfutable et fiable l'incertitude liée à chaque escale dans une grande majorité de cas. Il reste cependant possible que la visualisation des trajectoires des navires dévoile d'autres corrections nécessaires de l'algorithme, ou amènent les historiens à réviser leur jugement. Pour ce type d'exploration analytique, au cas par cas, il nous a semblé pertinent de l'associer à une carte (Ducruet, 2013). Mais nous souhaitons cette cartographie plus agréable à lire que celle de tronçons coupant en droite ligne des terres, comme le trait noir de la figure 1 reliant Saint-Malo à Saint-Brieuc.

3. Calcul de géométries reliant des escales par voie de mer

Pour calculer des géométries de tronçons qui suivent des voies maritimes, nous souhaitons une approche simple et reproductible, utilisable en dehors du contexte de cette recherche historique et qui n'exploite pas de sources historiques comme peut le faire le projet CLIWOC¹. Celui-ci exploite en effet les journaux de bord des capitaines pour retracer les grandes routes maritimes empruntés par les navires de la période couvrant 1750 à 1850. Notre approche est celle d'un calcul automatique d'évitement d'obstacle (*pathfinding*) sous contraintes, et s'inspire des travaux de (Poncet-Montanges, 2013). Si son code déposé sur une drop-box reste inaccessible aujourd'hui, son rapport dessine le fonctionnement dans les grandes lignes de son

¹ <http://www.ucm.es/info/cliwoc/>

algorithme qui, comme nous, n'est pas un algorithme de calcul de plus court chemin sur graphe. En effet, les techniques de recherche opérationnelle sur graphe (Chauveau, 2017) requièrent généralement une discrétisation prédéfinie de l'espace de recherche, alors que l'océan est un espace continu et qu'il n'y a pas de manière directe de le mailler (Dupuy, 2021). Souvent ces approches mobilisent un grand nombre de paramètres (météo, type de navire, danger de capture en mer, etc.) que nous souhaitons écarter pour plus de simplicité.

Le seul paramètre qui nous est nécessaire est l'expression de la contrainte, celle d'un trait de côte que le tronçon ne doit pas intersecter. Nous avons utilisé celui délivré par l'agence nationale américaine des océans et de l'atmosphère², sous la forme d'un shapefile nommé « GSHHS_i_L1.shp » qui est extrait de la base « *Global Self-consistent, Hierarchical, High-resolution Geography Database* » (GSHHG) dans sa version 2.3.7 (Wessel et Smith, 1996). Sa résolution est dite intermédiaire « i » selon l'agence, ce qui signifie généralisée suivant l'algorithme de Douglas-Peucker à un niveau de 80% par deux fois à partir du trait de plus haute résolution disponible, et projeté en WGS84. Ce choix nous permet de voir les contours des petites îles souvent citées dans nos sources comme l'île de Ré, l'île de Bouin, tout en optimisant les ressources nécessaires au calcul.

En effet, notre calcul dessine une ligne entre le début et la fin d'un tronçon, mais la partie du segment intersectant la côte est extraite pour construire un tampon le long de la côte, pour lequel un ensemble de points intermédiaires sont placés à espaces réguliers. Le croquis (fig. 1) expose son fonctionnement dans le cas relativement simple du tracé de la route maritime de Saint-Malo à Saint-Brieuc.

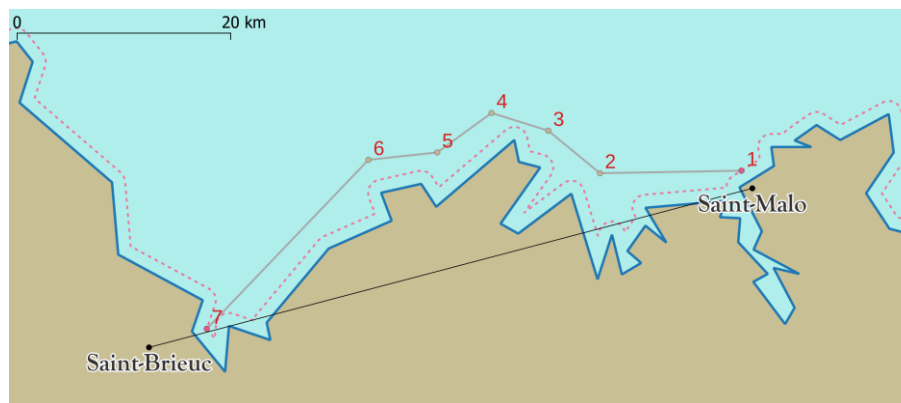


FIGURE 1. Tracé de Saint-Malo à Saint-Brieuc avec ses 7 points intermédiaires.

² Site Web d'accès aux données et métadonnées de la NOAA : <https://www.ngdc.noaa.gov/mgg/shorelines/shorelines.html>

La largeur du tampon (*offset*) ainsi que l'espacement entre les points (*distance_ratio*) sont des paramètres du programme qui s'ajustent automatiquement en fonction des distances à parcourir. Par exemple, sur Saint-Malo à Saint-Brieuc, le tampon est de 5 km et l'intervalle entre les points est de 10 km, mais entre Saint-Brieuc et les Côtes de Bretagne, ils augmentent respectivement à 20 km et 20 km. Ceci correspond à un souci d'économie de ressources, car en passant plus au large, il faut calculer moins d'intersections de côtes et moins de points. De même, les îles de moins de 1 km² sont traversées, comme l'île de Béniguet dans la figure 2. Cette approche correspond aux pratiques de navigation à voile depuis l'antiquité jusqu'à la fin de l'époque moderne : naviguer à vue le plus longtemps possible le long des côtes, avant de s'élancer en droiture sur de vastes étendues océanes seulement si nécessaire, en ligne droite suivant un cap pour rejoindre un amer proche de la destination et à nouveau caboter (Bochaca et Moal, 2019). L'invariant du programme est qu'il n'existe aucune intersection entre la terre et un tronçon reliant deux points intermédiaires ou finaux. Ceci explique que sur la figure 1, le tracé n'est pas réalisé des ports exactement (points noirs), mais depuis leur projection (points roses n° 1 et n°7) sur une ligne située à 1 mile nautique en pointillés rose (*i.e.* 1,852 km, soit la portée maximale d'un boulet de canon de l'époque).



FIGURE 2. Tracé à Saint-Brieuc à Côtes de Bretagne.

L'algorithme fonctionne récursivement afin de dépasser la difficulté liée à la forme fractale des côtes, mais également pour contourner les îlots qui se trouvent très souvent à l'entrée de zones portuaires. La figure 3 illustre comment le programme repousse automatiquement un point en mer lorsqu'un des points intermédiaires tombe malencontreusement sur une île. Le mauvais point P3 est remplacé par son symétrique P'3 par rapport à la côte. Ce tirage aléatoire réussit le plus souvent, mais sinon, l'algorithme reboucle pour déplacer P'3 jusqu'à ce qu'il soit placé en mer.

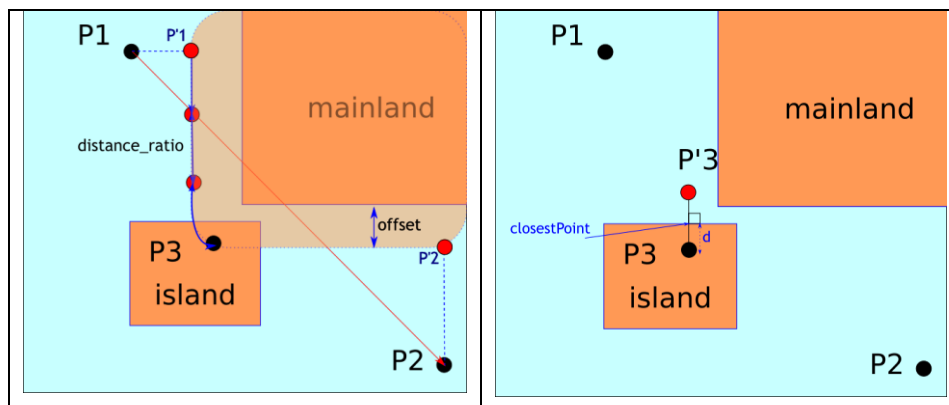


FIGURE 3. Déplacement du point P3 intermédiaire à P'3 pour éviter la terre.

L'itération est ainsi récursive, et permet de calculer un ensemble de routes maritimes qui n'intersectent jamais les parties côtières. Le chemin tracé fait ensuite l'objet d'une simplification, suivant un algorithme de complexité $n \log(n)$ afin de supprimer les boucles formées par le tirage de points hors mer. Le calcul est implémenté sous la forme d'une procédure PL/SQL³ qui renvoie une structure de données comprenant la liste ordonnée des points, mais également les valeurs ajustées des paramètres (tampon et espacement des points), ainsi que la durée du calcul en millisecondes. Sur les 7000 tronçons calculés, la durée moyenne varie entre 80 et 500 secondes, ceci en fonction de la distance à parcourir en ligne droite (trait noir sur les figures 1 et 2), qui conditionne les valeurs de tampon et d'espacement entre les points. Ce résultat nous permet de proposer d'explorer les parcours des navires et des capitaines via une cartographie interactive sur le Web.

4. Géovisualisation interactive de trajectoires de navires ou de capitaines

Cet outil de visualisation, disponible sur le Web⁴, permet de rechercher des itinéraires par le nom ou l'identifiant soit d'un capitaine soit d'un navire. La sélection peut s'affiner en fonction de différents descripteurs, tels que le pavillon, le port d'attache, le tonnage du navire ou les dates de navigation. Lorsqu'un itinéraire a été choisi (ou deux au maximum), l'affichage (fig. 4) comprend deux vues qui se complètent pour faire comprendre le scénario reconstruit pour l'itinéraire avec ses différentes alternatives.

³ https://gitlab.huma-num.fr/portic/porticapi/-/blob/master/sql/portic_routepaths_allfunctions.sql

⁴ <http://shiproutes.portic.fr>

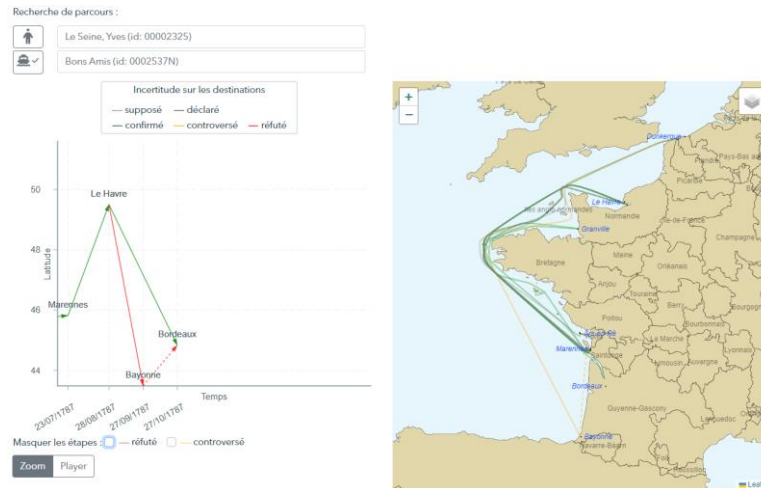


FIGURE 4. Affichage du chronogramme et carte de l'itinéraire du navire *Bons Amis*, (identifiant 0002537N), piloté par *Le Seine, Yves* dans <http://shiproutes.portic.fr>

A gauche de l'écran, un chronogramme déroule l'enchaînement des étapes dans le temps. L'axe vertical dispose les escales en fonction de leur latitude, l'axe horizontal les dispose en fonction de la date de leur réalisation. Il s'inspire de travaux de recherche récents sur les récits de migrations (Dubucs et al., 2010), (Bahoken, 2021) et les manières traditionnelles d'afficher les itinéraires des navires (Buti, 2003). En effet, l'existence de nombreux allers-retours entre les mêmes ports, fréquents à l'époque, nuit à la lisibilité des cartes, et l'interactivité permet de résoudre ce souci (Mac Eachren, 2005). Le chronogramme est interactif : en zoomant dans cette partie, les dates des escales très rapprochées peuvent être distinguées, comme pour les *Bons Amis* piloté par *Le Seine, Yves* (fig. 4), qui est parti de Marennes le 23 juillet 1787 pour le Havre. Ce chronogramme est associé à un outil de déroulement interactif, étape par étape qui déclenche la visualisation des détails de l'itinéraire dans la partie droite de l'écran, dédiée à la carte interactive.

La sémiologie pour les tronçons est commune à ces deux vues : elle vise à rendre compte de l'incertitude calculée des tronçons d'itinéraire (cf. tableau 4). La teinte d'une palette divergente est utilisée pour montrer le niveau de certitude (du vert pour « confirmé » au rouge pour « invalidé ») tandis que les lignes pointillées sont utilisées pour montrer les étapes qui ont été inférées par l'algorithme. Nos choix ont également été guidés par la littérature scientifique sur les signes d'incertitude (Arnaud et Davoine, 2009). Ainsi, la route calculée peut être cartographiée soit dans sa globalité comme sur la figure 4, soit dans le détail comme dans la figure 5. Par exemple, le navire les *Bons Amis* n'a pas réalisé le trajet du Havre à Bayonne, mais seulement un trajet du Havre à Bordeaux, confirmé en vert. Sur chaque

tronçon, le zoom de la carte s'adapte automatiquement, et une bulle au survol de la souris affiche les détails de l'armement du navire sur ce tronçon.



FIGURE 5. Déroulement de l'itinéraire dans la partie cartographique du navire *Bons Amis*, (identifiant 0002537N), dans <http://shiproutes.portic.fr>.

La finalité de l'application est double. D'une part, elle permet pour des passionnés d'histoire maritime de retrouver des personnages ou des navires cités dans d'autres sources, et nous avons un exemple avec ce marin, Pierre-Etienne Millié dont la biographie est commentée dans une revue locale d'histoire maritime (Millié, 2021), qui cite le navire « Comte de Forcalquier » mais sous une autre orthographe – « Comtede Forcalyner » - identifiant 0009151N. D'autre part, elle permet de vérifier visuellement les cas controversés comme l'arrêt à Gibraltar de la Suzanne (tab. 2), mais aussi de vérifier si l'identification de navires est cohérente, comme dans le cas suivant (fig. 6) présentant les trajets de deux navires distincts. Développée comme un client Web riche avec *React*, l'outil exploite l'API en ligne⁵.

⁵ <http://data.portic.fr>

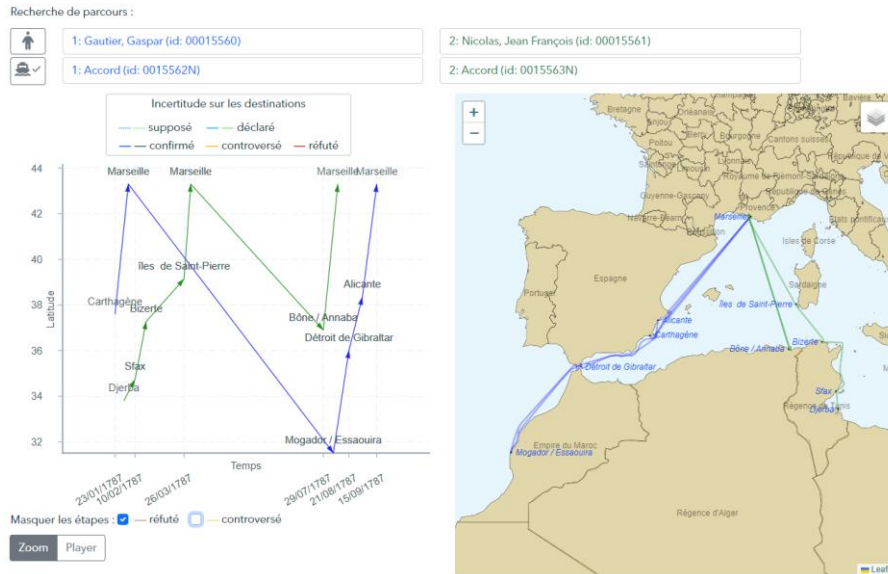


FIGURE 6. Comparaison d'itinéraires pour deux navires distincts nommés Accord.

5. Conclusion et perspectives

Outre la reconstitution des trajectoires logiques de navires communément identifiés dans différents corpus, l'outil proposé permet d'explorer différents scénarios concernant leur itinéraire. L'approche décrite pour qualifier et représenter des trajectoires maritimes incertaines se révèle utile pour l'histoire maritime. Plus largement, notre approche de contrôle du corpus fondée sur la formalisation de règles expertes interroge notre capacité à expliciter totalement la connaissance experte. L'expertise historique a été sollicitée plusieurs fois sur des questions incertaines, et cela soulève une question de confiance entre les chercheurs et leurs pairs ou leur public. À plusieurs reprises, les résultats de l'algorithme se sont avérés différents de ceux des historiens, et notre décision a été de le montrer, notamment avec les tronçons qualifiés de controversés dans la reconstitution des trajectoires de navires. Ceci ouvre la piste à de nouvelles analyses historiques.

L'algorithme proposé pour la fabrication de géométries linéaires contournant les côtes pour rejoindre deux escales peut intéresser un autre public. Le code est open-source et modulaire, accessible sur la forge d'Huma-Num⁶ tout comme les données sur une API en ligne. Cet algorithme devrait être enrichi prochainement afin de visualiser la navigation sur les fleuves, pour aider à comprendre la circulation sur la Seine de Rouen au Havre par exemple.

⁶ <https://gitlab.huma-num.fr/portic/porticapi> pour le modèle de données et l'API, <https://gitlab.huma-num.fr/portic/shiproutes> pour le client Web riche.

Par ailleurs, il peut s'adapter à d'autres bases de données. Le futur musée d'histoire maritime des Sables d'Olonne est intéressé par la représentation interactive de la biographie d'un capitaine sablais du XVIII^{ème} siècle. C'est un travail en cours qui a permis tout d'abord de produire des cartes de navigation illustrant la biographie d'un capitaine reconstruite dans une thèse (Retureau, 2020), et illustrant aussi un article de vulgarisation paru dans *Olonne*, une revue de société savante locale spécialisée sur l'histoire des Sables d'Olonne (Chambaud, 2023).

Bibliographie

- Arnaud, A., Davoine, P.-A., (2009). Approche cartographique et géovisualisation pour la représentation de l'incertitude, in: SAGEO. Paris, France, p. 12.
- Bahoken, F., (2021). La représentation graphique de narrations de mobilités spatiales, aspects formels, In: *Comment cartographier les récits / Mapping Methodologies*, Éditions de l'UMR Territoires, p. 22.
- Bohaca M., Moal L. (2019). *Le Grand Routier de Pierre Garcie dit Ferrande*, Ed. PUR, Rennes, France, 495 p.
- Butil, G., (2003). Cabotage et caboteurs de la France méditerranéenne (XVII^e-XVIII^e s.), in: *Rives nord-méditerranéennes* 13, p. 75-91. <https://doi.org/10.4000/rives.164>
- Chambaud, A/, Plumejeaud-Perreau C. (2023). La mer au XVIII^{ème} siècle : PORTIC et le numérique. *Olonne*, n°263, 2023
- Chauveau E., Jégou P., Prcovic N. (2017). Weather Routing Optimization: A New Shortest Path Algorithm. *29th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2017*, Nov 2017, Boston, United States. hal-01792118
- Dedieu, J.-P., Marzagalli, S., Pourchasse, P. Scheltjens, W., (2011). Navigocorpus, a database for shipping information. A methodological and technical introduction, in: *International Journal of Maritime History* XXIII:2, p. 241-262-
- Dupuy, M., d'Ambrosio C., Liberti L. (2021) Optimal paths on the ocean. (hal-03404586)
- Dubucs, H., Dureau, F., Giroud, M., Imbert, C., André-Poyaud, I., Bahoken, F., (2011). Les circulants entre métropoles européennes à l'épreuve de leurs mobilités. Une lecture temporelle, spatiale et sociale de la pénibilité, in: *articulo* 7 <https://doi.org/10.4000/articulo.1810>
- Ducruet, C., (2013). Histoire maritime et cartes en ligne, du XVI^e au XXI^e siècle. in: *Mappemonde* 109.
- Jeansoulin, R., Papini, O., Prade, H., Schockaert, S. (Eds.), (2010). *Methods for Handling Imperfect Spatial Information, Studies in Fuzziness and Soft Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-14755-5>
- MacEachren, A.M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., Hetzler, E., (2005). Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. in: *Cartography and Geographic Information Science* 32, p. 139–160. <https://doi.org/10.1559/1523040054738936>

- Marzagalli, S., Pfister, C., (2014). Les pratiques administratives des amirautés du XVIII^e siècle: entre spécificité locale et uniformisation. L'exemple de la gestion des congés, in: *Revue d'histoire maritime* 19, p. 259-280.
- Plumejeaud-Perreau, C., Mimouni, M., Bouju, A., Pfister-Langanay, C., Sauzeau, T., Marzagalli, S., 2021. Un gazetier des places portuaires françaises du xviii^e siècle. *revuehn*. <https://doi.org/10.4000/revuehn.1164>
- Plumejeaud-Perreau et Marzagalli, 2022. Plumejeaud-Perreau, Christine, Marzagalli, Silvia 2022, Qualifying and representing confirmed, possible, invalidated, controversial, inferred eighteenth-century ships' itineraries. Dealing with uncertainty with Navigocorpus database. *Atelier 2. [SI pour les Humanités Numériques](#)*, conférence INFORSID, 31 mai-3 Juin 2022, Dijon, France
- Poncet-Montanges, A., 2013, Final report of Mapping European Navy, on line: <https://mappingeuropeannavy.wordpress.com/2013/05/15/final-report/>
- Retureau, H., 2020, *Sociétés littorales, gens de mer et activités maritimes dans un port en mutation : l'exemple des Sables-d'Olonne (1747-1866)*. Thèse de l'Université de Nantes.
- Wessel, P., and W. H. F. Smith (1996), A global, self-consistent, hierarchical, high-resolution shoreline database, *J. Geophys. Res.*, 101(B4), 8741–8743.
- Millié J. (2021) Pierre-Etienne Millié, un marin oublié de Mornac, *de la Seudre à la Charente*, Ed. Société d'Histoire du Canton de Marennes et de ses environs, n°47

Variation d'estimation de température selon les échelles administratives

Didier Josselin¹, Matthieu Vignal¹, Nicolas Viaux¹,
Delphine Blanke², Céline Lacaux²

1. UMR ESPACE, 74, rue Louis Pasteur, 84029 Avignon Cedex, France
didier.josselin@univ-avignon.fr

2. Laboratoire de Mathématiques d'Avignon, 74, rue Louis Pasteur, 84029
Avignon Cedex, France

RÉSUMÉ. Dans cet article, nous explorons l'effet du support spatial sur la mesure des températures de la région Provence Alpes Côte d'Azur selon deux semis de points : les stations météorologiques (au nombre de 142) et les points de la grille des modèles Aladin (soit plus de 6500 points). À partir de ces mesures, il peut être utile de restituer à une échelle agrégée, de la commune au département, l'information météorologique. À l'aide d'une méthode déjà appliquée à des données de diversité écologique et de transactions immobilières, visant à éliminer l'effet du support spatial, nous montrons que le canton semble être, dans la quasi-totalité des cas, l'échelle de restitution qui s'éloigne le plus d'une répartition aléatoire des températures, c'est à dire avec un maximum d'homogénéité par entité et d'hétérogénéité entre entités. Ces résultats sont discutés et la méthode est mise en perspective de généralisation.

ABSTRACT. In this article, nous study the effect of the spatial support on the temperature measurement in the French region of Provence Alpes Côte d'Azur, according to two different sets of data: 142 meteorological channels and more than 6500 points of the Aladin's model grid. Using these measures, it can be helpful to provide relevant meteorological information at a certain aggregated scale, from the municipality to the « department ». Thanks to a method already applied to (bio)diversity and land transactions purpose, aiming at removing the spatial support effect, we show that the scale of the « canton » seems, in most of the cases, the scale the farthest from the random distribution of temperatures, this partitioning being the spatial entities with a maximum of homogeneity in the entities, and the maximum of heterogeneity between entities. The results are discussed and generalized.

MOTS-CLÉS : support spatial, changement d'échelle, agrégation de données, modèles climatiques, stations de mesure météorologique, scalogramme

KEYWORDS: change of spatial support, rescaling, aggregated temperature estimation, climate projection models, weather channel, scalogram

1. Introduction

Projeter le réchauffement climatique (Houghton, 2016) débouche souvent par la production d'une donnée globale (Aspinal, 2012), dont la variation s'applique sur un territoire dont l'étendue est généralement approximative, voire inconnue. Pourtant une des manières de résister au changement climatique est de rechercher et d'exploiter les poches locales de micro-climats (Barry, Blanken, 2016) qui, en fonction de l'altitude, de l'exposition, du couvert végétal ou de tout autre caractéristique, permettront aux êtres de vivant de s'adapter (GIEC, 2022). Ces approches sont développées en agronomie ou en géographie, par exemple sur le suivi des terroirs viticoles (Parker *et al.*, 2020 ; Morin *et al.*, 2021)

Mais le support spatial utilisé pour découper l'espace n'est pas neutre et génère des problèmes méthodologiques (Joly *et al.*, 2007 ; Li, Heap, 2014 ; Josse- lin, 2019). Cela amène notamment aux paradoxes de Simpson (Simpson, 1951) et de Robinson (Robinson, 1950), aux questions de l'erreur (ou inférence) écologique (King, 1997), au problème du changement du support (King *et al.*, 2004), dans le contexte méthodologique plus général des méthodes d' « up-scaling », « downscaling » (Kumar, Agarwal, 2020) et plus globalement de ré-échantillonnage ou « resampling » (MAUP en géographie (Openshaw, 1984 ; Louvet *et al.*, 2015)).

Pour obtenir des données climatiques à l'échelle régionale, nous disposons aujourd'hui d'un ensemble de modèles à l'échelle mondiale, et en particulier européenne. En France, deux types de sources sont accessibles : celles issues des stations météorologiques de Météo-France elles-mêmes, dont la localisation est historique mais qui ne couvre pas les territoires à densité égale, et les représentations maillées issues des modèles climatiques comme ALADIN, qui retranscrivent, sur une grille assez fine, des estimations à partir des relevés et des modèles physiques (cf. Figure 1). Les deux sources de données sont complémentaires et étudier leur usage à travers les échelles constitue deux points de vue croisés intéressants (Martin *et al.*, 2013).

L'objectif de l'article est d'étudier la relation entre des données statistiques climatiques et leurs échelles d'agrégation. Plus précisément, à l'aide d'une méthode de ré-échantillonnage spatial, nous mesurons l'éloignement des valeurs centrales de température au cas aléatoire (homogénéité), afin d'observer si certains découpages administratifs permettent une meilleure discrimination des hétérogénéités locales de température, et, en prolongement, une désagrégation fine et robuste des données climatiques (« downscaling »).

Administrativement, la région Provence-Alpes Côte d'Azur peut être découpée de plusieurs façons, en entités qui s'agrègent (peu ou prou) de la plus petite (la commune, « COM »), à la plus grande (le département, « DEP »), en passant par les cantons (petits agrégats de communes avec un chef-lieu, « CAN »), les structures intercommunales ou Établissements Publics de Coopération

Intercommunale (agrégation de communes à vocation de bassin de vie et de gestion, « EPC ») et les arrondissements (plus grandes entités regroupant des cantons, sous les départements, « ARR »). Chacun de ces découpages croise l'information source des relevés climatiques (ici les températures) à l'échelle des stations ou du modèle maillé (cf. Figure 1) avec des densités d'informations variables selon la source.

Après avoir présenté plus en détails les modèles et les données climatiques utilisées et leur emprise (§ 2), nous exposerons la méthode d'agrégation des températures à travers les échelles (§ 3). Nous expliquerons pourquoi nous réalisons des permutations pour éliminer l'effet du support spatial. La méthode se concentrera sur l'*Écart Relatif normalisé* de température, indiquant la partition spatiale la plus discriminante en termes d'estimation de température, après l'élimination de l'effet du support spatial. Les résultats sont présentés sous forme de « scalogrammes » et discutés (§ 4).

2. Description des données climatiques utilisées

Dans cet article, nous étudions les températures moyennes mensuelles minimales (« NORTNAV ») et maximales (« NORTXAV ») sur une période de 30 ans : 1976-2005. Ce corpus est stable du point de vue des modèles appliqués et nous disposons des valeurs mesurées dans les stations météorologiques et sur la grille Safran du modèle ALADIN de projection climatique. L'approche est largement transposable.

2.1. Stations de mesure

Les données des stations météorologiques proviennent de Météo-France¹. Au total, 142 stations sur l'ensemble de la zone d'étude ont été exploitées. Elles sont localisées dans des milieux aux caractéristiques topographiques et géographiques variées. Elles se caractérisent par la présence de données complètes et précises agrégées sur la période de référence pour les variables de température étudiées.

2.2. Modèles maillés

Le modèle ALADIN-Climat (Aire Limitée Adaptation dynamique Développement InterNational) du Centre National de Recherches Météorologiques (CNRM) est un modèle de circulation global et spectral à aire limitée d'une résolution spatiale initiale de 12 kilomètres. Le modèle ALADIN-Climat propose des données climatiques sur la période de référence, ainsi que sur des simulations du climat futur selon les scénarios établis par le GIEC. Dans le cadre de

1. <https://donneespubliques.meteofrance.fr>

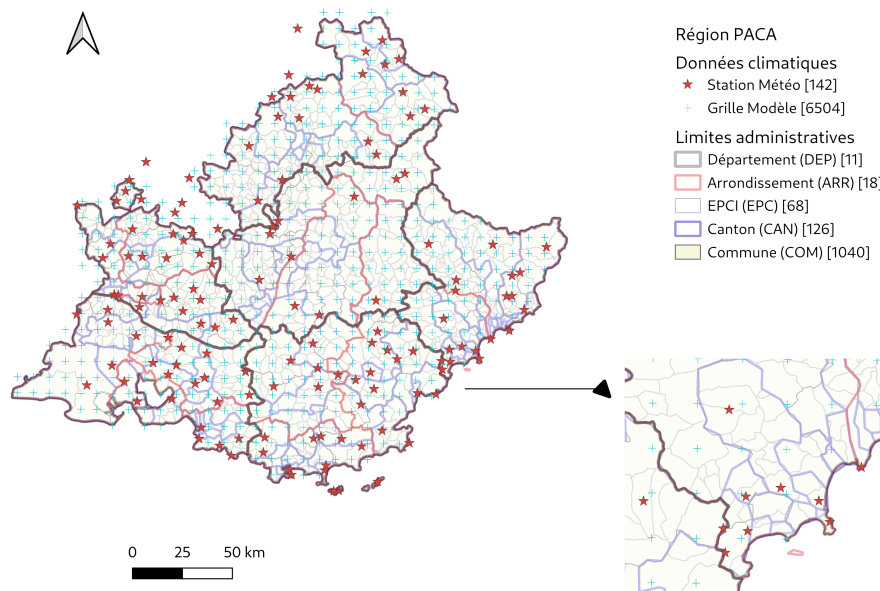


FIGURE 1. Localisation des stations météorologiques et des points maillés du modèle ALADIN-climat en région Provence Alpes Côte d’Azur (Sud de la France)

cette étude, deux sorties de ce modèle sur la période de référence ont été utilisées et la granularité spatiale a été abaissée à des mailles de 8km de côté. Les données ont été exportées sur l’ensemble de la zone d’étude, soit 6504 mailles de la grille du modèle ALADIN-Climat, depuis le site DRIAS-Climat².

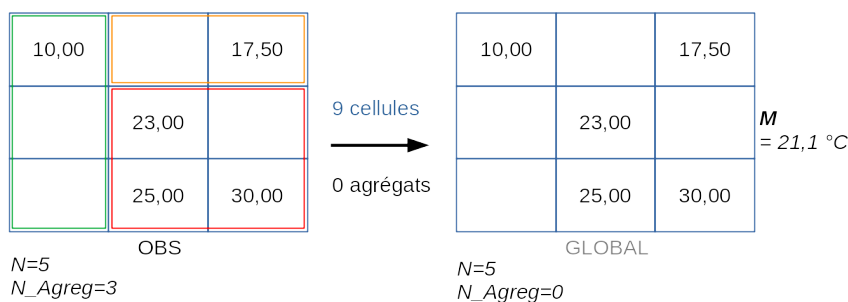
Dans la famille des modèles de climat du CNRM, ALADIN peut être vu comme la version à aire limitée du modèle global ARPEGE-Climat. Selon le CNRM³, l’intérêt principal des modèles régionaux de climat est de limiter le coût de calcul des simulations par rapport à des modèles de climat globaux de même résolution. Ils permettent d’explorer la représentation du climat et du changement climatique à plus fine échelle pour l’étude des forces et des

2. <http://www.drias-climat.fr>

3. <https://www.umr-cnrm.fr/spip.php?article125>

limites des modèles régionaux de climat, l'évaluation et l'amélioration de la représentation du climat moyen, la caractérisation et la compréhension des interactions aérosols-climat, l'étude et l'attribution des tendances passées, la caractérisation de l'évolution climatique au 21ème siècle aux échelles régionales et des incertitudes associées, les impacts régionaux du changement climatique sur les ressources hydrologiques.

Il ne s'agit ici aucunement de remettre en cause les résultats des modèles de projection du GIEC ou des organismes scientifiques associées à travers ces modèles. En revanche, dans un souci de restitution de l'information climatique à des échelles relativement fines pour des partitions administratives différant des sources de mesure, il convient de chercher à identifier l'impact potentiel des découpages sur la transcription de l'information climatique agrégée. C'est l'objet de cette proposition qui, d'une part, produit une information statistique sur les températures mensuelles minimales et maximales sur un période longue (30 ans) et d'autre part, via une procédure de permutations de Monte-Carlo, vise à extraire la valeur de la température, en éliminant l'effet du support spatial, à une échelle donnée.



Pas de ré-échantillonnage, on calcule la valeur sur la distribution de données observées sans partition

FIGURE 2. Agrégation simple des températures observées sans contrainte de découpage administratif

3. Mesures de température à l'échelle de différentes partitions administratives

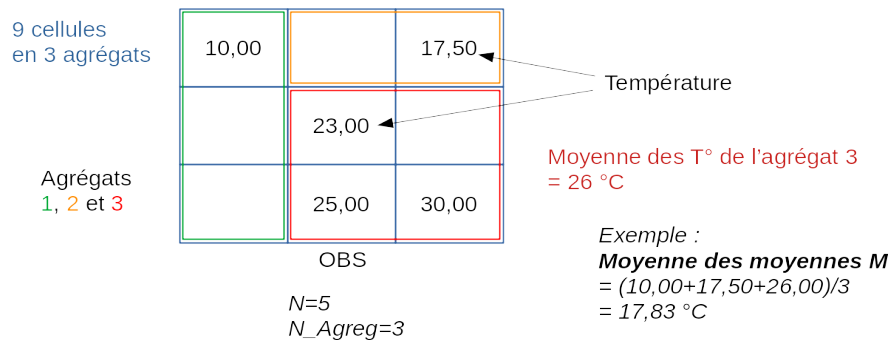
3.1. Principe de calcul : agrégation des mesures

La méthode que nous proposons (déjà publiée dans (Josselin *et al.*, 2008 ; 2009 ; Mahfoud *et al.*, 2007)) est aujourd'hui en phase de généralisation, via son application sur tout type de données géographiques. Pour cet article, elle associe la connaissance de la températures moyenne globale de l'ensemble des

relevés sur la zone, indépendamment de tout découpage territorial (cf Figure 2) et le processus d'agrégation des valeurs de températures mesurées ou produites par le modèle, ou, comme nous le verrons, redistribuées aléatoirement. Le calcul d'agrégation des valeurs de températures moyennes est globalisé sur l'ensemble de la région. Son principe est illustré de façon stylisée dans la figure 3).

Ainsi, l'on procède, pour une partition spatiale donnée, au calcul de :

- la moyenne des températures moyennes des agrégats (exemple : moyenne des températures minimales moyennes des points de la grille ALADIN par canton) ;
- la médiane des températures médianes des agrégats (exemple : médiane des températures médianes maximales des stations météorologiques de la région, par canton).



On calcule l'indicateur sur la distribution de données observées, sur chaque partition

FIGURE 3. Principe d'agrégation des données climatiques : on observe des valeurs de température localisées que l'on agrège à une échelle donnée

3.2. Ré-échantillonnage par permutations de Monte-Carlo

Pour pouvoir évaluer un ensemble de mesures de température sur le territoire, nous proposons de créer, en quelque sorte, un *tube-témoin a-géographique* avec lequel nous comparons nos observations spatialisées. Pour obtenir ce tube témoin, on procède à la suppression de toute auto-corrélation spatiale des températures, en effectuant une série importante de permutations spatiales aléatoires (1000⁴) sur nos deux corpus de données (stations météorologiques et grille du modèle ALADIN). Pour ce faire, on ne modifie pas la localisation

4. 1000 permutations permettent aisément, pour l'estimation du moment d'ordre 1, de converger vers l'espérance mathématique dans le cadre du théorème central limite.

géographique des mesures sur aucune des deux sources (stations météo *vs* modèle ALADIN). On ne fait que les inter-changer sur le même semis de points (cf. Figure 4). La procédure est proche des classiques indices d'autocorrélation spatiale de Moran (Salima, De Bellefon, 2010) (quadrants ou diagrammes de Moran avec distribution aléatoire simulée).

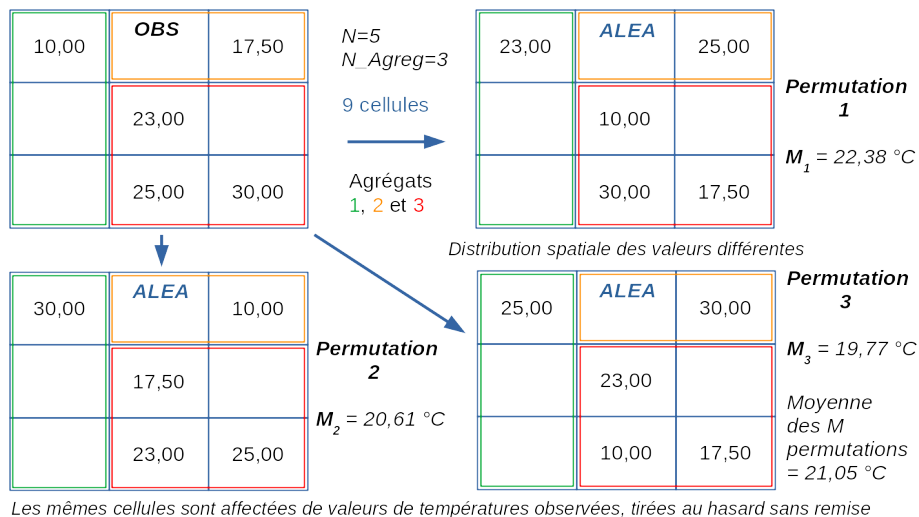


FIGURE 4. Permutations des valeurs de température et re-calcule des valeurs agrégées

3.3. Écart Relatif standardisé

Une fois la procédure de permutation appliquée, nous disposons alors de deux types de données :

- les données observées réellement, sur la grille ALADIN ou dans les stations météorologiques (« OBS »),
- les données permutées (aléatoires), 1000 fois plus nombreuses, qui ont perdu définitivement leur caractère géographique et leur auto-corrélation spatiale (« ALÉA »).

C'est alors que nous appliquons la procédure d'agrégation sur ces deux ensembles de températures, avec la même série de partitions spatiales (c'est à dire ici de découpages administratifs), nous permettant de croiser, dans un nuage de points appelé « scalogramme » les valeurs moyennes ou médianes des températures annuelles ou mensuelles (maximales, minimales) en fonction des échelles.

La comparaison des deux séries de températures (spatialisées *vs* « randomisées ») à travers les échelles nous indique l'écart entre l'observation (« OBS ») et le tube-témoin a-géographique constitué par les données permutées aléatoirement (« ALÉA »). Les partitions spatiales utilisées lors du processus d'agrégation des températures étant strictement similaires, la différence des valeurs agrégées des températures donne une idée, à une échelle donnée, du caractère aléatoire ou au contraire, géographique, de leur répartition. Elle permet également d'éliminer l'effet du support spatial, présent dans les deux types de données traitées (observées ou simulées avec l'aléa). La variation de la valeur observée de la température peut en effet être expliquée par deux grandes composantes : le support spatial lui-même et la valeur intrinsèque mesurée, indépendante du support.

Rapportée à la valeur aléatoire de référence, la différence entre observation et aléa devient l'Écart Relatif normalisé ($ER(\%)$, « ÉCART ») et identifie, selon les échelles, l'éloignement de la distribution des températures (T_{obs}) par rapport à leur répartition aléatoire ($T_{aléa}$).

$$ER(\%) = 100 * \frac{|T_{obs} - T_{aléa}|}{T_{aléa}}.$$

En synthèse, si l'on conserve pour l'observation comme pour la redistribution aléatoire des valeurs de température :

- les mêmes découpages selon les échelles,
- les mêmes méthodes statistiques d'agrégation,
- la même série de données observées (même distribution statistique),
- les mêmes localisations des points de mesures (les permutations n'agissant que sur la localisation des valeurs de température),

cela signifie qu'en soustrayant aux valeurs calculées avec les observations les valeurs « randomisées », c'est à dire redistribuées par permutation, on élimine l'effet du support (celui-ci jouant son rôle dans les deux cas), ne conservant que la valeur de la température réellement portée par la *géographie*. En divisant cette différence par les valeurs du cas aléatoire, on standardise la mesure, la rendant comparable selon les variables considérées (dans notre cas, les types de températures, recensées sur les stations de mesure ou estimées sur les mailles des modèles climatiques).

4. Résultats

Les résultats sont donnés selon les deux types de sources de données, en associant les deux valeurs centrales que sont la moyenne et la médiane des températures minimales et maximales. À des fins de comparaison et compte tenu des formats des données utilisées, nous avons considéré les cumuls à deux granularités temporelles : l'année et le mois, selon les cas.

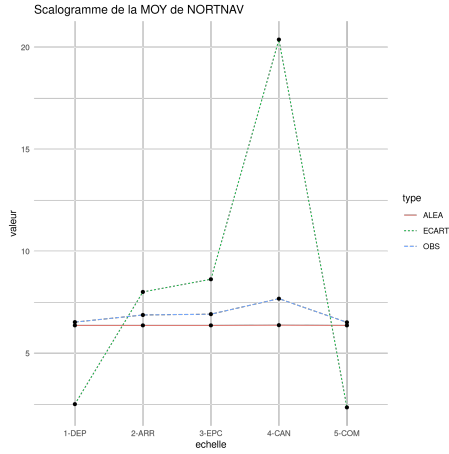


FIGURE 5. Moyenne des températures minimales annuelles (stations) avec Écart Relatif

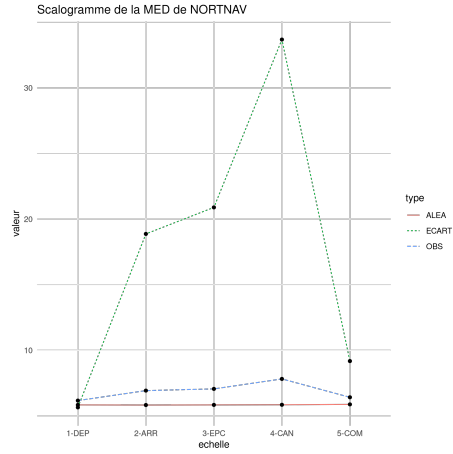


FIGURE 6. Médiane des températures minimales annuelles (stations) avec Écart Relatif

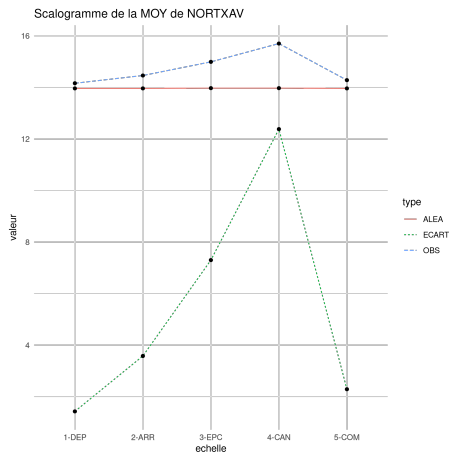


FIGURE 7. Moyenne des températures maximales annuelles (stations) avec Écart Relatif

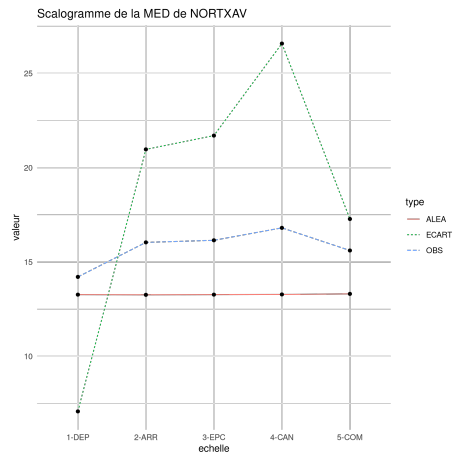


FIGURE 8. Médiane des températures maximales annuelles (stations) avec Écart Relatif

4.1. Résultats des mesures sur les stations météorologiques

Ce qui est intéressant sur les figures 5, 6, 7 et 8 est le pic commun avec les deux métriques que sont la moyenne et la médiane, même si les valeurs d'Écart

Relatif (courbe verte) varie de 12 à 35% environ. Quelle que soit la donnée de température considérée (minimale ou maximale), le canton est l'échelle pour laquelle la restitution de la température est la plus contrastée spatialement, car éloignée d'une répartition aléatoire. Les autres échelles ne montrent pas de particularités très nettes. Cela doit être bien entendu considéré en regard du semis de stations météorologiques.

4.2. Résultats des mesures à partir de la grille du modèle ALADIN-Climat

En passant à un semis de mesure régulier sur la grille ALADIN, pour les températures minimales mensuelles de janvier (cf. Figures 9 et 10) ou d'octobre (cf. Figures 11 et 12), nous obtenons des formes de la courbe d'Écart Relatif relativement similaires aux précédentes (excepté pour la moyenne des températures minimales de janvier, avec un point saillant au niveau des EPCI).

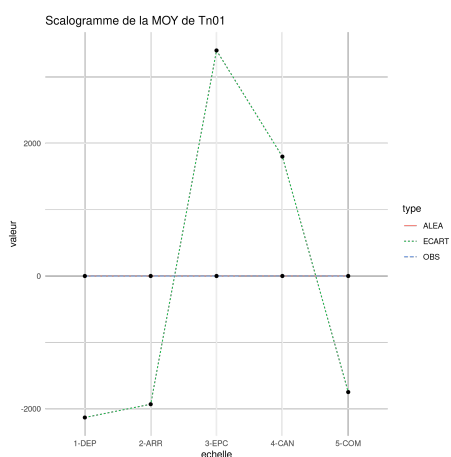


FIGURE 9. Moyenne des températures minimales mensuelles de janvier (ALADIN) avec Écart Relatif

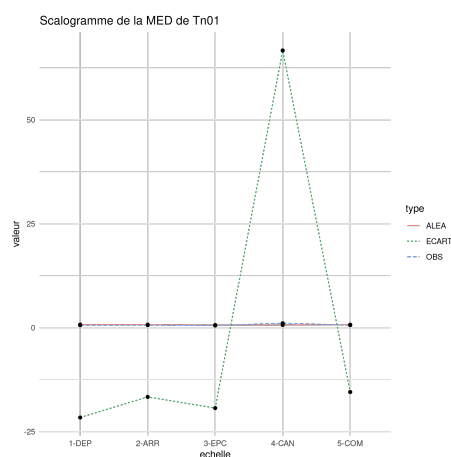


FIGURE 10. Médiane des températures minimales mensuelles de janvier (ALADIN) avec Écart Relatif

En revanche, on note une chute des pics au niveau de l'échelle du canton, pour les températures maximales mensuelles d'avril (cf. Figures 13 et 14) et de juillet (cf. Figures 15 et 16). Ces différences montrent que pour certaines séries de valeurs, les niveaux d'échelle sont peu discriminants pour restituer les températures en question. Elles corroborent la généralisation et l'uniformisation de l'augmentation des températures et ce, quel que soit le découpage considéré.

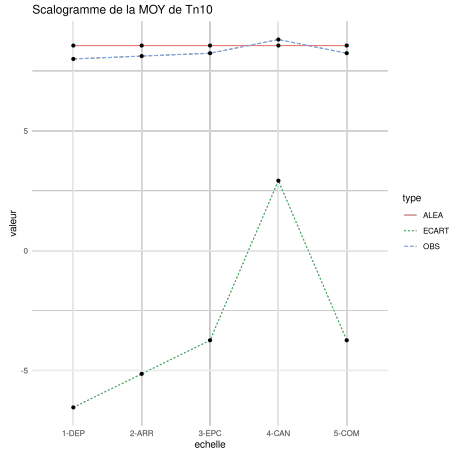


FIGURE 11. Moyenne des températures minimales mensuelles d'octobre (ALADIN) avec Écart Relatif

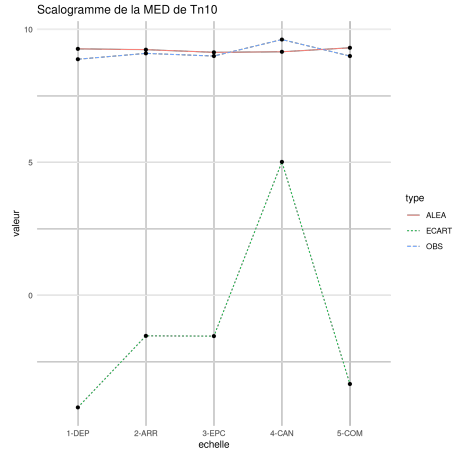


FIGURE 12. Médiane des températures minimales mensuelles d'octobre (ALADIN) avec Écart Relatif

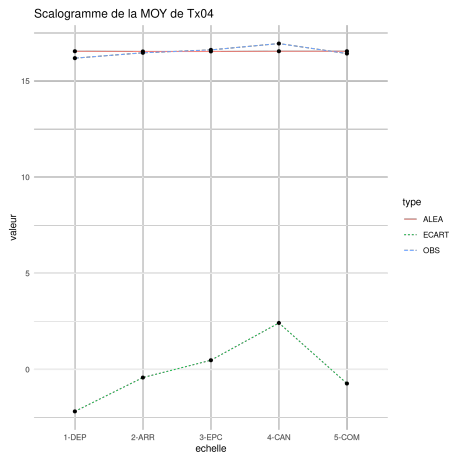


FIGURE 13. Moyenne des températures maximales mensuelles d'avril (ALADIN) avec Écart Relatif

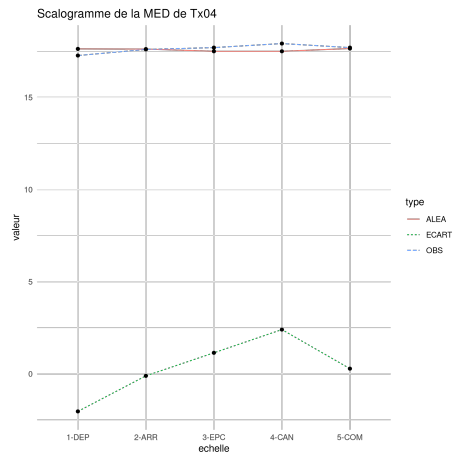


FIGURE 14. Médiane des températures maximales mensuelles d'avril (ALADIN) avec Écart Relatif

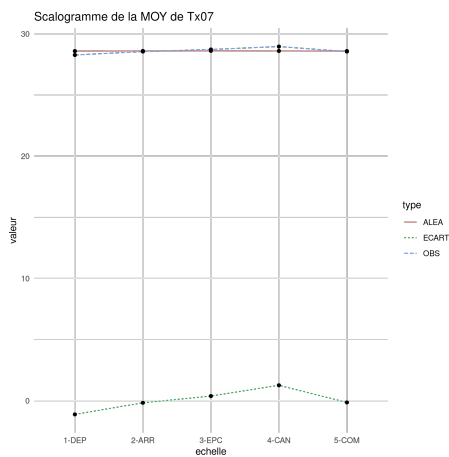


FIGURE 15. Moyenne des températures maximales mensuelles de juillet (ALADIN) avec Écart Relatif

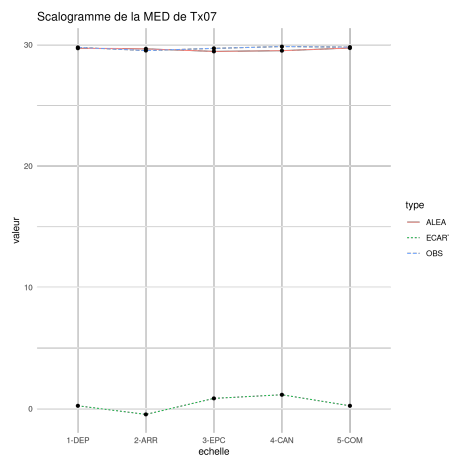


FIGURE 16. Médiane des températures maximales mensuelles de juillet (ALADIN) avec Écart Relatif

Conclusion et discussion

On constate d'une part une certaine convergence dans les pics d'Écart Relatif à l'échelle du canton. Cela nous renseigne sur le fait, qu'en région française Provence Alpes Côte d'Azur, avec des sources de données différentes, il semble que l'échelle dont les entités concentrent le plus de valeurs de température homogène (et différentes entre les entités) soit le canton. Il n'y a *a priori* pas de raison géographique directe, mais, parmi les découpages proposés, c'est celui qui met le mieux en évidence les différences locales, avec une augmentation sensible des valeurs centrales agrégées. Une hypothèse réside dans la possible relation morphologique entre cantons et bassins physiques. D'autres résultats récents (sous presse) sur les mutations immobilières montrent que c'est la commune qui cette fois est la plus discriminante statistiquement.

Au-delà de ces résultats, c'est bien l'Écart Relatif qui est un indicateur pertinent pour savoir quelle est l'échelle la plus appropriée que l'on doit considérer selon notre objectif :

- quand ER est proche de 0, il n'y a aucun intérêt à utiliser un partitionnement spatial et agréger l'information, la valeur moyenne globale suffit,
- quand ER a une valeur significativement positive, cela suggère que pour cette échelle, les valeurs ne sont pas aléatoires et que les fournir à l'entité du découpage en question est utile, voire nécessaire.

La méthodologie proposée, appliquée aujourd'hui à plusieurs types de données, doit être généralisée géographiquement, en particulier pour la désagrégation des données climatologiques. Ces travaux sont actuellement mis en relation avec les approches, citées en introduction, de « downscaling » basées sur des méthodes statistiques régressives à échelle fine, sensibles notamment à la densité et à la répartition des semis de capteurs au regard de la taille de vallées des territoires étudiés (cas notamment de la région Provence Alpes Côte d'Azur).

Bibliographie

- Aspinal R. (Ed.). (2012). *Geography of climate change*. Routedledge, Taylon Fancis.
- Barry R. G., Blanken P. D. (Eds.). (2016). *Microclimate and local climate*. Cambridge University Press.
- GIEC (Ed.). (2022). *Changements climatiques 2013 les éléments scientifiques - contribution du groupe de travail i au cinquième rapport d'évaluation du groupe d'experts intergouvernemental sur l'évolution du climat - résumé à l'intention des décideurs*. OMM, PNUE.
- Houghton J. (Ed.). (2016). *Le réchauffement climatique. un état des lieux complet*. Planète en jeu, De Boeck.
- Joly D., Brossard T., Cardot H., J., Hilal M., Wavreski P. (2007). Densité des points de mesure, types et limites des modèles d'interpolation. In *Journées de Climatologie, Climat et société : Mesures et modèles*, p. 49-65. Grenoble, 22-23 mars 2007.
- Josselin D. (2019). Impact of the Scale on Several Metrics Used in Geographical Object-Based Image Analysis: Does GEOBIA Mitigate the Modifiable Areal Unit Problem (MAUP)? *ISPRS International Journal of Geo-Information*, vol. 8, n° 3, p. 156.
- Josselin D., Mahfoud I., Fady B. (2008). Impact of a change of support on the assessment of biodiversity with shannon entropy. In *Spatial Data Handling, SDH'2008*, p. 109-131. Montpellier, June, 23-25.
- Josselin D., Mahfoud I., Fady B. (2009). Analyse exploratoire des effets de support spatial et de robustesse statistique sur la fiabilité de la mesure de la (bio)diversité. *Photo-interprétation / European Journal of Applied Remote Sensing*, vol. 45, p. 3-11;35-41.
- King G. (1997). *A solution to the ecological inference problem. reconstructing individual behaviour from aggregate data*. Princeton University Press.
- King G., Rosen O., Tanner A. M. (Eds.). (2004). *Ecological inference. new methodological strategies*. Cambridge University Press.
- Kumar A. P. V., Agarwal S. (Eds.). (2020). *Statistical downscaling of climatic parameters. forecasting regional scale rainfalland temperature using statistical downscaling model under vartious climatic scenarios*. Lambert Academic Publisher.
- Li J., Heap A. D. (2014). Spatial interpolation methods applied in the environmental sciences : A review. *Environmental Modeling Software*, vol. 53, p. 173-189.

- Louvet R., Josselin D., Genre-Grandpierre C., Aryal J. (2015). Impact du changement d'échelle sur l'étude des causes des feux de forêts du sud-est de la France. In *Actes de la 11ème conférence internationale annuelle Spatial Analysis and GEomatics, SAGEO 2015, Hammamet, Tunisia, November 23th, 2015.*, p. 254-268. Consulté sur <http://ceur-ws.org/Vol-1535/paper-18.pdf>
- Mahfoud I., Josselin D., Fady B. (2007). Sensibilité des indices de diversité à l'agrégation. *Revue Internationale de Géomatique*, vol. 3-4, p. 293-308.
- Martin N., Carrega P., Adnes C. (2013). Downscaling à fine résolution spatiale des températures actuelles et futures par modélisation statistique des sorties ALADIN-Climat sur les Alpes-Martimes (France). *Climatologie*, vol. 10, p. 51-72.
- Morin G., Le Roux R., Lemasle P.-G., Quénoil H. (2021). Mapping bioclimatic indices by downscaling modis land surface temperature: Case study of the saint-emilion area. *Remote Sensing*, vol. 13, 4.
- Openshaw S. (1984). *The modifiable areal unit problem*. Norwich: Geo Books, CAT-MOG 38.
- Parker A. K., Cortázar-Atauri I. García de, Trought . D. A., Michael C.T., Agnew R., Andrew S., Leeuwen C. van. (2020). Adaptation to climate change by determining grapevine cultivar differences using temperature-based phenology models. *Oeno-One*, vol. 4, p. 955-974.
- Robinson W. (1950). Ecological correlations and the behaviour of individuals. *American Sociological Review*, vol. 15, p. 351-357.
- Salima B. A., De Bellefon M.-P. (2010). *Indices d'autocorrélation spatiale*. Rapport technique. INSEE. Consulté sur <https://www.insee.fr/fr/statistiques/fichier/3635442/imet131-g-chapitre-3.pdf>
- Simpson E. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society - Series B (Methodological)*, vol. 13-2, p. 238-241.

Méthodologie d'identification du domaine public par traitements géomatiques en creux du cadastre pour une analyse spatio-temporelle

Thibault Lecourt^{1,2}, Laure Casanova Enault^{1,3},
Didier Josselin^{1,4}

1. Avignon Université, UMR CNRS 7300 ESPACE, France

2. thibault.lecourt@univ-avignon.fr

3. laure.casanova@univ-avignon.fr

4. didier.josselin@univ-avignon.fr

RÉSUMÉ. Le domaine public en France correspond à des propriétés inaliénables de personnes publiques, hors marchés fonciers et immobiliers, au service d'un intérêt public. N'étant généralement pas cadastré, il est difficile de l'identifier et d'en étudier l'évolution à travers les données cadastrales. Afin de remédier à cette méconnaissance, on propose une méthodologie d'identification du domaine non-cadastré en creux du domaine cadastré. Cette opération de différence spatiale est rendue possible par une série de traitements géomatiques visant à fiabiliser et optimiser le processus. La méthode s'appuie sur le SGBD PostgreSQL, est expérimentée sur le département d'Ille-et-Vilaine (35) et permet d'observer l'évolution du domaine public entre 2009 et 2021.

ABSTRACT. In France, public bodies can own land from the public domain which is inalienable and excluded from the land and real estate markets, serving a public interest. It usually does not appear in any land registry and therefore it can't be identified and its evolutions can't be approached by the analysis of land data. In order to solve this lack of knowledge, we propose a methodology for identifying the non-registered domain in counter-relief from the registered domain. This spatial difference operation is made possible by a geomatic process making the workflow more reliable and optimized. The method is performed by PostgreSQL DBMS, tested on the Ille-et-Vilaine department (France) and allows to observe the evolution of the public domain between 2009 and 2021.

MOTS-CLÉS : foncier, domaine public, non-cadastré, PostGIS, différence spatiale

KEYWORDS: land, public domain, non-registered land, PostGIS, spatial difference

1. Introduction

Cet article propose de contribuer à la connaissance du foncier public en France en en révélant une dimension méconnue : le domaine public. Ce régime de propriété propre aux personnes publiques concerne des biens qui, contrairement à ceux de leur domaine privé, leur sont inaliénables et imprescriptibles - ils ne peuvent être cédés ni appropriés - et doivent revêtir un caractère d'utilité publique (articles L. 2111-1 et L. 2111-2 (CG3P, 2007)). Le domaine public n'est pas assujéti à la taxe foncière, et donc non cadastré. Il est d'ailleurs souvent nommé abusivement « domaine non-cadastré », bien que certains pans du domaine privé ne soient pas cadastrés non plus comme les autoroutes, certains cours d'eau ou zones de montagne. Ces cas restent rares et on peut résumer la correspondance entre propriété, domaine et cadastre ainsi :

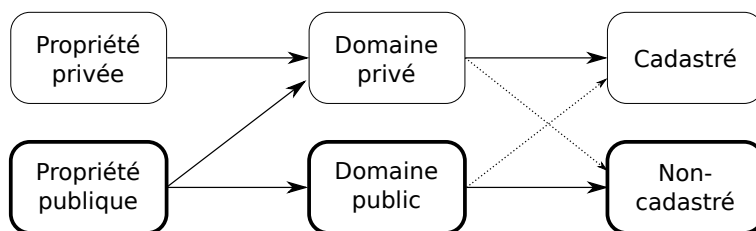


FIGURE 1. Schéma récapitulatif des correspondances entre propriété, domaine et situation cadastrale. En gras, ce que la méthode cherche à identifier.

Le Centre National de l'Information Géolocalisée (CNIG) estime le domaine public à 3% de la surface du territoire métropolitain, mais avec une variance telle que ce taux atteint 38 % à Paris. Cette estimation chiffrée reste non géolocalisée : « la parcellisation généralisée du domaine non cadastré ne revêt pas d'intérêt déterminant au regard de la mission cadastrale [...] et des moyens dont elle dispose » (CNIG, 2016). L'enjeu de connaître l'état et l'évolution du domaine public est pourtant primordial aujourd'hui. La crise économique de 2008 a renforcé la nécessité pour tous les échelons de l'administration publique de prendre des mesures d'austérité budgétaire (Peck, 2012) menant notamment à la vente du foncier public, le patrimoine des acteurs publics représentant une manne financière d'autant plus importante que les prix ont atteint des niveaux historiques. La cession des biens publics est par ailleurs vivement encouragée par l'Etat en France afin de favoriser la production de logements (Dufлот, 2013 ; Rebsamen, 2021). Ce contexte propice à une privatisation du foncier public a déjà été observé au Canada (Whiteside, 2020), au Royaume-Uni (Christophers, 2018 ; Artioli, 2021), en Suède (Olsson, 2018) et en France (Adisson, Artioli, 2020), et il concerne les deux domaines. En effet, malgré son caractère inaliénable, le domaine public peut être déclassé et certains cabinets de conseils se sont faits une spécialité d'accompagner les acteurs publics dans cette démarche

afin de « valoriser » leur patrimoine - c'est-à-dire de le vendre (Bonneau *et al.*, 2020). Il est donc crucial d'évaluer la potentielle cession de domaine public qui échapperait aux analyses des données cadastrales fondées uniquement sur le domaine privé, les bases de données existantes produites par le CEREMA (Fichiers Fonciers, Demande de Valeurs Foncières, Référentiel du Foncier Public) ne portant que sur le domaine cadastré. Dans une optique d'aide à la décision, la méthode développée ici remplit trois objectifs. Le premier est de révéler le domaine non-cadastré en creux du cadastre (Figure 2). Le deuxième est d'identifier le « domaine public d'intérêt », c'est-à-dire le domaine public mobilisable pour des opérations d'aménagement ou pour être vendu, dégagé de la voirie et d'espaces résiduels inexploitable. Le troisième enjeu consiste à mettre en lumière des dynamiques de transformation de ce domaine public d'intérêt, pour permettre aux acteurs publics d'en maîtriser le développement.



FIGURE 2. Principe général de la méthode d'identification du domaine public

On expérimente la méthode sur le territoire de Rennes, connu pour sa forte maîtrise foncière (Dormois, 2006 ; Vergriete, 2013 ; Serrano, Demazière, 2016). Cette réputation se base sur l'action de la Métropole et de la Ville sur les marchés fonciers, donc sur le domaine privé uniquement : on cherche ici à la vérifier concernant la gestion du domaine public. On retient l'échelle du département de l'Ille-et-Vilaine (35) afin d'observer une diversité de situations géographiques. Pour définir le domaine privé, on utilise les données cadastrales de la base MAJIC retraitées par le CEREMA, les *Fichiers Fonciers*. Ces données couvrant la période de 2009 à 2021, on propose une comparaison diachronique entre le plus récent et le plus ancien millésimes à travers des traitements géomatiques, principalement la différence spatiale¹. On explique d'abord comment préparer et fiabiliser les données cadastrales (2), et on aborde ensuite l'enjeu de reconstitution des surfaces dédiées aux infrastructures routières (3). C'est alors que peut être mené le processus de différence spatiale pour en déduire le domaine public (4), dont les résultats sont ensuite exposés (5) et enfin discutés pour guider leur interprétation (6).

1. Afin de rendre possible et accessible cette opération coûteuse en temps de calcul, on a recours à un serveur libre et *open-source*, PostgreSQL 13 et sa cartouche spatiale PostGIS, avec des traitements optimisés pour être réalisés sur un poste de travail récent et non un serveur de production (processeur 2,90GHz 2x8 cœurs, 32GB de mémoire vive).

2. Préparation des données cadastrales

Pour lire le domaine public en creux du domaine privé, il est d'abord nécessaire de disposer d'informations fiables, complètes et correctement préparées sur le domaine privé, qu'on connaît à travers le cadastre.

2.1. Fiabilisation de la géométrie surfacique

Les Fichiers Fonciers offrent deux géométries pour localiser les parcelles cadastrales : l'une ponctuelle (*Point*) correspondant au localisant de la parcelle ; l'autre surfacique (*Multi-Polygon*), correspondant à son emprise réelle. Néanmoins, cette dernière n'a pas toujours été vectorisée, et dans ce cas, la géométrie surfacique ne correspond pas à la forme réelle de la parcelle mais à un carré dont la surface est égale à la valeur *surface* saisie dans la base de données. Ce cas pose problème dans notre démarche car on ne peut pas soustraire cette forme géométrique théorique à la surface départementale pour en déduire le domaine non-cadastré : ce dernier serait largement faussé (Figure 3 à gauche). Il est possible de repérer ces parcelles en étudiant leur forme et en comparant l'aire de la surface de l'entité à l'aire du rectangle minimal l'enveloppant : si les aires sont parfaitement égales et que deux côtés non-opposés ont la même longueur, c'est qu'il s'agit d'un carré parfait.

Listing 1 – Code SQL pour repérer les parcelles ayant une géométrie surfacique théorique carrée

```
SELECT * FROM parcelles_cadastrales_2009
WHERE ST_Area(ST_OrientedEnvelope(geom)) = ST_Area(geom)
AND ST_Distance(ST_PointN(ST_ExteriorRing(ST_OrientedEnvelope(geom)),1),
  ST_PointN(ST_ExteriorRing(ST_OrientedEnvelope(geom)),2)) =
  ST_Distance(ST_PointN(ST_ExteriorRing(ST_OrientedEnvelope(geom)),3),
  ST_PointN(ST_ExteriorRing(ST_OrientedEnvelope(geom)),4))
```

Ces approximations sont fréquentes en 2009 : dans le département, elles concernent 932 parcelles (128,01 ha, soit 0,02 % de la surface départementale cadastrée) ; mais aucune en 2021. Par ailleurs, une autre imprécision est à prendre en compte : certaines parcelles n'ont aucune géométrie, seule une surface est saisie dans l'attribut numérique correspondant ; 1 042 parcelles sont concernées en 2009 (602,36 ha, soit 0,09 % de la surface départementale cadastrée), contre 338 en 2021 (115,35 ha, soit 0,02 % de la surface départementale cadastrée). On remarque une nette amélioration de la qualité des données entre les deux dates. Afin de remédier à l'incertitude propre au millésime 2009, on propose de s'appuyer sur cette amélioration continue des données cadastrales. On met en place une logique itérative recherchant, pour chaque parcelle du département en 2009, la dernière géométrie disponible dans les différents millésimes de la base de données. Cette opération est possible car une parcelle ne subissant aucune mutation morphologique (fusion ou division) conserve un même identifiant d'une année à l'autre. On retrouve ainsi pour chaque parcelle sa géométrie surfacique la plus récemment saisie. Cette méthode élimine plus

de 98 % (- 126,02 ha) des géométries surfaciques théoriques carrées en 2009 (Figure 3 à droite), seules 15 parcelles la conservent encore (1,99 ha). Elle permet aussi de réduire de 10 % (- 61,26 ha) la superficie de parcelles sans géométrie surfacique en 2009, pour la ramener à 604 parcelles (541,10 ha, soit 0,08 % de la surface départementale cadastrée).

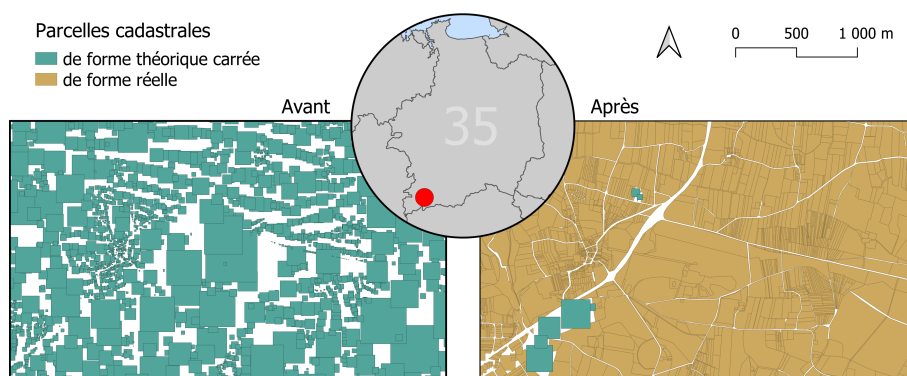


FIGURE 3. Illustration de la méthode de fiabilisation morphologique des parcelles cadastrales en 2009 (Sources : DGFIP-DGALN-CEREMA Fichiers Fonciers 2009-2021. Réalisation : LECOURT, 2023)

2.2. Correction et simplification des géométries

La couche des parcelles ainsi fiabilisée doit ensuite être simplifiée afin de préparer l'opération de différence spatiale, car plus le nombre d'entités et de sommets est grand, plus l'opération de différence est lourde (Obe, Hsu, 2021) ; nous cherchons donc à les réduire par une succession de fonctions complémentaires. Il s'agit d'éclater les éventuels multi-polygones correspondant à une seule parcelle, puis d'extraire les géométries de type surfacique pour écarter les points et lignes renvoyés par l'opération. On réduit ensuite la précision de chaque entité en calant les sommets sur une grille géométrique (*ST_SnapToGrid()*) avant de simplifier leur géométrie avec une tolérance de 0,01 mètre. Une fois cette couche parcellaire simplifiée obtenue, on doit encore corriger certaines erreurs présentes dans la base de données cadastrales : les limites de parcelles limitrophes ne sont parfois pas parfaitement alignées entre elles si bien que des espaces résiduels peuvent apparaître entre deux parcelles. En plus de fausser la détection du domaine non-cadastré, ces écarts rendent plus lourds les traitements de différence spatiale car ils complexifient les objets à soustraire. Pour les éliminer, on réalise un tampon de 2 mètres sur les parcelles. On unit ensuite la couche ainsi obtenue en une seule entité afin d'optimiser l'opération

prochaine de différence. Au sein de cette entité unique du parcellaire, on observe des « trous » et des bandes de superficie variable, pouvant correspondre à du domaine public d'intérêt, à de la voirie non-cadastrée, ou du bruit généré par la méthode de correction des géométries. A l'instar d'autres travaux de détection des « gisements fonciers », on considère les espaces inférieurs à 200m² comme non exploitables pour l'aménagement (Casanova Enault *et al.*, 2021) et on choisit de les intégrer au parcellaire afin de concentrer notre étude sur le domaine public d'intérêt. La fonction de détection et suppression des trous dans un polygone en fonction de leur surface n'existe pas nativement dans PostGIS 3 mais peut être créée². On appelle la couche produite *parcelles_surf*.

3. Reconstitution des surfaces de voirie

Pour identifier le domaine public d'intérêt, il convient d'isoler la voirie, qui ne répond pas à une règle stable au regard de sa situation cadastrale. Cette incertitude est problématique car la création de voirie entre 2009 et 2021 est importante dans un contexte de péri-urbanisation rennaise : on doit être capable de distinguer l'évolution du domaine public d'intérêt de l'évolution de la voirie.

3.1. Estimation des largeurs

On établit une méthode de reconstitution des surfaces dédiées à la voirie en s'appuyant sur la BD TOPO de l'IGN, disponible pour les années 2009 et 2021. Elle comprend une couche *routes* sous formes de lignes pour lesquelles une information de largeur est disponible. Néanmoins, comme en témoigne le Centre d'Études sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques, la donnée n'est pas renseignée de façon homogène et semble souvent incohérente (CERTU, 2013). On s'appuie sur une méthode alternative que l'organisme propose afin de reconstituer une largeur moyenne des infrastructures routières prenant en compte le rang de la route (communale, intercommunale, départementale...), sa nature (bretelle, autoroute...), le nombre de voies et les accotements (bandes dérasées et bermes). En complément, on choisit de multiplier par 1,33 la largeur totale ainsi estimée par le CERTU afin de prendre en compte plus largement les bas-côtés de la voirie correspondant souvent à des espaces délaissés, comme les bords de voies rapides ou des espaces de circulation pour piétons. On est alors en mesure de créer un tampon adapté aux caractéristiques de différentes infrastructures sur les lignes de voirie. De fait, ce tampon vient souvent se superposer à celui effectué sur les parcelles : on parvient ainsi à éliminer l'essentiel des décalages entre la voirie et le parcellaire dans le cadastre, et à concentrer notre méthode sur le domaine public d'intérêt.

2. Son principe nous a été suggéré par Martin Davis et Cyril Mikhalchenko (2022) sur ce fil de discussion en accès libre : <https://gis.stackexchange.com/questions/431664/deleting-small-holes-in-polygons-specifying-the-size-with-postgis>

3.2. Correction des espaces résiduels

Quelques espaces dédiés à la voirie ne sont pas pris en compte par la méthode, en particulier les terrepleins centraux de rond-points ou entre deux voies. Ils prennent la forme d'entités surfaciques de forme longitudinale et/ou de petite surface, imbriqués entre des entités de la couche *routes* jusqu'ici produite. On propose de les détecter par une combinaison d'indicateurs de taille, d'élongation et de compacité (Guérois, 2003 ; Maignant, 2005) : l'aire A , exprimée en m^2 ; la largeur l , déterminée par l'aire de l'entité après un tampon négatif de -3 m ; l'indice de forme i , dont la valeur augmente avec l'étirement de la forme ; l'indice de circularité de Miller M , dont la valeur 1 correspond à un cercle et s'approche de 0 quand la forme s'allonge.

$$i = \frac{Aire}{Périmètre} \qquad M = \frac{4 \cdot \pi \cdot Aire}{Périmètre^2}$$

La combinaison optimale de ces quatre indicateurs pour sélectionner les entités surfaciques est déterminée empiriquement par tests successifs afin d'intégrer un maximum d'entités résiduelles dans la voirie tout en veillant à ne pas inclure de larges portions qui correspondraient à du domaine public d'intérêt. On sélectionne ces entités sur la base d'une des conditions suivantes (une seule suffit) : l'aire A est inférieure à $200m^2$; la largeur l est inférieure à 3m ; l'indice i est inférieur ou égal à 2 et l'indice M à 0,3 ; l'indice i est inférieur ou égal à 5 et l'indice M à 0,1. On ajoute ainsi, pour le millésime 2009, 12,58 ha de voirie pour atteindre un total de 30 118,77 ha de surface dédiée aux infrastructures routières dans le département. On appelle cette couche *routes_surf*.

4. Un jeu de différences

La méthode développée ici fonctionne comme la technique de l'estampe en art : il s'agit de creuser le domaine privé afin de révéler, par différence, le domaine public. Cette opération implique quelques ajustements, et permet finalement de comparer les résultats entre deux dates. Les différentes étapes sont résumées dans la Figure 4 et mesurées dans la Table 1.

4.1. Optimisation des différences spatiales

Le support de base de l'opération est une entité surfacique pleine correspondant à l'ensemble du département considéré, qu'on appelle *departement_surf*. On a ensuite deux fois recours à la fonction PostGIS de différence spatiale : il est nécessaire de commencer par soustraire à la couche *departement_surf* la couche des routes *routes_surf*, puis celle des parcelles *parcelles_surf*. L'inverse rend l'opération impossible en coût de traitement avec les moyens à notre disposition car la forme du parcellaire est bien plus complexe et volumineuse que celle des routes. D'après notre expérience, soustraire une forme complexe (les parcelles) à une forme relativement simple (le département moins les routes)

est techniquement plus réalisable que l'inverse. Il convient pour ce faire de systématiquement créer un index spatial sur les couches utilisées, y compris les couches intermédiaires. Par ailleurs, les couches utilisées doivent toujours être unies en une seule entité afin d'optimiser les opérations de différence spatiale, qui dans le cas contraire serait répétées entité après entité.

4.2. *Espaces résiduels et erreurs d'affectation*

Le résultat de ces opérations génère des espaces résiduels, provoqués par les imprécisions dans la vectorisation des parcelles, par les traitements de préparation ci-dessus (tampons, simplifications, corrections de géométrie...) et par l'incertitude dans l'estimation des surfaces de voirie. Comme pour la reconstitution des routes, ces espaces résiduels prennent la forme d'entités surfaciques soit de forme longitudinale et/ou de petite taille, principalement le long des routes. Il convient de les exclure du calcul du domaine public d'intérêt et de les considérer comme faisant partie de la voirie. On les détecte avec la même méthode que dans la partie 3.2. Par ailleurs, notre estimation du domaine public est entachée d'erreurs liées à des lacunes dans les données : l'absence de géométrie d'une part, et d'autre part la superposition des entités surfaciques de certaines parcelles mal vectorisées (partie 2.1). Ces erreurs génèrent des espaces non-cadastrés dans les données, mais cadastrés en réalité. Pour redresser ces biais, on soustrait à l'estimation de la surface de domaine non-cadastré la sur-

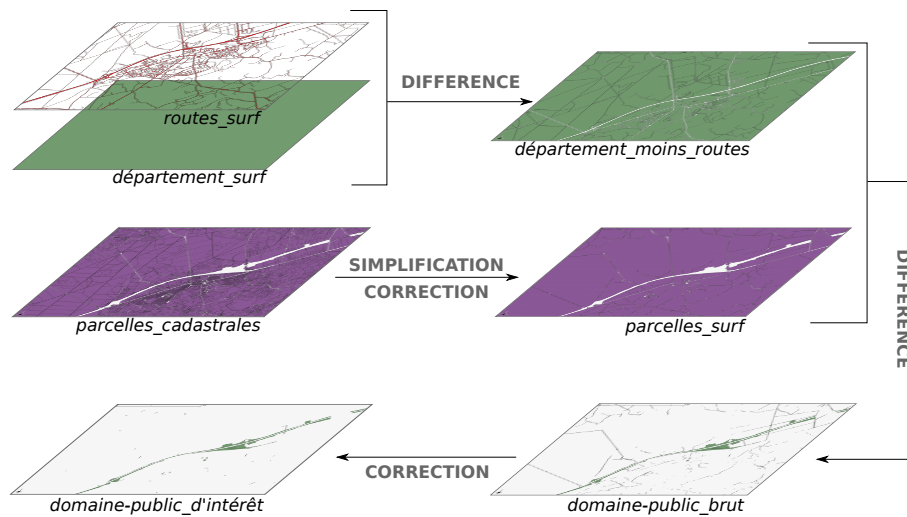


FIGURE 4. *Illustration des principales étapes de la méthode d'identification du domaine public d'intérêt pour une année*

face cumulée des superpositions de parcelles, ainsi que la somme des surfaces saisies numériquement pour les parcelles sans géométrie (Table 1, « Erreurs d'affectation »). Ces redressements permettent de corriger l'estimation quantitative du domaine public dans le département, mais pas de les localiser : les représentations cartographiques ne peuvent donc pas les exclure.

TABLE 1. *Mesure et comparaison de l'apport des différentes étapes à l'identification du domaine public en 2009 et 2021*

	Surfaces soustraites	Evolution de surf. (ha)		Surf. restante (ha)	
		2009	2021	2009	2021
	<i>Surface de départ</i>			<i>683 169,44</i>	<i>683 169,44</i>
Différences spatiales	routes	- 30 118,77	- 30 689,36	653 050,67	652 480,08
	parcelles	- 647 227,81	- 647 490,50	5 822,86	4 989,58
Espaces résiduels	aire < 200m ²	- 530,70	- 551,86	5 292,16	4 437,72
	largeur -3m	- 403,53	- 418,94	4 888,63	4 018,78
	$i \leq 2$ & $M \leq 0,3$	- 603,20	- 594,33	4 285,43	3 424,45
	$i \leq 5$ & $M \leq 0,1$	- 912,55	- 962,92	3 372,88	2 461,53
Erreurs d'affectation	superpositions	- 638,72	- 16,09	2 734,16	2 445,44
	pas de géométrie	- 541,10	- 115,35	2 193,06	2 330,09

4.3. Différence spatio-temporelle

La méthode permet d'aboutir à l'identification du domaine public d'intérêt en 2009 et en 2021. Ces deux données peuvent enfin être associées en une seule table afin d'observer à la fois l'espace et le temps dans un SIG, et ainsi étudier des processus de changement (Cheylan, 1999; Mathian, Sanders, 2014). On peut déterminer la mise en domaine public ou la sortie du domaine public de portions de territoire entre 2009 et 2021 par différence spatiale entre les deux dates, à laquelle on soustrait à nouveau les espaces résiduels liés à des légers décalages entre les deux couches, qu'on assimile à du bruit.

5. Résultats

En Ile-et-Vilaine, le domaine non-cadastré (hors voirie) représente en 2009 4 643 ha, soit 0,68 % de la surface départementale; et 4 858 ha en 2021, soit 0,71 % du territoire. On le distingue en 3 catégories (Figure 5) : les *erreurs d'affectation* correspondent aux superpositions parcellaires et aux parcelles sans géométrie qui induisent des surfaces injustement considérées comme non-cadastrées; grâce à l'amélioration des données cadastrales, leur volume est beaucoup moins important en 2021 qu'en 2009. Les *espaces résiduels* sont les portions de domaine non-cadastré que la faible superficie ou la forme longitudinale amène à considérer comme du bruit résultant des opérations de différence spatiale; ils représentent environ la moitié du domaine non-cadastré en 2009 comme en 2021. Entre ces deux dates, leur superficie augmente légèrement (+ 3 %, soit + 78,07 ha), à peu près proportionnellement à la voirie qui augmente de 2 %, ce qui tend à confirmer que les espaces résiduels sont corrélés au

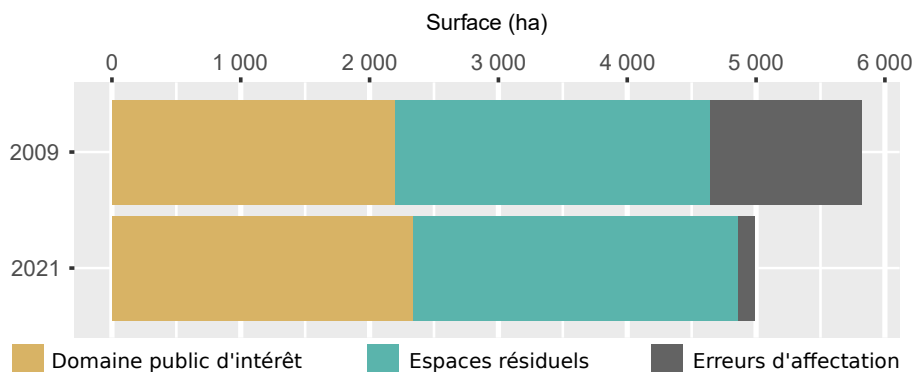


FIGURE 5. Surfaces de domaine non-cadastré (hors voirie) en Ille-et-Vilaine en 2009 et 2021 (Sources : Th. Lecourt 2023, DGFIP-DGALN-CEREMA Fichiers Fonciers 2009-2021)

développement d'infrastructures routières. Le reste du domaine non-cadastré, le *domaine public d'intérêt*, peut ainsi être mesuré indépendamment de l'augmentation de la surface de voirie et de ses espaces résiduels d'une part, et de la réduction des erreurs d'affectation d'autre part : il est en légère augmentation globale dans le département sur la période (+ 6 %, soit + 137,03 ha). Cette statistique départementale cache néanmoins des disparités géographiques : le domaine public est plus répandu en zone urbaine, et sa disparition y est plus concentrée : il représente par exemple 2,84 % (142,96 ha) de la superficie de la commune de Rennes en 2009, et 2,51 % (126,15 ha) en 2021, soit une diminution importante (- 12 %, ou - 16,81 ha). La carte en Figure 6 permet de visualiser la répartition spatiale et l'évolution du domaine public d'intérêt autour de Rennes.

Les résultats de l'identification du domaine non-cadastré en 2021 étant produits à partir de données disponibles en *opendata* par la DGFIP³ et l'IGN⁴, nous sommes en mesure d'en proposer une version téléchargeable en accès libre⁵ au format *geopackage*, comprenant la géométrie polygonale des espaces non-cadastrés, leur caractérisation comme résidu ou non et les indicateurs utilisés pour cette caractérisation. En revanche, les résultats de 2009 et *a fortiori* l'évolution spatio-temporelle ne peuvent être diffusés publiquement car la DGFIP n'a pas à ce jour rendu publiques les données cadastrales historicisées.

3. Plan Cadastral Informatisé : <https://cadastre.data.gouv.fr/data/dgfip-pci-vecteur/>

4. BD Topo : <http://files.opendatarchives.fr/professionnels.ign.fr/bdtopo/>

5. Lien de téléchargement et DOI : <https://doi.org/10.34847/nkl.b43eab01>

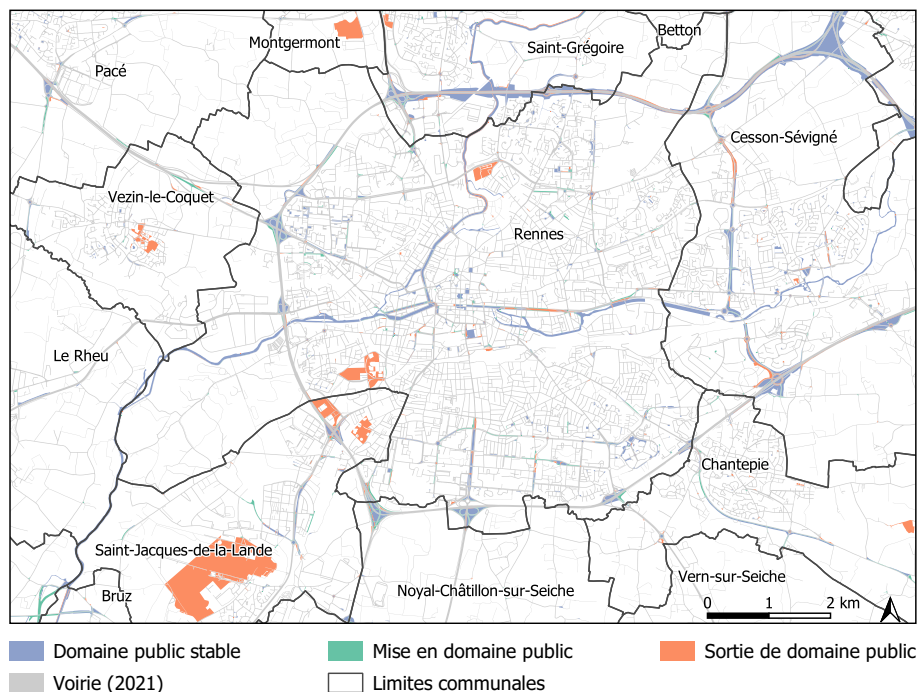


FIGURE 6. Carte de l'évolution du domaine public d'intérêt entre 2009 et 2021 à Rennes. (Sources : Th. Lecourt 2023, DGFIP-DGALN-CEREMA Fichiers Fonciers 2009-2021, IGN BD TOPO 2009-2021. Réalisation : LECOURT, 2023)

6. Discussion et perspectives

A l'échelle du département, on n'observe pas de disparition du domaine public d'intérêt, mais il s'agit bien d'un phénomène observable à l'échelle de la Ville de Rennes, en mettant à part le développement de la voirie et des espaces résiduels conséquents. Le *domaine public stable* entre 2009 et 2021 (Figure 6) est principalement constitué d'interstices entre et autour d'échangeurs (auto-)routiers, et du lit de la Vilaine. Le domaine public fluvial n'étant pas mobilisable pour des opérations d'aménagement, il serait pertinent de l'identifier spécifiquement à partir d'un croisement avec la BD Topo de l'IGN selon une procédure analogue à celle de la détection des routes.

La *mise en domaine public* concerne les portions de territoire cadastrées en 2009 mais qui ne le sont plus en 2021 : là encore, il s'agit principalement d'espaces situés autour d'infrastructures routières. Malgré la méthode d'identification des espaces dédiés à la voirie, une part non-négligeable de ceux-ci

n'est donc pas correctement détectée. Ce constat met en lumière la nécessité d'évaluer la qualité des données produites pour mesurer les faux-positifs et les faux-négatifs dans le domaine public identifié. L'absence d'autre base de données recensant le domaine public empêche d'envisager une validation croisée autrement qu'à partir d'un travail empirique de recension à établir directement avec les acteurs locaux, collectivités locales et services déconcentrés de l'Etat.

De vastes espaces sont *sortis du domaine public* entre les deux dates ; on les remarque notamment au sud-ouest de Rennes. Sur la commune de Saint-Jacques-de-la-Lande, il s'agit de l'aéroport : il a donc été cadastré, mais les Fichiers Fonciers informent que la Métropole et le Département en sont propriétaires ; on peut émettre l'hypothèse que sa déclassification prépare sa privatisation, sur le modèle de l'aéroport de Toulouse-Blagnac et dans la lignée de la gestion foncière des grands ports maritimes (Magnan, 2014). Au nord de l'aéroport, les trois zones au bord de la rocade correspondent au quartier militaire de la Courrouze qui a été rétrocédé par l'Etat à la collectivité locale pour y réaliser un éco-quartier « à fort portage et à visibilité institutionnels » (Faburel, Girault, 2017). On observe bien ici l'effet de la vente du foncier de domaine public d'Etat (Adisson, Artioli, 2020). Ces pistes d'analyses posent plusieurs questions aux décideurs comme aux chercheurs : au-delà de ces zones de grande superficie, y a-t-il un motif qui se répète dans la déclassification du domaine public ? Que devient-il, et qui le possède désormais ? Quels en sont les facteurs et les conséquences ? Cette étude ouvre de nouvelles perspectives stimulantes pour analyser les dynamiques de transformation du foncier de domaine public.

Bibliographie

- Adisson F., Artioli F. (2020). Four types of urban austerity: Public land privatisations in French and Italian cities. *Urban Studies*, vol. 57, n° 1, p. 75–92.
- Artioli F. (2021). Sale of public land as a financing instrument. The unspoken political choices and distributional effects of land-based solutions. *Land Use Policy*, vol. 104.
- Bonneau O., Rivière M., Jarroux N. (2020). *Guide des cessions des biens immobiliers publics*. Dalloz Public.
- Casanova Enault L., Popoff T., Debolini M. (2021). Vacant lands on french mediterranean coastlines: Inventory, agricultural opportunities, and prospective scenarios. *Land Use Policy*, vol. 100.
- CERTU. (2013). *Surfaces occupées par les infrastructures routières*. Consulté sur https://piece-jointe-carto.developpement-durable.gouv.fr/REG074B/FONCIER_SOL/N_OCCUPATION_SOL/L_EMPRISE_ROUTE_R74/Fiche1-7-1.pdf
- CG3P. (2007). *Guide pratique d'utilisation du code général de la propriété des personnes publiques*.
- Cheylan J.-P. (1999). *Représentation de l'espace et du temps dans les SIG* (Revue internationale de géomatique éd., vol. 9). Lavoisier-Hermès.

- Christophers B. (2018). *The New Enclosure. The Appropriation of public land in Neoliberal Britain* (Verso éd.). Brooklyn, NY.
- CNIG. (2016). *Définitions et gestion du domaine non cadastré par la DGFIP - Commission Données (diaporama)*. Consulté sur http://www.cnig.gouv.fr/wp-content/uploads/2016/10/CNIG_Commission_Donn%C3%A9es_12102016_Domaine-non-cadastr%C3%A9_DGFIP-1.pdf
- Dormois R. (2006). Structurer une capacité politique à l'échelle urbaine. Les dynamiques de planification à Nantes et à Rennes (1977-2001). *Revue française de science politique*, vol. 56, n° 5.
- Duflot C. (2013). *Loi n°2013-61 du 18 janvier 2013 relative à la mobilisation du foncier public en faveur du logement et au renforcement des obligations de production de logement social*.
- Faburel G., Girault M. (2017). *Les modes de vie dans les écoquartiers. Phase 2 : les éco-quartiers à fort portage et à visibilité institutionnels (ZAC de Bonne à Grenoble et La Courrouze à Rennes)*. Rapport de recherche. Bureau AD4 (Ministère du Logement, de l'Égalité des Territoires et de la Ruralité).
- Guérois M. (2003). *Les formes des villes européennes vues du ciel. Une contribution de l'image CORINE Land cover à la comparaison morphologique des grandes villes d'Europe occidentale*. Thèse de doctorat non publiée, Université Panthéon-Sorbonne - Paris I.
- Magnan M. (2014). La gestion foncière dans les grands ports maritimes français entre mission d'utilité publique et activité commerciale : un modèle en transition. In *Colloque international DEVPORT*.
- Maignant G. (2005). Compacité et forme urbaine, une analyse environnementale dans la perspective d'un développement urbain durable. In *Actes du colloque développement urbain durable, gestion des ressources et gouvernance*.
- Mathian H., Sanders L. (2014). *Objets géographiques et processus de changement. Approches spatio-temporelles* (Iste éd.).
- Obe R. O., Hsu L. (2021). *PostGIS in Action* (3e éd.). Manning Publications.
- Olsson L. (2018). The Neoliberalization of Municipal Land Policy in Sweden. *International Journal of Urban and Regional Research*, vol. 42, n° 4.
- Peck J. (2012). Austerity urbanism. *City: analysis of urban trends*, vol. 16.
- Rebsamen F. (2021). *Rapport de la Commission pour la relance durable de la construction de logements - Tome 1*. Rapport technique.
- Serrano J., Demazière C. (2016). Le foncier des espaces périurbains dans la planification spatiale : une construction intercommunale et interterritoriale. *Revue d'Économie Régionale & Urbaine*, n° 4, p. 737-766.
- Vergriete P. (2013). *La ville fiscalisée : politiques d'aide à l'investissement locatif, nouvelle filière de production du logement et recomposition de l'action publique locale en France (1985-2012)*. Thèse de doctorat non publiée, Université Paris-Est.
- Whiteside H. (2020). Privatizing Canadian government land and real estate: Railroads, reconciliation, and rip-offs. *Land Use Policy*, vol. 99.

Optimiser la localisation des services publics

Développement méthodologique et application à la Région SUD

Quentin Godoye¹, Cyrille Genre-Grandpierre²

1. UMR 7300 ESPACE, Avignon Université
74 rue Louis Pasteur, 84000 Avignon, France
quentin.godoye@univ-avignon.fr

2. UMR 7300 ESPACE, Avignon Université
74 rue Louis Pasteur, 84000 Avignon, France
cyrille.genre-grandpierre@univ-avignon.fr

RESUME. En France, la mise en place depuis les années 2000 de politiques d'austérité budgétaire a eu pour conséquence la fermeture de services publics, qui a pu entraîner localement une diminution de leur accessibilité. Pour limiter, ou même revenir sur cette baisse d'accessibilité, tout en s'inscrivant dans la logique d'austérité budgétaire, cet article propose un modèle (basé sur celui de p-médiane) permettant d'explorer les optimisations possibles du ratio nombre de services publics / accessibilité de la population aux services. Le modèle permet d'explorer, pour un ensemble de 7 services publics en région SUD (France), les optimisations du ratio nombre de services / qualité de l'accessibilité en simulant successivement : le choix optimal de nouvelles implantations de services à fermer, en relocalisant tout ou partie des services actuels et en combinant fermetures et relocalisations. Les résultats montrent que des optimisations substantielles peuvent être réalisées et prouvent, a minima, tout l'intérêt de développer des méthodes d'analyse spatiale dans une perspective d'aide à la décision pour la localisation de services publics ou autres ressources localisées.

ABSTRACT. In France, the implementation of budgetary austerity policies since the 2000's has resulted in the closure of public services, leading to a local decrease in their accessibility. In order to limit or even reverse this decrease in accessibility while taking into account the logic of budgetary austerity, this article presents a model (based on the p-median model) that makes it possible to explore possible optimisations with regard to the ratio of the number of

public services to the population's accessibility. The model explores, for a set of 7 public services in the SUD region (France), the possible optimisations of this ratio by simulating successively the optimal choice of new locations of services to be closed, by relocating all or part of the current services and by combining closures and relocations. The results show that substantial optimisations can be achieved and prove at least the interest of developing spatial analysis methods in a decision support perspective for the location of public services or other localised resources.

MOTS-CLES : optimisation, services publics, accessibilité, simulation

KEYWORDS: optimization, public services, accessibility, simulation

1. Introduction

Depuis les années 1990, une succession de réformes encadrant le fonctionnement et l'organisation des services publics a vu le jour en France, par exemple la LOLF (Loi Organique relative aux Lois de Finance) en 2001 et la RGPP (Révision Générale des Politiques Publiques) en 2007. Globalement ces réformes, qui s'inscrivent dans un contexte général « d'austérité financière » (Gourgues et Houser, 2017), visent à « une adaptation des missions, de l'organisation et du format de l'État, mais aussi au rétablissement des comptes publics, avec la volonté de dégager des économies dans tous les secteurs d'intervention de l'État » (Sénat, 2011). Cette succession de réformes s'est notamment traduite par une diminution des implantations de nombreux services publics, en particulier dans les zones les moins peuplées (Chouraqui, 2020). Ce repli et ses effets sur leur accessibilité, définie comme la facilité pour accéder à une ressource localisée, au départ d'un lieu donné, en un temps donné, avec un mode de transport donné (Handy et Niemeier, 1997), ont fait l'objet de multiples travaux qui décrivent, pour un type de service particulier (les maternités : Bergonzoni et Simon, 2021 ; les services de justice : Cahu, 2017 ; les forces armées et de l'ordre : Artioli, 2017 ; Blot et Courcelle, 2012), ou pour un territoire donné, local ou national (Taulelle, 2012 ; Courcelle *et al.*, 2017, Barczak et Hilal, 2017), les pertes d'accessibilité consécutives aux fermetures.

Rares sont en revanche les travaux qui se risquent à proposer des solutions pour stabiliser, ou même améliorer, le niveau d'accessibilité aux services publics. Quand des propositions sont formulées, elles consistent à suggérer le développement des e-services (télémedecine, téléadministration, Desbois, 2018), la création de maisons multiservices, qui regroupent en un lieu différents types de services (à l'image des maisons France Services soutenues par E. Macron en 2019), ou tout simplement à stopper les fermetures, voire à proposer des (ré)ouvertures de services. Si ces deux dernières solutions s'imposent comme une évidence théorique, elles s'opposent en revanche pratiquement au contexte d'austérité financière, qu'il soit justifié ou pas, et sont donc peu envisageables.

C'est dans ce contexte que se situent nos travaux, qui visent à savoir si (outre les téléservices) d'autres solutions peuvent être trouvées pour répondre à la fois aux injonctions de l'austérité financière et à l'aspiration à une meilleure accessibilité aux services publics. Plus précisément, nous nous intéressons à la question de l'optimisation de la localisation des services publics actuels. En effet, sachant que les fermetures des services publics s'arbitrent administrativement à l'échelle de l'implantation elle-même - on ferme une maternité faute de naissances, un collègue faute de collégiens- mais qu'elles se font sans vision globale de la qualité de service restante puisque rien dans la loi n'impose de regarder l'existence d'alternatives crédibles (Godoye, 2022) ; sachant que les fermetures se font sans concertation entre les différents types de service, ce que dénonce la Cour des comptes (2019) pour qui « les processus de restructuration des services ont été jusqu'à présent conduits en silo par chaque administration, sans cohérence et concertation transversale aux différents services » ; sachant que les localisations des services publics actuels sont très souvent des héritages, et que les services ont chacun leur propre logique (il est

par exemple historiquement obligatoire d'avoir une gendarmerie par canton rural et la poste, bien que privatisée en 2010, doit toujours s'assurer que 90% de la population soit à moins de 5 km ou de 20 minutes en voiture d'une présence postale, c'est-à-dire soit d'une vraie poste, soit d'un relais poste commerçant dans lequel les services de base de la poste sont proposés) ; sachant enfin que les localisations des services publics ont peu évolué depuis 40 ans et que, dans le même temps, la population a au contraire connu de très fortes dynamiques spatiales, on peut raisonnablement penser qu'une optimisation de l'accessibilité des populations aux services publics est possible, au regard du ratio [nombre d'implantations des différents services / qualité de l'accessibilité produite]. Plus précisément, on peut se demander s'il est possible 1/d'optimiser l'accessibilité en relocalisant tout ou partie des services publics 2/ de continuer à fermer des implantations de services publics, mais en les choisissant de façon « pertinente » pour en limiter les impacts sur l'accessibilité, en s'inscrivant donc sciemment dans une perspective d'austérité 3/ de combiner fermetures et relocalisations de services, pour à la fois diminuer les coûts et stabiliser, voire améliorer, l'accessibilité. En un mot faire mieux avec moins.

Pour localiser optimalement une ressource dans l'espace en fonction d'une demande localisée des méthodes éprouvées existent. Elles visent soit à assurer la meilleure couverture spatiale possible, ce qui correspond à la logique d'équité, qui vise à une minimisation de la variance des situations des individus en termes de distance à la ressource, soit à assurer la meilleure efficacité, qui vise à minimiser la somme des distances parcourues par les individus (Béguin, 1989). Pour l'efficacité, le modèle le plus utilisé depuis les années 80 est celui de la p-médiane (Thomas, 1984 ; 1986). Il est aujourd'hui accessible dans les SIG du marché. Ses applications se limitent toutefois à de petites instances : peu d'éléments à relocaliser, une demande agrégée et peu de sites candidats pour accueillir les localisations. En effet, dès lors que les instances grossissent, on se heurte à des problèmes de capacités de calcul qui rendent les problèmes insolubles, car le nombre de combinaisons à tester explose de façon combinatoire.

C'est dans ce cadre général que se situent nos travaux, qui sont à la fois thématiques et méthodologiques. Ils visent à développer une méthode opérante pour de grandes instances permettant 1/d'explorer les gains possibles en termes d'accessibilité pour des relocalisations totales ou partielles de services publics 2/d'explorer la possibilité de continuer à fermer des services publics en minimisant les impacts sur l'accessibilité 3/de faire « aussi bien sinon mieux » en termes d'accessibilité en combinant fermetures et relocalisations de services publics.

Après une présentation des données mobilisées, nous exposons la méthode développée pour explorer la possibilité d'une optimisation du ratio [nombre d'implantations des services / accessibilité de la population], puis nous présentons quelques résultats pour la région française Provence-Alpes Côte d'Azur pour un ensemble de sept services publics.

2. Données et terrain d'étude

Les services publics considérés sont les collèges, les lycées, les gendarmeries, les maternités, les services d'urgence et la poste. Ils ont été choisis afin de couvrir des thématiques variées et de travailler avec des services aux logiques de répartition différentes. En effet, comme vu précédemment les postes et gendarmeries ont leurs règles spécifiques de répartition, alors que les localisations des collèges, lycées et maternités ne sont pas régies par une règle rigide, mais sont fonction de la demande.

Les données sur les services publics sont issues de la Base Permanente des Equipements (BPE) de 2017 qui recense chaque année à la commune pour toute la France, et géolocalise depuis 2013, des services de diverses natures, mais sans en préciser les dates d'ouverture ou de fermeture, ni les capacités d'accueil.

TABLE 1. *Effectifs des services publics traités en région SUD*

Nom du service	Effectif 2017
Urgence	60
Lycée	352
Hôpital	335
Lycée	352
Collège	478
Gendarmerie	174
Maternité	37
Poste	681

Le terrain d'étude est la région Provence-Alpes Côte d'Azur (ou SUD). Sa dimension, 31 400 km² et 5 millions d'habitants, et sa grande variété de contextes géographiques (littoraux, plaines, zones de montagnes) et de dynamiques démographiques en font un terrain propice pour travailler avec des distributions de services très hétérogènes, tout en rendant réaliste le calcul à fine échelle des accessibilités.

Pour mesurer précisément l'accessibilité des populations résidentes, nous avons utilisé les données carroyées de l'INSEE avec des mailles de 1 km de côté. Ces données n'étant pas disponibles pour 2017, le millésime utilisé est le plus proche de cette date à savoir celui de 2015. L'accessibilité de la population a été calculée en traçant des isochrones de 5, 10, 20, 30, 40, 50, 60 minutes en voiture autour des implantations des services publics, puis en calculant la population présente dans chacun d'eux. Les isochrones sont calculés à partir du réseau routier de la BD TOPO pour des conditions fluides de circulation, en tenant compte des vitesses autorisées et des sens uniques.

3. Optimiser le nombre et la localisation des services publics au regard de l'accessibilité fournie

3.1. Optimiser la localisation des services publics : le modèle théorique de la p-médiane

Pour choisir les implantations à fermer et/ou à relocaliser, nous avons utilisé la logique d'efficacité, plutôt qu'une logique d'équité, car elle correspond mieux au principe de rationalisation porté par les politiques d'austérité. Le modèle utilisé est celui de la p-médiane, développé dans les années 1980 (Hanjoul et Peeters, 1985 ; Thomas, 1986) et qui est toujours très employé aujourd'hui (Daskin et Maass, 2015), car il correspond bien à cette logique rationalisation économique (Beaumont, 1981).

Localiser j services publics selon une logique d'efficacité avec le modèle de la p-médiane revient à trouver les localisations qui permettent de minimiser le temps total d'accès des i individus représentant la demande aux j services. La fonction d'objectif du modèle consiste donc à minimiser la somme des distances D_{ij} en testant les différentes combinaisons de localisation des j services et en retenant la meilleure. Selon les utilisations, il est possible d'ajouter une contrainte au modèle qui porte sur la capacité d'accueil des services, en fixant le nombre d'utilisateurs maximum qui peut leur être affecté. Pour utiliser pratiquement le modèle de la p-médiane, il faut donc connaître la répartition de la demande, le nombre de services à localiser, le cas échéant leur capacité, mais aussi la localisation des « points candidats » susceptibles d'accueillir les équipements (et le cas échéant leur capacité d'accueil). En effet, la surface ayant une dimension topologique de 2 et le point une dimension 1, il existe théoriquement une infinité de possibilités pour localiser n points dans l'espace. Il est donc nécessaire de limiter les possibilités qui seront testées par le modèle. Enfin, le dernier input du modèle est une matrice de distances entre les points de demande i et les points candidats à la localisation des j services.

Une limite du modèle de la p-médiane est qu'il n'est pas en mesure de prendre en charge de grandes instances, car la combinatoire de solutions devient trop grande, ce qui est notre cas puisque le nombre de services publics peut être grand (753 postes), la demande est très désagrégée (il y a 15 863 carreaux de un kilomètre de côté dans lesquels il y a de la population en PACA), et les points candidats sont également potentiellement très nombreux si on veut une approche fine, nécessaire pour obtenir des localisations précises. Par ailleurs, la p-médiane a été conçue pour localiser des services ou équipements, pas pour choisir lesquels fermer en minimisant l'impact des fermetures sur l'accessibilité. Aussi, sur la base théorique du modèle de la p-médiane, nous avons développé notre propre approche opérationnelle, implémentée dans un outil d'optimisation (CPLEX Optimizer) en langage Julia.

3.2. Le modèle pratique développé sur la base du modèle de la p-médiane

3.2.1. Le choix des points candidats

Comme le travail est effectué pour une vaste zone d'étude et de nombreux services, il n'était pas envisageable de choisir manuellement les sites candidats (par exemple en fonction du prix des terrains, de leur proximité aux transports publics, etc.), comme on pourrait le faire pour des travaux portant sur des petits terrains d'étude. La solution retenue, qui est un compromis entre la « tractabilité » du modèle et la précision des localisations obtenues, a consisté à créer un semi régulier de points couvrant l'ensemble du territoire régional, avec un espacement entre les points de 500 m dans les espaces peuplés (là où il y a de la population dans les carreaux de 1 km de côté de l'Insee) et de 1,5 km dans les zones non peuplées. L'espacement de 500 mètres permet d'avoir des localisations précises, et celui de 1,5 km dans les zones peu peuplées n'est pas problématique, car la probabilité qu'un processus de relocalisation optimale visant à minimiser des sommes de distances localise des aménités dans des zones désertes est plus que faible. Au final nous obtenons un jeu de 70 000 points candidats pour l'ensemble de la région SUD.

3.2.2. Comment attribuer des capacités d'accueil aux antennes des services publics ?

La BPE ne fournissant pas les capacités d'accueil des services (qui sont en outre très difficiles à estimer pratiquement, car elles dépendent aussi de l'organisation interne, de la productivité des équipes etc.) et comme nous souhaitions, pour obtenir les simulations les plus réalistes possible, intégrer la question des capacités (sinon rien n'empêche théoriquement d'avoir un seul lycée à Marseille...), nous avons estimé les capacités des services sur la base de polygones de Voronoï. L'hypothèse est que la capacité des services actuels est bien calibrée et qu'elle correspond, à un facteur près, à la population résidente présente dans l'aire de « chalandise naturelle » de chaque service. Les usagers sont donc supposés fréquenter le service le plus près de leur domicile, ce qui est en pratique globalement vérifié, mais pas toujours (notamment pour ce qui a trait aux soins).

Puis, pour permettre la prise en charge de tous les usagers pour les simulations de fermetures de services, les capacités « Voronoï » ont été augmentées arbitrairement de 30%, ce qui donne une certaine « marge de manœuvre » au modèle pour choisir les implantations à fermer. Des augmentations de plus de 30% paraissent peu envisageables pratiquement et des augmentations plus faibles (de l'ordre de 10%) n'auraient sans doute pas laissé suffisamment de marge au modèle pour autoriser des fermetures et assurer la prise en charge de tous les usagers. Précisons toutefois que pour des simulations de fermetures trop importantes (40-50% des services), il est possible qu'aucune solution ne soit trouvée, car les antennes restantes n'ont pas les capacités nécessaires, même augmentées de 30%, pour accueillir tous les usagers.

Précisons qu'il serait possible d'estimer la capacité des différents services, non pas sur la base de la population totale, mais sur la base de leurs usagers théoriques,

en prenant par exemple la tranche des 12-15 ans pour les collèges, des 16-18 ans pour les lycéens, les femmes en âge de procréer pour les maternités etc. Toutefois, cela complexifierait l'approche sans pour autant apporter un gain très notable de précision (au moins à l'échelle de toute la région SUD et pour ce travail à portée méthodologique avant tout) car les taux de scolarisation varient spatialement, de même que les taux de criminalité, de fécondité etc. Aussi, l'option de baser l'estimation de toutes les capacités sur la population totale a-t-elle été retenue, mais rien n'empêche dans le modèle de travailler par la suite avec des capacités plus proches du réel.

Une fois les capacités des services actuelles définies, il est nécessaire lorsque l'on effectue des relocalisations simples ou des fermetures couplées à des relocalisations, de définir les capacités des points candidats susceptibles d'accueillir les services relocalisés. Pour les relocalisations, la méthode utilisée a consisté à réutiliser les capacités d'accueil Voronoï et à les attribuer aléatoirement aux 70 000 points candidats susceptibles d'accueillir un service. Par exemple, dans le cas des maternités, au nombre de 37 en région Sud en 2017, les 37 capacités Voronoï ont été attribuées aléatoirement à l'ensemble des points candidats : on a donc 1 891 points ($70\,000/37=1\,891$) avec la capacité de la maternité 1, 1 891 autres points avec celle de la maternité 2, etc. Le modèle s'assure que les capacités totales des sites retenus au final n'excèdent pas les capacités initiales « réelles » des services de 2017 (on ne crée pas de nouvelle capacité), et qu'aucun site n'a à accueillir plus d'usagers que ce que lui permettent ses capacités.

Précisons enfin que le semis de points candidats étant dense, et leur capacité ayant été définie aléatoirement, il existe de très fortes probabilités d'avoir localement des points candidats aux capacités différentes, ce qui permet au modèle de trouver les meilleures configurations possible.

3.2.3. *L'heuristique développée*

Nous décrivons tout d'abord le modèle développé pour réaliser des relocalisations totales, puis celui pour les relocalisations partielles et enfin celui qui permet de combiner fermetures et relocalisations partielles, car chacun s'appuie partiellement sur le précédent.

Le modèle de relocalisation totale

Les instances traitées étant de grande taille, il n'est pas possible d'utiliser les applications commerciales du modèle de la p-médiane (par exemple celle de Network Analyst de ESRI). Il est donc nécessaire de développer une heuristique, c'est-à-dire une approche qui permette d'obtenir une bonne solution, mais sans avoir la certitude qu'elle est la meilleure (il faudrait pour cela tester toutes les possibilités et retenir la meilleure). L'heuristique développée se déroule en 4 étapes.

Supposons que l'on veuille relocaliser optimalement la totalité d'un effectif de 100 maternités et que l'on ait 70 000 points candidats.

– étape 1 : on commence par réaliser des sélections aléatoires (appelées sous-problèmes), ici 100, comprenant chacune 800 points candidats parmi les 70 000 possibles. Le nombre de sous-problèmes et de points par sous-problème (800) a été fixé en nous appuyant sur les travaux de (Mu, Tong, 2020) et après la réalisation de tests visant à trouver le meilleur compromis (nombre de sous-problèmes – temps de calcul – qualité des résultats). À titre indicatif, le passage à 200 sous-problèmes améliore en moyenne les résultats de 5%, mais peut conduire (selon le type de service) à plus que tripler les temps de calcul.

– étape 2 : pour chaque sous-problème, on fait tourner une heuristique « classique » de la littérature en Recherche Opérationnelle : l'heuristique TB (Teitz et Bart, 1968). Elle est initialisée avec un choix aléatoire de 100 points tirés dans chaque sous-problème, que nous appellerons la solution 1. Ensuite, l'heuristique effectue une « recherche de descente la plus rapide » pour chaque sous-problème en remplaçant itérativement un emplacement sélectionné de la solution 1, par l'un des 700 autres restant dans le sous-problème. À chaque remplacement la fonction d'objectif (la somme des D_{ij}) est recalculée. Au total il y a donc 100×700 solutions à calculer et à comparer pour sélectionner au final la meilleure que nous appellerons la solution 2. Cette solution 2 est ensuite prise comme solution de base pour relancer la même procédure que celle vue pour la solution 1 (remplacement itératif de chaque point par un des 700 autres et recalcul de la fonction d'objectif, etc.). L'opération se répète ensuite de façon itérative pour une solution 3, une solution 4, etc. Le problème est que cette procédure peut être très gourmande en temps de calcul (entre plusieurs dizaines d'heures et plusieurs jours selon le nombre de services), aussi l'avons-nous arrêtée lorsque le gain en termes de fonction d'objectif devenait marginal pour une solution n par rapport à la solution $n-1$. Concrètement, après des tests, le seuil de 10^{-4} a été choisi, car il permet un bon compromis entre le temps de traitement et la pertinence des résultats, ce qui signifie que si une itération fournit par exemple une valeur pour la fonction d'objectif de 10 000, il faut que pendant les 10 itérations suivantes la solution trouvée n'ait pas une valeur inférieure à 9 999 (ce qui représente une amélioration de 10^{-4}), pour stopper la recherche. À ce stade de la méthode, il est donc possible de trouver les 100 meilleurs points candidats susceptibles d'accueillir les 100 maternités pour chaque sous-problème de 800 points.

– étape 3 : évaluation de tous les points candidats au regard de leur capacité à accueillir un service. Elle consiste à attribuer à chaque point candidat A_i une note en fonction de sa distance à chaque emplacement composant la solution des sous-problèmes. La note pour un point i par rapport à un point j faisant partie de la solution du sous-problème k est : $v(i,j,k) = 1/(\exp(\text{Dist}(a_{kj}, A_i)^2))$ où $\text{Dist}(a_{kj}, A_i)$ est la distance entre l'emplacement candidat A_i et l'emplacement a_{kj} . On effectue ensuite pour chaque point la somme des notes obtenues pour chaque sous-problème, puis la somme des notes obtenues pour tous les sous-problèmes. Les emplacements fréquemment retenus dans les solutions de sous-problèmes pour lesquelles $\text{Dist}(a_{kj}, A_i) = 0$, ou situés à proximité des emplacements souvent retenus auront donc une note élevée.

– étape 4 : une fois tous les points notés, ils sont triés en fonction de leur note et les 800 mieux notés sont sélectionnés comme candidats pour faire partie de la

solution pour résoudre le problème d'origine. L'heuristique TB est relancée une dernière fois de façon à trouver la solution définitive, à savoir les 100 meilleurs points pour accueillir les 100 maternités à localiser au regard de la fonction d'objectif.

Le modèle de relocalisation partielle

Supposons que l'on veuille à présent relocaliser optimalement 10% d'un total de 100 maternités (on a donc 10 localisations optimales à trouver). On identifie grâce au modèle de la p-médiane classique en les supprimant successivement, les 10 maternités actuelles les « moins bien placées », c'est-à-dire celles qui contribuent le moins bien à la minimisation de la fonction d'objectif. Ces 10 maternités seront celles qui seront à relocaliser. On bloque la localisation des 90 autres en leur affectant « virtuellement » une demande très forte pour s'assurer que le modèle les retienne dans la solution finale. Puis, les 10 maternités sont relocalisées selon la même procédure que pour la relocalisation totale.

Le modèle fermeture - relocalisation

Le modèle est très proche de celui de la relocalisation partielle. Supposons que l'on veuille fermer 10% d'un total de 100 maternités (donc 10) et en relocaliser 10% du total initial (donc 10 autres). Comme pour le modèle de relocalisation partielle, on identifie grâce au modèle de la p-médiane les 20 maternités les « moins bien placées », c'est-à-dire celles qui contribuent le moins bien à la minimisation de la fonction d'objectif. Parmi ces 20 maternités, les 10 « pires » maternités sont supprimées et les 10 autres vont être celles qui seront à relocaliser. Les localisations des 80 maternités restantes vont être conservées et dans cette perspective on va leur affecter, comme précédemment, virtuellement un très grand nombre d'utilisateurs à proximité directe. La suite reprend le processus, déjà vu, basé sur l'utilisation de sous-problèmes.

4. Résultats

Les résultats présentés ne sont pas exhaustifs (pour plus de détails cf. Godoye, 2022), mais ils ont été choisis pour illustrer les enseignements principaux des différentes simulations et ouvrir de nouvelles pistes de travail.

4.1. Fermetures de services : peut-on encore aller plus loin ?

Les simulations de fermetures de services ont consisté à fermer 5 ; 10 ; 20 ; 30 ; 40 et 50% des effectifs de chaque type de service en les choisissant de façon à limiter l'impact des fermetures sur l'accessibilité de la population. Les résultats montrent tout d'abord qu'il n'existe pas d'effet de seuil marqué, puisque la dégradation de l'accessibilité se fait progressivement en fonction de l'augmentation du pourcentage de fermetures. Mais, ils montrent aussi, et surtout, qu'il est possible de fermer jusqu'à 30% des effectifs d'un service sans impact lourd sur l'accessibilité

(rappelons qu'on postule dans le même temps une augmentation de 30% des capacités des services restants). Par exemple sur la figure 1 pour les services d'urgence, on peut voir que la part de la population dans chaque tranche d'isochrone varie très peu en fonction du pourcentage de fermetures. On illustre ici que l'accessibilité est bien fonction du nombre de services, mais aussi de leur localisation et capacité.

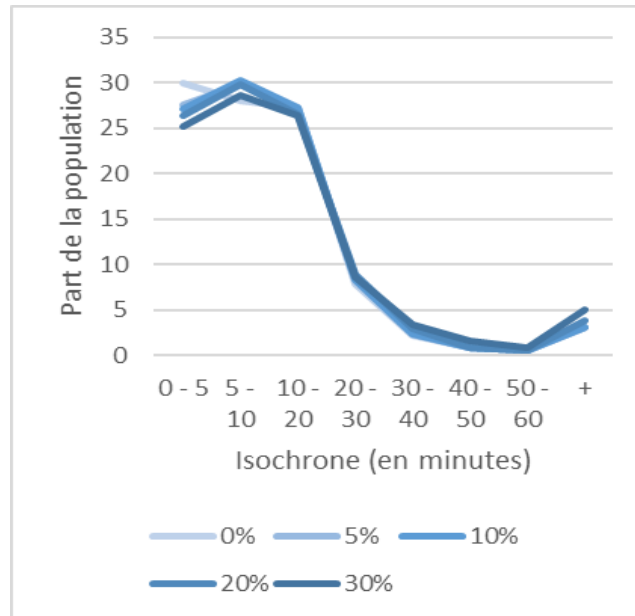


FIGURE 1. Part d'individus par isochrone en fonction de la réduction du nombre de services d'urgences

Pour avoir une vision plus agrégée des effets des fermetures choisies, nous avons calculé la part de la population et la part de la surface régionale qui perdent de l'accessibilité à 1, 2, 3, etc. jusqu'à 7 services, en fonction de la part des services de chaque type qui ferment. On constate alors voir (table 2) que pour 20% de fermeture de tous les services, seule 22,6% de la population perd de l'accessibilité à un service et que cette part tombe à 7,9% pour 2 services (elle est de zéro ou presque pour 5 services et plus). Les pertes en temps apparaissent par ailleurs limitées (inférieures à 5 minutes dans l'immense majorité des cas).

Spatialement, les antennes choisies pour être fermées par le modèle sont situées, de façon assez contre-intuitive, dans les zones denses et bien équipées lorsque les services sont nombreux et concentrés (collèges, lycées, hôpitaux), et plus dans les zones plus rurales pour les autres services (gendarmerie par exemple). Ce résultat tend à montrer que tout se passe comme si la « rationalisation » de l'implantation des services publics engagée dès les années 90 était allée trop loin, en tous cas plus

loin que nécessaire, pour assurer l'efficacité, puisque le modèle choisit des antennes à fermer en zones denses et non en zones peu peuplées, contrairement à ce que commanderait théoriquement la logique d'efficacité.

TABLE 2. *Impacts sur la population et le territoire de la cumulativité des pertes d'accessibilité selon le % simulé de fermetures*

		Perte d'accessibilité à n service					
		0	1	2	3	4	5
10% de fermeture	Part de la population	75,4	17,4	6,1	0,9	0,2	0
	Part du territoire	73	22,4	4,1	0,5	0,1	0
20% de fermeture	Part de la population	68	22,6	7,9	1,2	0,2	0
	Part du territoire	61,1	28,8	8,8	1,2	0,1	0
30% de fermeture	Part de la population	61,5	25,5	10,7	1,9	0,4	0
	Part du territoire	47,1	32,9	15,2	4,1	0,6	0,1

Ces simulations de fermetures tendent donc à montrer qu'il existe de larges marges de manœuvre pour continuer à fermer des services publics sans pour autant dégrader l'accessibilité, à condition toutefois d'augmenter les capacités des services restants. Pour une utilisation pratique de ces résultats, reste à savoir s'il est moins coûteux de fermer certains services et d'augmenter la capacité d'autres ou de maintenir stable le nombre de services ?

Outre les questions économiques, les fermetures de services étant en général mal vécues par les populations qui les subissent (Fijalkow, Taulelle, 2020), même si les conséquences sur l'accessibilité demeurent limitées, nous avons ensuite cherché à mesurer les effets de relocalisations de services, et de fermetures couplées à des relocalisations, dans l'hypothèse qu'elles pourraient être mieux vécues par les populations.

4.2. Quels bénéfices attendre de relocalisations totales ou partielles des services ?

4.2.1. Les relocalisations totales

Même si une relocalisation totale des services est concrètement non envisageable, elle a été testée afin de faire office de « tube témoin » pour les autres simulations. Ainsi, en se donnant la liberté de relocaliser toutes les implantations de tous les services, on observe que ces relocalisations se font en défaveur des zones denses (mais avec des différences d'intensité selon les services) et au profit des périphéries des grandes agglomérations et des espaces peu peuplés. Les services se

« desserrent ». Par exemple les distances moyenne et médiane au plus proche voisin pour les lycées passent de 1 629 et 426 mètres, à respectivement 4 642 et 3 818 mètres. Pour les services initialement plus répartis spatialement en raison de règles relevant globalement de l'équité (comme les postes et les gendarmeries), on observe au contraire une « reconcentration ». La répartition de ces services présente donc un « déficit » d'efficacité quand, au contraire, les autres apparaissent « trop concentrés », même au regard d'une recherche d'efficacité.

En termes d'accessibilité, on peut prendre l'exemple des lycées qui illustre bien la situation majoritaire des services concentrés. Pour une relocalisation totale, rien ne change pour 68,4% de la population, mais 28,4% des habitants gagnent en accessibilité, ainsi que 65,7% de la surface régionale (pas de changement pour 1/3 de la surface). Pour les services plus répartis, la relocalisation totale favorise les espaces fortement peuplés au détriment des zones peu denses, mais les effets globaux négatifs sur l'accessibilité globale sont faibles, car les populations concernées sont réduites.

Ainsi, la relocalisation totale des services montre qu'une grande marge d'optimisation existe quant à la localisation des implantations des différents services, et ce particulièrement pour les services initialement les plus concentrés. Les distributions actuelles apparaissent donc globalement comme n'étant pas régies par la logique d'efficacité, ou alors les implantations héritées ne correspondent plus à la distribution actuelle de la population.

4.2.2. *Les relocalisations partielles*

Les simulations de relocalisations partielles, plus envisageables que les relocalisations totales, montrent qu'il est possible d'améliorer localement l'accessibilité, parfois avec des gains importants, notamment dans les espaces ruraux ou en marge des centres urbains, mais que globalement les résultats sont assez décevants et éloignés en termes d'intensité de ceux des relocalisations totales. Les espaces qui bénéficient des relocalisations partielles sont globalement les mêmes que pour les relocalisations totales à savoir les espaces peu peuplés, mais l'impact global sur l'accessibilité est limité. Sur la Figure 2 qui concerne les lycées et qui donne en abscisse les tranches d'isochrones et en ordonnée le pourcentage de population qu'elles contiennent, on peut voir cet effet limité des relocalisations partielles, même si elles portent sur 40% des effectifs. Les résultats en termes d'accessibilité de toutes les simulations demeurent proches de la situation actuelle, à l'exception de la simulation avec 100% de relocalisations.

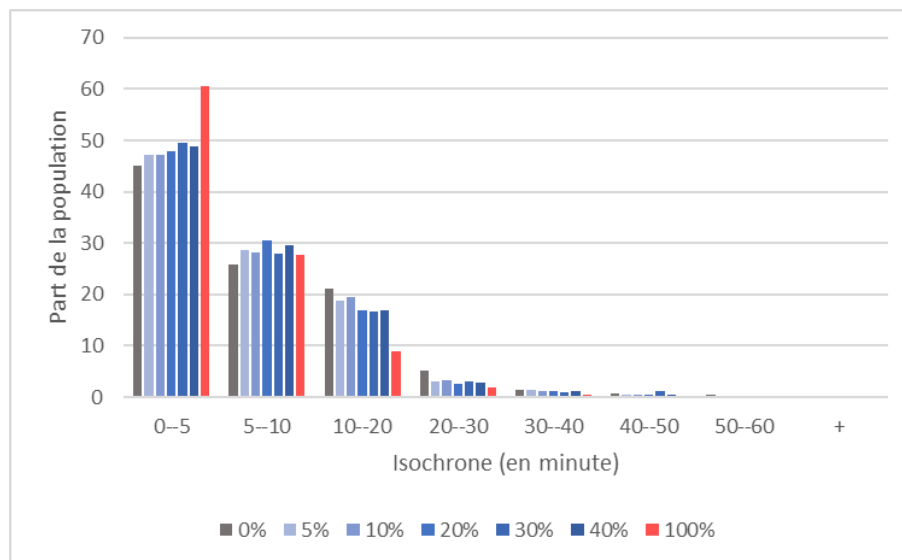


FIGURE 2. Population par isochrone en fonction de la part de relocalisation des lycées

Dans la même veine, on observe que la relocalisation de 10% des collèges bénéficie seulement à moins de 3 % de la population et à 9% du territoire régional et que les chiffres n'atteignent respectivement que 3,4% et 10,5% pour 30% de relocalisation. Cet effet limité des relocalisations s'observe aussi avec les mêmes ordres de grandeur pour les services plus régulièrement répartis comme les postes ou maternités.

L'analyse agrégée pour tous les services montre aussi que, certes les relocalisations partielles des services publics peuvent avoir un bénéfice en termes d'accessibilité, mais que celui-ci reste modeste à la fois en intensité (en général les usagers gagnent un isochrone, soit 5 – 10 minutes maximum) et quant aux volumes des populations concernées (seule 5,3% de la population gagnerait de l'accessibilité à 3 services ou plus pour une relocalisation de 30% de tous les services).

TABLE 3. *Gain d'accessibilité à n services en fonction du pourcentage de relocalisation des services*

		Gain d'accessibilité à n services					
		0	1	2	3	4	5
10% de relocalisation	Part de la population	54,8	35	8	1,7	0,4	0
	Part du territoire	48,9	28,6	14,3	6,6	1,5	0,1
30% de relocalisation	Part de la population	54,4	27,3	13,1	4,2	0,9	0,2
	Part du territoire	43,1	26,6	19,4	7,2	3,2	0,5

Le bilan des simulations de relocalisation partielle apparaît donc un peu « décevant », à tout le moins au regard de ce qui a été vu pour les fermetures, sans même parler du fait que les relocalisations seules ne correspondent pas à une logique d'austérité, puisqu'elles sont plus un coût qu'une économie. En effet, même en allant jusqu'à 30% de relocalisation, ce qui aurait un coût financier et social fort, les impacts en termes d'accessibilité paraissent trop faibles. Il paraît donc très peu justifiable de promouvoir des relocalisations massives, quand bien même elles favoriseraient des zones peu pourvues en équipements, pour améliorer globalement de façon très faible l'accessibilité, qu'il s'agisse de celle de la population ou du territoire.

4.3. Fermetures et relocalisations : le bon compromis ?

On l'a vu, des fermetures bien choisies paraissent plus pertinentes que des relocalisations pour optimiser le ratio nombre de services / accessibilité, mais elles ne peuvent être que mal vécues localement. Aussi, dans un dernier jeu de simulations, qui se voudrait un « compromis », nous avons combiné fermetures et relocalisations pour voir s'il était possible de gagner sur tous les tableaux, à savoir les coûts et l'accessibilité. Quatre combinaisons ont été réalisées pour chaque service : 10% de fermetures (F) et 10% de relocalisations (R), puis 10 F et 30 R, 30 F et 10 R et 30 F et 30 R.

Les résultats montrent qu'associer fermetures et relocalisation ne produit pas strictement les mêmes effets sur l'accessibilité selon les services considérés. Trois profils se révèlent.

Pour les maternités, peu nombreuses, il n'est pas possible en combinant fermetures et relocalisations de retrouver une situation en termes d'accessibilité équivalente à celle de 2017. Si avec les différentes combinaisons envisagées la situation évolue peu pour les populations lointaines des maternités, elle se dégrade

en revanche pour ceux qui ont une maternité à proximité en 2017. Ici le faible nombre d'aménités ne permet pas vraiment l'optimisation.

Pour les hôpitaux, postes, gendarmeries, urgences, les relocalisations parviennent au contraire à compenser les effets des fermetures et limitent l'augmentation des temps de trajet. Même avec 30% de fermeture et 10% de relocalisations on peut retrouver la situation de 2017. Globalement les populations proches des services en 2017 perdent un peu en termes d'accessibilité, au profit de celles qui sont plus lointaines qui peuvent gagner beaucoup.

Enfin, pour les services les plus concentrés que sont les collèges et lycées, il est possible de fermer une grande part des services, avec un minimum de relocalisation, tout en améliorant l'accessibilité par rapport à 2017. Sur la figure ci-dessous on voit par exemple que le scénario 30F10R (30% de fermetures et 10% de relocalisation) donne de meilleurs résultats en termes d'accessibilité que la situation en 2017 !

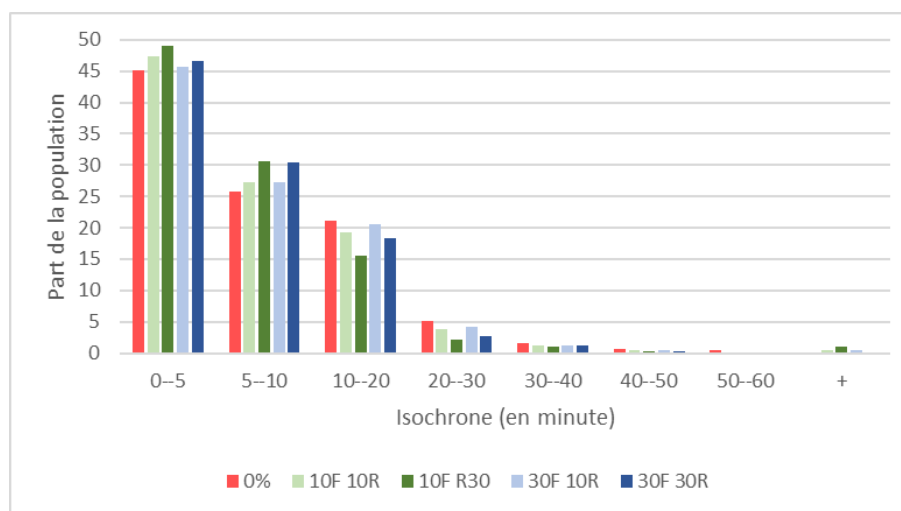


FIGURE 3. Accessibilité des individus aux lycées en fonction des parts de fermetures et de relocalisations simulées selon une logique d'efficacité

En considérant simultanément tous les services, on note que les fermetures combinées à des relocalisations donnent de bons résultats en termes d'accessibilité, puisque par exemple pour 30% de fermetures et 10% de relocalisation, 68% de la population voit son accessibilité à au moins un service s'améliorer et 64% pour la surface de la région.

5. Conclusion

Pour conclure, il apparaît difficile de répondre simplement à la question initiale de savoir s'il est possible de faire mieux avec moins ? De manière générale et au

sens strict, ce n'est pas possible. Les simulations montrent en effet qu'en combinant fermetures et relocalisations, il est difficile d'obtenir des résultats d'accessibilité meilleurs pour toutes les populations qu'avec la disposition réelle des services de 2017, et ce particulièrement pour les services à faible effectif. Pour autant, certaines simulations combinant fermetures et relocalisations donnent des résultats très comparables au niveau d'accessibilité de 2017, mais avec des effectifs de services qui peuvent être réduits d'un tiers. On fait donc alors aussi bien avec moins.

Un taux de relocalisation élevé combiné à un faible taux de fermeture est logiquement la combinaison qui permet pour la plupart des services d'obtenir les meilleurs résultats en termes de minimisation de temps de trajet des individus à l'équipement le plus proche.

Enfin, et au sens strict, pour certains services très concentrés dans les espaces urbains, il apparaît effectivement possible de réellement faire mieux avec moins (toujours en acceptant l'idée de l'augmentation des capacités des services restants). C'est notamment le cas des lycées, et dans une moindre mesure des collèges. On mesure bien ici l'intérêt d'explorer grâce à des méthodes d'analyse spatiale, sous-utilisées en pratique, les marges de manœuvre qui existent dans l'optimisation du nombre et de la distribution des services publics (en précisant que les applications peuvent concerner d'autres thématiques).

D'un point de vue plus thématique, nous avons aussi pu mettre en évidence que la plupart des services étudiés apparaissent trop concentrés dans les espaces densément peuplés, même au regard d'une logique d'efficacité. En effet, cette méthode d'optimisation tend théoriquement à favoriser les espaces avec une population nombreuse, au détriment des zones moins peuplées. Or, observer dans la quasi-totalité des simulations menées, que le modèle basé sur une logique d'efficacité favorise les périphéries des espaces urbains et les zones plus isolées, tend à montrer que les services sont concentrés à un point tel que cela en devient contre-productif pour l'accessibilité globale.

Bien entendu, pour tendre vers plus d'opérationnalité, il serait nécessaire de travailler avec les capacités réelles des services et/ou leurs spécificités (tous les services d'urgence ne proposent par exemple pas le même niveau de service), et il serait important de savoir quels sont les coûts globaux et respectifs des fermetures – relocalisations - augmentation des capacités, mais la base théorique du modèle n'en serait en rien modifiée.

Concernant la poursuite de ces travaux, elle va se faire dans un premier temps en intégrant aux méthodes permettant de simuler fermetures et relocalisations une dimension « multiéchelle avec une contrainte de couverture ». Il est en effet important d'envisager la répartition des services à différents niveaux d'échelle, pour différents découpages. Intégrer une dimension multiéchelle au cœur de la méthode se fera en formulant des contraintes spécifiques telles que : il est obligatoire que chaque intercommunalité ait au moins un service de tel type et chaque canton en ait un de tel autre ; ou alors : il est impossible de fermer une antenne d'un service dans une commune s'il n'existe pas d'alternative dans une commune contiguë ou si la commune de fermeture ne compte plus, suite à la fermeture, un minimum de nombre

de services (qui peut être fonction de la population). Ces développements doivent permettre de renforcer l'opérationnalité de la méthode et de sortir de la logique de gestion en silo qui caractérise les services publics.

Bibliographie non numérotée et références

- Artioli F. (2017). Les politiques du retrait territorial de l'État : Réformes de la carte militaire et gestion des mobilisations locales (1989-2012), *Gouvernement et action publique* 1 (6), 81-106.
- Barczak A., Hilal M. (2017). Quelle évolution de la présence des services publics en France, in Courcelle, Fijalkow and Taulelle, *Services publics et territoires : adaptations, innovations et réactions*, Presses Universitaires de Rennes
- Béguin H. (1989). Efficacité et équité en aménagement du territoire. *L'Espace géographique*, 18, 4, p. 335-336.
- Beaumont J.R. (1981). Location-allocation problems in a plane a review of some models, *Socio-Economic Planning Sciences*, 15, 5, p. 217-229.
- Bergonzoni A., Simon M. (2021). *La part des femmes en âge de procréer résidant à plus de 45 minutes d'une maternité augmente entre 2000 et 2017*, Direction de la Recherche, des Études, de l'Évaluation et des Statistiques, juillet 2012, n°1201, 6 p.
- Blot F., Courcelle T. (2012). Bien observer sans se faire remarquer, est-ce toujours possible avec la RGPP ? Le rôle social et l'implantation territoriale de la gendarmerie en question, *Sciences de la société*, 86 / 2012 : 14-33.
- Cahu E. (2017). *Géographie de la justice pénale en France. L'équité à l'épreuve des territoires*. Thèse de doctorat en géographie, Normandie Université.
- Chouraqui J. (2020). Les réformes et l'État et la transformation des services publics : des tendances internationales au local, quels effets pour les territoires ? *Annales de Géographie*, 732 : 5-30.
- Courcelle T., Fijalkow Y., Taulelle F. (2017). *Services publics et territoires : adaptations, innovations et réactions*, Presses universitaires de Rennes (Collection « Espace et territoires »), 253 p.
- Cour des Comptes (2019). *L'accès aux services publics dans les territoires ruraux*, Cours des comptes, Paris.
- Daskin M.S, Maass K.L. (2015). The p-Median Problem, in Laporte G, Nickel S, Saldanha da Gama F (dirs.), *Location Science*, Cham, Springer International Publishing, p. 21-45.
- Desbois.D. (2018). La télémédecine, un instrument susceptible de corriger les inégalités de santé ? *Terminal. Technologie de l'information, culture & société*, 122, pp.1-10.
- Fijalkow Y., Taulelle F. (2020). La mobilisation en faveur du maintien des services publics : la colère des territoires ruraux de faible densité, *Géographies et cultures*, n°114-été 2020 : 17-35.
- Godoye Q. (2022). *Austérité et aménagement du territoire. Diagnostic des inégalités d'accès aux services publics et développement de méthodes d'optimisation pour limiter les*

inégalités territoriales. Application à la région SUD. Thèse de doctorat de Géographie, Avignon Université.

- Gourgues G, Houser M. (dir) (2017). *Austérité et rigueur dans les finances locales : une approche comparative et interdisciplinaire*, L'Harmattan, Paris, 295 p.
- Hanjoul P., Peeters D. (1985). A comparison of two dual-based procedures for solving the p-median problem, *European Journal of Operational Research*, 20, 3, p. 387-396.
- Handy S.L., Niemeier D.A. (1997). Measuring Accessibility: An Exploration of Issues and Alternatives, *Environment and Planning A*, 29 : 1175-1194.
- Mu W., Tong D. (2020). On solving large p-median problems, *Environment and Planning B: Urban Analytics and City Science*, 47, 6, p. 981-996.
- Sénat (2011). *Rapport d'information sur les conséquences de la RGPP pour les collectivités territoriales et les services publics locaux*, N°666.
- Taulelle F. (2012). *Le délaissement du territoire : quelles adaptations des services publics dans les territoires ruraux ?* Presses universitaires du Mirail, Toulouse.
- Teitz M.B., Bart P. (1968). Heuristic methods for estimating the generalized vertex median of a weighted graph », *Operations research*, 16, 5, p. 955-961.
- Thomas I. (1984). Towards the simplification of location models for public facilities: The case of the postal service », *Papers of the Regional Science Association*, 55, 1, p. 47-58.
- Thomas I. (1986). Analyse spatiale et modèles de localisation optimale. Outils opérationnels d'aide à la décision, *Actes des colloques de l'AIDELF*, p. 139-153.

CentipedeRTK

Le Real Time Kinematic collaboratif, low-cost et open source

Positionnement GNSS temps réel, cinématique, collaboratif et en accès libre et à faible coût

Ancelin Julien¹, Ladet Sylvie², Heintz Wilfried²

1. INRAE, UE 0057 DSLP

Domaine expérimental de Saint-Laurent-de-la-Prée, France

Julien.Ancelin@inrae.fr

2. INRAE, Université de Toulouse, UMR 1201 DYNAFOR

Castanet-Tolosan, France

Sylvie.Ladet@inrae.fr; Wilfried.Heintz@inrae.fr

Le système GNSS (Géolocalisation et Navigation par un Système de Satellites) avec correction différentielle temps réel offerte par le RTK fournit une précision sub-métrique ou centimétrique, même dans des environnements difficiles, grâce à l'augmentation du nombre de satellites et à l'accès facilité à des technologies et matériels dédiés. Cependant, l'accès à ces technologies de géolocalisation précises a longtemps été limité par les coûts élevés des équipements. Avec l'émergence de solutions libres et ouvertes, ce frein est désormais levé. Le réseau CentipedeRTK, créé par des agents issus d'organismes français de recherche publique, vise à fournir un maillage national d'antennes fixes qui fournissent un signal de correction ouvert. Ce réseau d'acteurs scientifiques et techniques offre également des modes opératoires en libre accès pour construire et utiliser des appareils RTK à bas coût. En conséquence, les utilisations de CentipedeRTK se sont multipliées dans divers domaines tels que l'agriculture de précision, la recherche forestière ou océanique.

The GNSS (global navigation satellite systems) system with real-time differential correction offered by RTK provides sub-meter or centimeter precision, even in challenging environments, due to the increasing number of satellites and easier access to dedicated technologies and equipment. However, access to these precise geolocation technologies has long been limited by the high cost of equipment. With the emergence of free and open solutions, this barrier has now been lifted. The CentipedeRTK network, created by agents from the Public Research sector, aims to provide a nationwide network of fixed antennas that provide an open correction signal. This network of scientific and technical actors also provides open access to operational modes for building and using low-cost RTK devices. As a result, uses of CentipedeRTK have multiplied in various fields such as precision agriculture, forest or oceanic research.

MOTS-CLES : CentipedeRTK, GNSS, géolocalisation, collaboratif, RTK, correction différentielle, précision, satellites, technologies, matériels, réseau, techniques, modes opératoires, libre accès, code ouvert, matériel libre, usages.

KEYWORDS: CentipedeRTK, GNSS, geolocation, collaborative, RTK, Differential Correction, Precision, Satellites, Technologies, Equipment, Network, Technical, Operational Modes, Open-Access, Open-Source, Open-Hardware, uses cases.

1. Introduction

La géolocalisation est un processus qui permet de déterminer la position géographique d'un objet fixe ou mobile sur la Terre. Le système de localisation par satellite utilise la trilatération, où la position de l'objet est déterminée à partir des signaux échangés entre le récepteur de l'objet et au moins quatre satellites d'une constellation. La mesure du temps de parcours de ces signaux permet de déterminer la longitude, la latitude et l'élévation de l'objet, avec une précision qui varie en fonction de la technologie utilisée, allant de quelques dizaines de mètres avec un système de positionnement par satellite (GNSS) conventionnel (Bossier, 2012 ; Bossier, 2014 ; Duquenne *et al.*, 2005 ; Lahaye et Ladet, 2014a) à des résultats centimétriques avec une correction différentielle temps réel (RTK) ou différé (DGSS).

La correction différentielle est une technique utilisée pour améliorer la précision de géolocalisation. Elle consiste en l'obtention en temps réel d'un signal de correction spécifique pour rectifier les déviations de position XYZ causées par les couches atmosphériques (Lahaye et Ladet, 2014b). La cinématique temps réel (Real Time Kinematic ou RTK en anglais) en est une forme spécifique qui utilise des informations sur les différences de phase des signaux GPS, GLONASS ou Galileo entre autres pour fournir des corrections en temps réel à l'aide d'une station de référence au sol. Cette utilisation de la correction différentielle permet d'atteindre une précision de géolocalisation de l'ordre du centimètre pour l'utilisateur.

La technologie RTK est conçue pour des applications telles que la surveillance, la mesure et le guidage de précision. La station de base est située à un emplacement connu et fonctionne comme point de référence pour les récepteurs mobiles. Ceux-ci peuvent alors fournir des positionnements précis en mesurant leur position relative à ce point de référence. La qualité des résultats dépend de la densité et des capacités du réseau de stations de référence, qui doit être accessible et proche du lieu de collecte. Cependant, l'utilisation de RTK nécessite du matériel coûteux, généralement propriétaire, ainsi que des abonnements à un service de diffusion de correction pour chaque machine connectée.

Les objectifs du projet CentipedeRTK sont (1) de développer une chaîne de traitement innovante reposant sur des logiciels et des composants matériels en libre accès pour créer des solutions fiables, légères, peu coûteuses et faciles d'utilisation et (2) d'offrir une couverture complète du territoire (en commençant par la France métropolitaine). Cet article présente CentipedeRTK en tant que réseau collaboratif

de bases GNSS ouvertes et disponibles pour toute personne se trouvant dans la zone de couverture du système. Un effort important a été fourni pour partager en ligne un ensemble de ressources documentaires (mode opératoire, code source, tutoriels), tout en répondant à l'enjeu de FAIRisation des objets de la Recherche (<https://doi.org/10.5281/zenodo.5814772>). Depuis son déploiement en France en 2019, le réseau s'est considérablement étendu avec plus de 400 stations opérationnelles et une variété de domaines d'application, notamment dans le secteur public (17%) et privé (66%), en particulier du monde agricole.

2. Description du réseau et de ses différents segments

L'origine du nom et du logo du réseau CentipedeRTK provient du nom des mille-pattes, une classe d'arthropodes appelée chilopodes ou centipèdes (Fig. 1a). Cette désignation s'inspire de la pratique courante dans le monde de l'open-source de choisir des logos basés sur des animaux. L'analogie entre un mille-pattes et CentipedeRTK peut être établie ainsi : tout comme les nombreuses pattes d'un mille-pattes coopèrent pour atteindre un objectif commun, CentipedeRTK utilise plusieurs stations de références GNSS pour augmenter sa disponibilité par rapport à ce qui serait possible avec une seule référence.

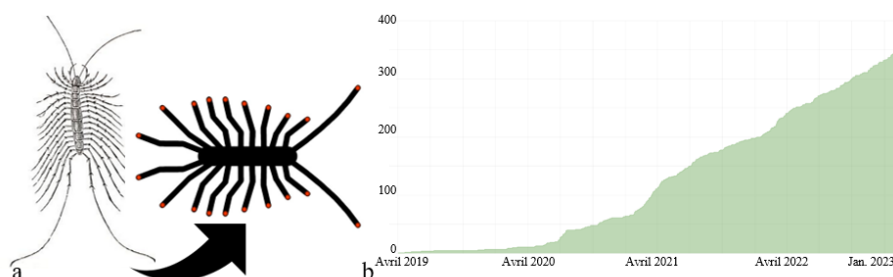


FIGURE 1. Logo CentipedeRTK (a) et évolution des contributeurs au réseau (b).

En outre, les pattes d'un mille-pattes peuvent s'adapter à différents terrains et obstacles, tout comme CentipedeRTK peut s'ajuster à différentes conditions environnementales pour maintenir une précision de positionnement maximale. Enfin, la coopération entre les pattes d'un mille-pattes assure une mobilité rapide et efficace, de même CentipedeRTK assure une fiabilité grâce à son réseau de contributeur : de 2 à 350 stations de référence installées en 4 ans (Fig. 1b).

CentipedeRTK, déployé à partir de 2019, peut être utilisé avec des composants matériels (récepteurs et/ antennes) spécifiquement conçus pour les besoins d'utilisation ou des composants matériels propriétaires. Il vise à éliminer les obstacles financiers au déploiement du GNSS RTK pour un grand nombre d'utilisateurs finaux, sans dépendre de systèmes propriétaires ou de l'abonnement à

un réseau de bases GNSS (Real Time Network - RTN). Il s'appuie sur des moyens mutualisés entre des agriculteurs, des instituts de recherche, des organismes publics et des entreprises privées pour l'installation de stations de référence. INRAE (en particulier l'équipe projet du CATI GEDEOP¹) assure la maintenance du serveur central CentipedeRTK, une partie de la production de la documentation et la formation des utilisateurs. La force de ce projet réside dans le fait que chacun peut devenir acteur du réseau en contribuant à l'installation autofinancée d'une station de référence grâce au projet RTKbase² ou en participant à la mise à jour de la documentation et à l'assistance aux utilisateurs. L'obtention d'une géolocalisation haute précision nécessite une combinaison de plusieurs composants ou segments dans le système. (Fig. 2) :

- une base GNSS de référence, fixe (Fig. 2.1);
- un rover GNSS, mobile (Fig. 2.2);
- un caster (Fig. 2.3), système de centralisation/diffusion des différentes corrections disponibles et de communication entre une base et un rover *via* le protocole NTRIP (Network Transport of RTCM *via* Internet Protocol).

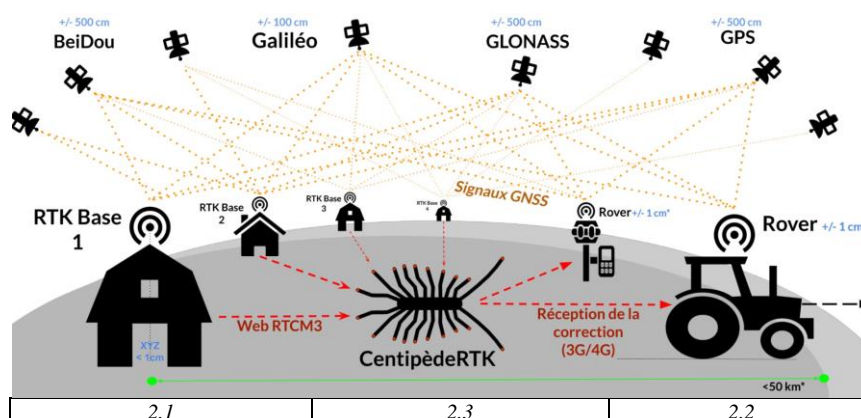


FIGURE 2. CentipedeRTK en 3 segments : 2.1 base ; 2.2 rover et 2.3 caster.

2.1. Le segment « base RTK »

Pour obtenir une géolocalisation de haute précision, il est nécessaire d'installer (ou d'utiliser) une base au sol que l'on fabrique simplement grâce au projet RTKbase. La position de l'antenne est connue avec une grande précision en trois dimensions (coordonnées X, Y et Z). Le calcul de cette position est réalisé par post-

¹ Cati Gedeop : Centre Automatisé de traitement de l'Information « GEstion des Données d'Expérimentations, d'Observations et de Pratiques sur les agro-socio-éco-systèmes », réseau d'informaticiens Inrae : <https://gedeop.cati.inrae.fr/>

² <https://github.com/Stefal/rtkbase>

traitement des données des satellites observables sur 24 heures et est effectué par rapport aux antennes de référence du réseau GNSS permanent de l'IGN³ en France, et par le service en ligne NRCAN qui utilise la méthode PPP (Precise Point Positioning) pour le reste du monde⁴. Une fois cette position définie, la station de base reçoit en continu les signaux des différentes constellations de satellites de positionnement et génère des trames de correction entre sa position réelle et sa position calculée pour chaque satellite. Le résultat obtenu est un flux de données RTCM⁵ (Radio Technical Commission for Maritime Services), qui est utilisé par tout autre récepteur GNSS se déplaçant à proximité de la base, appelé « rover », pour corriger sa position.

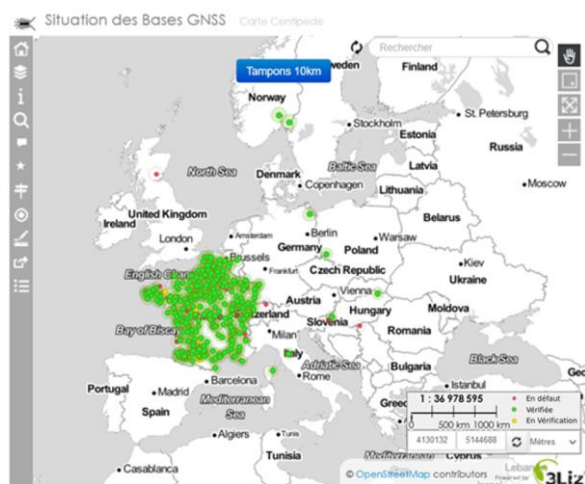
Selon les recommandations de CentipedeRTK, l'antenne de la station de référence devra être installée sur un support stable avec une vue dégagée vers le ciel (Fig. 3a), sans obstruer la vue en dessous de 10 degrés par rapport à l'antenne. Depuis 5 ans de mise en œuvre de CentipedeRTK, ce sont plus de 400 bases déployées en Europe avec la majorité sur le territoire français (Fig. 3b). Il est donc possible pour un nouvel utilisateur de CentipedeRTK d'utiliser une base existante active en recherchant celle la plus proche de sa zone d'étude *via* la consultation de la carte web des bases (Fig. 3c).



a



c



b

Lien à la carte consultable en ligne:

<https://centipede.fr/index.php/view/map/?repository=cent&project=centipede>

³ <https://rgp.ign.fr/STATIONS/antennes.php>

⁴ pour le Canada : Le service Positionnement ponctuel précis du Système canadien de référence spatiale (SCRS-PPP) a été lancé en 2003. Il permet aux utilisateurs de systèmes mondiaux de navigation par satellite (GNSS) de recueillir des données sur le terrain et de téléverser ces données dans des serveurs de Ressources naturelles Canada (RNC) https://webapp.csrscs-scrcs.nrcan-rncan.gc.ca/geod/tools-outils/sample_doc_files/V3/NRCan%20SCRS-PPP-v3_Tutorial%20FR.pdf

⁵ https://en.wikipedia.org/wiki/RTCM_SC-1041.

FIGURE 3. Schéma de montage d'une base CentipedeRTK (a) et carte web de localisation des bases RTK en Europe à la date du 25/01/2023, symbolisées en fonction de leur état de fonctionnement rouge, vert, orange (b) et identification et calcul de proximité des bases voisines du siège Inrae à Castanet avec les tampons concentriques de rayons à tous les 10km (c).

2.2. Le segment « rover »

Un rover RTK, appareil électronique mobile consacré à une variété d'utilisations GNSS, doit être facile à configurer pour l'utilisateur de terrain (Fig. 4a). Ce rover peut équiper un arpenteur, un tracteur ou un drone volant et permettre de le localiser avec une précision centimétrique ou proche du centimètre grâce aux trames de correction envoyées par la station de référence choisie et *via* l'algorithme présent dans la puce GNSS RTK. Il est important de noter que plus la distance à la base RTK choisie est grande, plus la précision sera affectée. On estime que l'on ajoute 1 mm d'incertitude en planimétrie pour chaque kilomètre supplémentaire dans l'éloignement de la base RTK. Il est possible de construire soi-même son propre « rover » RTK. La documentation en ligne fournit les instructions étape par étape pour le faire⁶. Par exemple, ce rover peut être composé d'une antenne GNSS mobile montée sur une canne d'arpentage, connectée à un module GNSS alimenté par une batterie et relié à une puce Bluetooth. Il est ensuite possible de connecter un smartphone et une application NTRIP Client au rover *via* Bluetooth pour traiter les trames de correction de la base la plus proche et récupérer la position géographique corrigée. Ainsi, toutes les applications utilisant la géolocalisation sur le smartphone bénéficieront d'une précision de localisation proche du centimètre. Il existe également des rovers RTK prêts à l'emploi et disponibles à un coût raisonnable (Fig. 4b).



FIGURE 4. Différents types de rover en fonction des cas d'utilisations (a) et rover Sparkfun équipant une canne (b).

⁶ https://docs.centipede.fr/docs/make_rover/

CentipedeRTK propose également, *via* sa carte en ligne des positionnements des bases du réseau, d'aider l'utilisateur à se localiser et à identifier la ou les bases les plus proches en utilisant l'outil « Tampons 10km ». Ce dernier trace à l'écran des cercles concentriques de rayon 10 km centrés sur la position de l'utilisateur (Fig. 3c). Il est à noter que certains clients NTRIP disposent également de la recherche de la station de référence la plus proche en fonction de sa position.

2.3. Le segment « caster »

Le système CentipedeRTK est conçu pour la diffusion en temps réel de corrections GNSS différentielles. Cette diffusion se fait par l'intermédiaire d'un serveur central appelé NTRIP Caster, qui utilise le protocole NTRIP pour diffuser les trames RTCM des stations de référence sur Internet *via* des réseaux cellulaires 3G ou 4G.

L'architecture logicielle de CentipedeRTK est basée sur des logiciels open source tels que BKG Caster, PostgreSQL/PostGIS, Lizmap et Grafana. BKG Caster est un serveur de diffusion de données RTK qui permet de transmettre en temps réel les corrections GNSS différentielles *via* Internet, jouant le rôle de passerelle entre les bases et les rovers. PostgreSQL/PostGIS est une plateforme de gestion de bases de données spatiales qui enregistre les informations des stations GNSS de références, mises à jour en continu. Lizmap est un outil de cartographie open source qui permet de visualiser les données géographiques et les métadonnées des stations, tandis que Grafana aide à comprendre graphiquement les performances et la qualité des données diffusées par les stations RTK. Ces différents logiciels open source permettent d'assurer la reproductibilité du projet, de gérer l'ensemble des données géographiques et de visualiser les performances du réseau CentipedeRTK. Ensemble, ils forment une architecture efficace pour la diffusion optimale des données RTK *via* Internet en connectant les stations RTK de manière cohérente.

L'infrastructure de CentipedeRTK est gérée par une équipe composée d'une dizaine d'administrateurs privés et publics. La communauté d'utilisateurs peut accéder à une documentation complète, ainsi qu'un forum et un fil de discussion sur Telegram regroupant actuellement 670 membres⁷. Les utilisateurs peuvent ainsi bénéficier de l'aide de la communauté et participer aux débats sur les développements futurs.

3. Résultats au travers de quelques cas d'utilisation

3.1. L'autoguidage d'un tracteur pour une agriculture de précision

CentipedeRTK a été largement adopté pour l'agriculture de précision en France, avec de nombreux agriculteurs ayant installés des bases RTK (65 % du réseau) chez eux afin de gagner en autonomie tout en réduisant leur budget dédié à la

⁷ https://t.me/Centipede_RTK

géolocalisation. Les agriculteurs ainsi que les conseillers en agriculture (coopératives, chambres d'agriculture) considèrent que le CentipedeRTK a deux avantages clés par rapport aux corrections satellitaires standards telles que EGNOS ou RTX de Trimble : une précision extrême de 1 à 2 cm à l'antenne, idéale pour les travaux les plus exigeants en combinaison avec un autoguidage, et une meilleure répétabilité de la position, permettant de revenir exactement au même endroit même après plusieurs mois d'intervalle (par exemple pour les semis et le binage)⁸.

La chambre d'agriculture de la Somme⁹ a par exemple installé en juin 2020 quatre bases de référence (Amiens, Abbeville, Villers Bocage et Estrées Mons) qui couvrent ce département. En 2021, elle a publié un rapport technique, accessible en ligne, sur l'évaluation de la qualité du RTK libre Centipede comparée au signal RTK du commerce Orphéon. La qualité du signal RTK a été définie par trois composantes: la précision, la répétabilité et la stabilité (Dupont, 2021) et l'autrice conclut en 2020 « Centipede est un signal RTK libre (accessible gratuitement) équivalent à un signal RTK du commerce comme Orphéon, leader du marché de corrections RTK, en terme de précision et répétabilité. La position donnée est à +/- 2cm de la position d'un RTK de référence (Orphéon) jusqu'à 20km autour de la base. Jusqu'à 30km on a une différence de +/- 3cm entre Orphéon et Centipede (...). La répétabilité d'Orphéon est d'en moyenne +/-5cm d'un mois sur l'autre et celle de Centipede (sur les bases situées à moins de 30km) en moyenne de +/-5.2cm soit équivalente à celle d'Orphéon. Les bases Centipede présentent également une bonne stabilité de connexion, les coupures sont rares et courtes. Les résultats sur la qualité du signal Centipede sont très encourageants ».

3.2. Une Balise Open Source pour le monitoring des océans : le cas du projet de recherche Physalia (UMR LIENSs¹⁰ CNRS La Rochelle et UE SLP Inrae, Saint-Laurent de la Préé)

La surveillance du niveau de la mer et des marais est un enjeu majeur dans la compréhension des dynamiques côtières et hydrologiques. Plusieurs études ont montré l'efficacité des bouées GNSS appliquées à la mesure du niveau de la mer. Toutefois, leur usage est généralement limité par le coût élevé d'un tel dispositif.

Trois prototypes de bouées GNSS-RTK ont été développés pour répondre à des problématiques liées à la surveillance temps réel des variations du niveau de la mer et des marais littoraux. La réalisation des bouées avec des composants semblables aux bases CentipedeRTK permet une réduction importante des coûts et s'élève à près de 600€ pièce (contre 8000€ pour un modèle propriétaire). Les bouées ont été construites autour d'une structure centrale imprimée en 3D et sont équipées de trois flotteurs montés sur des bras en acier inoxydable (Fig. 5a.). La réception des signaux

⁸ <https://centre-valdeloire.chambres-agriculture.fr/produire-innover/machinisme/dossier-technique-machinisme/le-rtk-centipede/> et multitudes de vidéos You Tube comme celle-ci: <https://www.youtube.com/watch?v=WhL2JaARfVk>

⁹ <https://hautsdefrance.chambre-agriculture.fr/chambre-agriculture-somme/>

¹⁰ <https://lienss.univ-larochelle.fr/> et <https://physalia.centipede.fr/>

GNSS est assurée par un récepteur bi-fréquences Ublox-F9P, utilisable avec les principales constellations GPS, GLONASS, Galileo et Beidou. Les corrections RTK sont obtenues par le réseau de bases CentipedeRTK, et traitées en temps réel par un processeur Raspberry-Pi. La liaison est assurée par un module 4G et deux antennes externes

Des essais réalisés sur l'île d'Aix (France) ont permis d'évaluer la précision des bouées en comparant les données obtenues avec celles d'un marégraphe radar. La comparaison est réalisée avec les données corrigées en temps réel et celles corrigées en post-traitement. Les résultats altimétriques obtenus montrent une erreur quadratique moyenne (RMSE) comprise entre 1,93 et 3,03cm validant ainsi l'utilisation des bouées GNSS-RTK low-cost pour un usage marégraphique (Pira *et al.*, 2022).

En 2022, deux nouvelles missions ont été menées pour estimer les pentes du géoïde terrestre pour la validation des missions de mesures altimétriques. Cette expérience a validé la méthodologie et la précision des bouées GNSS Physalia tout en offrant des résultats probants dans le domaine de la géophysique (Tranchant *et al.*, 2023 ; Fig. 5b.)

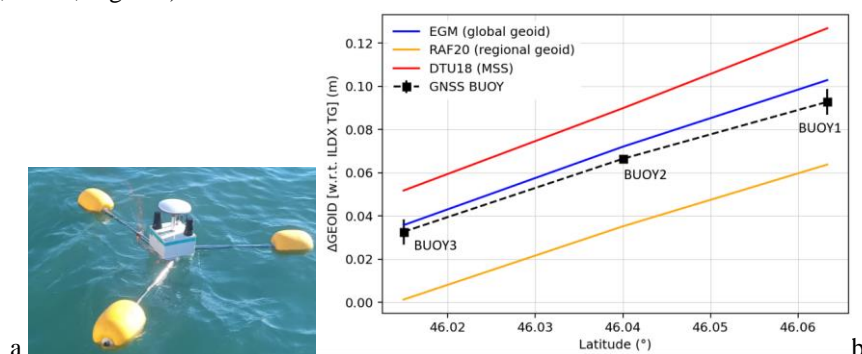


FIGURE 5. Bouée Physalia (a) et levés GNSS pour la cartographie du géoïde dans une mer macrotidale impliquant un USV marin, des bouées et modélisation hydrodynamique. EGM et DTU18 sont des modèles de géoïde global d'emprise mondiale alors que RAF20 est régional avec un pixel d'1km². Les bouées BUOY ont été placées à 3km les unes des autres validant la mesure haute précision de la pente de géoïde. (b) (Pertuis Charentais, France).

3.3. Localiser précisément les cavités basses des arbres pour y étudier la biodiversité hébergée : cas du projet de recherche BloBiForM (UMR Dynafor, INRAE)

La géolocalisation précise d'arbres en peuplement forestier est une demande récurrente en sciences forestières. Le projet BloBiForM débuté en 2020 vise à développer une palette d'outils pour évaluer les effets des choix d'aménagement d'un massif forestier sur la biodiversité. Il s'intéresse aux organismes inféodés aux

"dendro-microhabitats", des structures singulières des arbres telles que les cavités de tronc (Larrieu *et al.*, 2019 ; Fig. 6). Suivre et gérer la distribution spatiale des dendro-microhabitats à l'échelle du massif pourrait permettre de comprendre l'effet fort sur la biodiversité des espèces associées, et constituer un levier pour atteindre des objectifs de conservation. Cette hypothèse ne peut être testée qu'à l'aide d'analyses de grands jeux de données spatialisées à échelle fine.

Le recueil de telles données spatialisées est chronophage, et donc coûteux, quand il est mis en œuvre avec des moyens optiques : par exemple avec un théodolite ou en utilisant simultanément une station totale avec une boussole et un télémètre. Il est difficile d'intégrer de telles mesures dans un plan d'échantillonnage avec un grand nombre de placettes. De plus, ces moyens optiques fournissent une localisation relative à un point de référence dont il n'est pas aisé, en contexte forestier, de connaître la position exacte. Ils ne permettent pas non plus d'estimer la précision de la localisation.

L'enjeu est donc de géolocaliser des arbres au sein d'un peuplement forestier en disposant d'un système de positionnement des arbres plus rapide et préservant néanmoins des erreurs-types idéalement de l'ordre ou inférieures au mètre. Le système de positionnement par satellites (GNSS) apporte des réponses potentielles à ce besoin. Cependant le couvert forestier nuit fortement à la précision de la géolocalisation et ne permet pas en général d'obtenir une précision inférieure à une dizaine de mètres avec un système de positionnement par satellites conventionnel (Gaudin *et al.*, 2006 ; Lahaye et Ladet, 2014b). Nous avons donc utilisé la solution CentipedeRTK pour géolocaliser précisément ces arbres. Nous avons également montré que cette solution montre une performance au moins égale sinon supérieure à une solution standard sous licence propriétaire (Fig. 6).

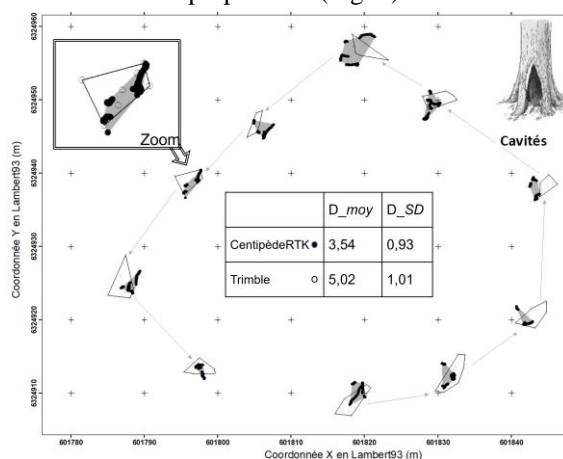


FIGURE 6. Visualisation des données de position d'arbres par le Trimble Geo 7x (cercle blanc) et par CentipedeRTK (cercle noir). Chaque amas de points correspond à un arbre, collectés lors de passages successifs le long du parcours. Les enveloppes englobantes des 2 nuages de points par arbre ont été obtenues par SIG (en noir sur les acquisitions CentipedeRTK et en blanc sur celles du Trimble). Les caractéristiques géométriques des enveloppes ont été calculées à savoir la

distance la plus longue entre deux sommets de l'enveloppe convexe notée D (moyenne et écart-type en mètres) pour caractériser la dispersion des points (Forêt de Grésigne, France).

4. Discussion et perspectives

Nous avons souhaité présenter dans cet article de manière synthétique l'intérêt du réseau CentipedeRTK (Fig. 7) et les détails matériels de son fonctionnement, et de montrer un éventail d'applications pour lesquelles ce réseau a fourni des performances satisfaisantes, en comparaison un système de positionnement par satellite (GNSS) conventionnel ou un système propriétaire avec correction différentielle temps réel (RTK) ou différé (DGSS).

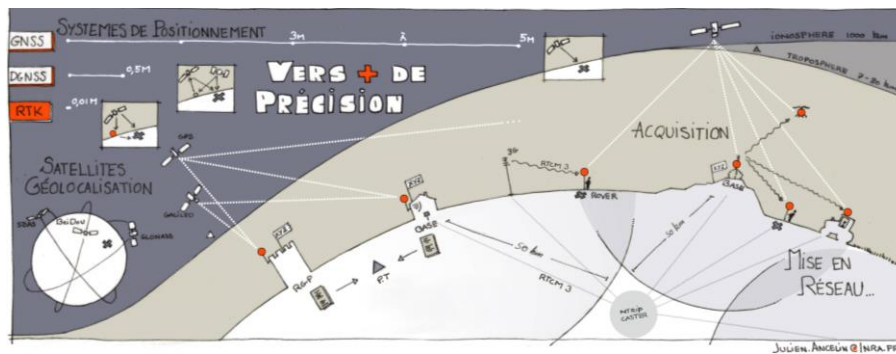


FIGURE 7. Fonctionnement du système de positionnement par satellite et comparaison des précisions de mesure entre GNSS, DGSS et RTK (famille à laquelle Centipede appartient).

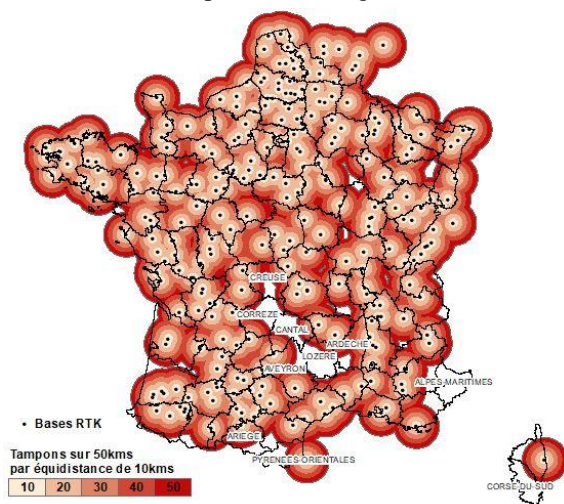
En complément nous avons produit une analyse spatiale de la couverture en France métropolitaine des bases existantes du réseau CentipedeRTK début 2023, en considérant deux tampons, celui à 30 kilomètres (signal peu dégradé du fait du rayon court) et celui à 50 kilomètres (signal éligible sur rayon maximal de réception) pour mieux évaluer le taux de couverture des départements français hors DROM-COM (départements et régions d'outre-mer et collectivités d'outre-mer) (Fig. 8). Depuis le réseau s'est déployé aussi dans l'Océan Indien et s'est densifié en France avec de nouvelles antennes récemment installées.

Ceci nous a permis :

- de repérer les départements français avec un faible taux de couverture en bases RTK (inférieur à 20% de la surface du département) comme les Alpes-Maritimes, la Corse du Sud où l'effort d'installation de nouvelles bases pourra être intensifié à court terme ;

– de repérer les départements français presque intégralement couverts par les bases existantes (supérieur à 80% de la surface du département) si l'on veut plutôt achever la couverture intégrale des départements ;

– d'identifier les départements français intégralement couverts par les bases existantes (100% de la surface du département) (Fig. 8).



Nombre de départements	taux_couverture<20%	taux_couverture>80%	taux_couverture=100%
tampon30kms	6	6	34
tampon50kms	2	35	52

FIGURE 8. Carte de la couverture France (hors Drom-Com) en bases RTK jusqu'à 50 kilomètres (d'après les positions des 303 bases actives le 30/01/2023 : <http://caster.centipede.fr:2101/>)

Selon nous, vu ce taux de couverture actuel en constante progression, CentipedeRTK peut ainsi être vu en tant que géocommun. Nous proposons ici l'idée de considérer CentipedeRTK comme un bien commun, en raison de son caractère fondamentalement coproduit. Les géocommuns sont des ressources produites collectivement par une communauté d'acteurs hétérogènes, tels que des secteurs publics et privés, des clients et fournisseurs, des individus et des organisations, aussi bien partenaires que concurrents. En moins de 5 ans, CentipedeRTK a rassemblé une communauté dédiée à l'amélioration du géopositionnement dont la mise en œuvre technique a été entièrement collaborative. La gouvernance de la communauté, un aspect fondamental des géocommuns, est claire et transparente, et elle inclut des règles inclusives de gouvernance et d'utilisation, ainsi que des chartes, des licences ouvertes et des accords de contribution. Ses caractéristiques sont bien celles discutées et listées collectivement en mai 2022¹¹, à savoir bien définir le périmètre du géocommun, ses règles de contribution et ses règles d'utilisation. CentipedeRTK

¹¹ <https://forum.geocommuns.fr/t/une-definition-partagee-dun-geo-commun/57/>

a d'ailleurs depuis peu un forum dédié sur la plateforme des géocommuns : <https://forum.geocommuns.fr/t/a-propos-de-la-categorie-rtk-centipede/544>

Les perspectives stratégiques de CentipedeRTK incluent l'optimisation de la performance technique autant que l'expansion de la communauté de contributeurs et d'utilisateurs. Elle vise également la mise en place de mécanismes de contrôle de qualité plus rigoureux, la promotion de technologies ouvertes reposant sur des logiciels et des composants matériels en libre accès, ainsi que le partage du savoir et la vulgarisation auprès des utilisateurs. Les objectifs clés sont la robustesse et la résilience du système, l'internationalisation, l'essaimage de collectifs de contributeurs et d'utilisateurs. Il s'agit également de renforcer le contrôle qualité et la supervision pour garantir une qualité de service optimale et la fiabilité des informations produites. Enfin, les développeurs de CentipedeRTK veillent, au travers d'une grande panoplie de supports de communication, à faciliter et accompagner l'utilisation de ces technologies complexes pour offrir des solutions accessibles à tous.

Remerciements

Nous remercions l'ensemble des collaborateurs et contributeurs pour leur engagement au profit de l'amélioration du réseau CentipedeRTK. Grâce à leur dévouement, le système a atteint des niveaux inespérés de performance et de fiabilité. Leur contribution a permis de démocratiser l'accès à la géolocalisation centimétrique, tout en renforçant la capacité du réseau à fournir des solutions GNSS fiables, pour une multitude d'applications scientifiques et industrielles. Nous souhaitons aussi remercier ici certains de nos soutiens financiers qui se sont mobilisés pour nous permettre d'aller présenter ce travail à la conférence francophone SAGEO édition 2023 à Québec : à INRAE, nos unités de recherche SLP et Dynafor, la direction pour la Science Ouverte DIPSO, le département de recherche Ecologie et biodiversité des milieux forestiers, prairiaux et aquatiques ECODIV.

Bibliographie

- Bosser P. (2012). *GNSS : Systèmes globaux de positionnement par satellite*, Support de cours en ligne, Bibliothèque de ressources pédagogiques de l'ENSG, École Nationale des Sciences Géographiques <http://cours-fad-public.ensg.eu/course/view.php?id=69>
- Bosser P. (2014), RTKlib, est une librairie open-source proposant un ensemble de programmes pour le positionnement standard et précis par GNSS, École Nationale des Sciences Géographiques, <http://cours-fad-public.ensg.eu/course/view.php?id=85#section-1>
- Dupont A. (2021) Evaluation de la qualité du RTK libre Centipede. Rapport de la Chambre d'agriculture de la Somme, Rapport Casdar : https://hautsdefrance.chambre-agriculture.fr/fileadmin/user_upload/Hauts-de-France/171_Inst-Hauts-de-France/10-Actualites/2021/Rapport_qualite_Centipede_CA80.pdf

- Duquenne F., Botton S., Peyret F., Bétaille D., Willis P. (2005). *GPS. Localisation et navigation par satellites*, éd. Hermès, Paris.
- Gaudin S., Pont S., Gantillon M. (2006) Le GPS : un outil efficace pour se diriger en forêt. *Revue forestière française*, 58 (2), pp.141-154. <https://hal.science/hal-03449284>
- Lahaye R., Ladet S. (2014)a. Les principes du positionnement par satellite : GNSS. *Les Cahiers des Techniques de l'INRA*, p. 9-18. <https://www6.inrae.fr/novae/Les-articles-parus/Les-n-Speciaux/GPS-SIG/chap1-ns-GPS-SIG-2014/Art01-ns-GPS-SIG-2014>
- Lahaye R., Ladet S. (2014)b. Les réseaux de correction différentielle. *Les Cahiers des Techniques de l'INRA*, p. 36-43. <https://www6.inrae.fr/novae/Les-articles-parus/Les-n-Speciaux/GPS-SIG/chap1-ns-GPS-SIG-2014/Art04-ns-GPS-SIG-2014>
- Larrieu L., Gosselin F., Archaux F., Chevalier R., Corriol G., Dauffy-Richard E., Deconchat M., Gosselin M., Ladet S., Savoie J. M., Tillon L., Bouget C. (2019). Assessing the potential of routine stand variables from multi-taxon data as habitat surrogates in European temperate forests. *Ecological Indicators* 104:116-26. doi: 10.1016/j.ecolind.2019.04.085.
- Pira A., Ancelin J., Coulombier T., Dausse D., Ballu V., Testut L., Mzali L., Gaugue A. (2022) Physalia : Plateforme HYdrographique pour la Surveillance Altimétrique du Littoral. *Lettre d'information Résif*, pp.13-14. <https://hal.inrae.fr/hal-03508116v2>
- Tranchant Y.T., Ballu V., Testut L., Ancelin J., Coulombier T., Dausse D., Kerebel V., Chupin C. (2023) GNSS Survey for Geoid Mapping in a Macrotidal Sea Involving Marine USV, Buoys and Hydrodynamic Modeling (Pertuis Charentais, France) 13th Coastal Altimetry Workshop, Coastal Altimetry Training, 6-10 February 2023, Universidad de Cádiz, Spain.

Unlocking Insights into Business Trajectories with Transformer-based Spatio-temporal Data Analysis

Muhammad Arslan¹, Christophe Cruz²

1. *Laboratoire Interdisciplinaire Carnot de Bourgogne (ICB),*

9 Av. Alain Savary, 21000 Dijon, France

muhammad.arslan@u-bourgogne.fr

2. *Laboratoire Interdisciplinaire Carnot de Bourgogne (ICB),*

9 Av. Alain Savary, 21000 Dijon, France

christophe.cruz@u-bourgogne.fr

ABSTRACT. The world of business is constantly evolving and staying ahead of the curve requires a deep understanding of market trends and performance. With the increasing availability of text data, it has become possible to gain insights into the health and trajectory of businesses and help decision-makers to make informed decisions. This work is based on the application of the Transformer-based spatio-temporal data analysis for analyzing vast amounts of business-related text data. The proposed method processes news articles that were retrieved from multiple sources from 2017 to 2021 inclusive. It categorizes them into different business categories (e.g. manufacturing, healthcare, construction, etc.) based on the company's interest. A company's taxonomy is used to understand their domains of interest. It then identifies the main topics in each article, which give us information about the events (e.g. launch of a company, regulatory changes, etc.) and their spatio-temporal evolution over time. The insights generated from this analysis can be used by businesses to make informed decisions, by investors to identify opportunities and assess risk, and by policymakers to understand the impact of regulations and economic policies on the business community.

KEYWORDS: MULTI-LABEL TEXT CLASSIFICATION, TOPIC MODELING, TRANSFORMERS, BUSINESS TRAJECTORIES

1. Introduction

Transformer-based spatio-temporal data analysis is a cutting-edge approach that leverages the power of machine learning and natural language processing to analyze vast amounts of business-related text data (Fan et al. 2022). This approach allows us to not only analyze the performance of businesses over time but also understand how trends and performance vary across geographic territories. By combining data analysis with the latest advancements in natural language processing, we can gain a comprehensive view of business trends (Braşoveanu and Andonie 2020). It offers a powerful tool for unlocking insights into business trajectories, providing valuable information for businesses, investors, and policymakers.

To perform business data analysis, we need to develop a news data analyzer (Alawadh et al. 2023). A news data analyzer refers to a system that processes and analyzes news articles to extract relevant information and insights (Lau et al. 2021). The goal of a news data analyzer is typically to gain a better understanding of the events and trends in the news and to make sense of the large amounts of news data that are generated daily. A news data analyzer can use various techniques and algorithms (Ashtiani and Raahmei 2023), such as Natural Language Processing (NLP), Machine Learning (ML), and Data Mining, to process news articles and extract valuable information such as topics, etc. The analyzer can then present this information in a meaningful and useful way, such as through visualizations. The specific features and capabilities of a news data analyzer can vary depending on the use case and requirements, but the overall goal is to help users make sense of the news data and gain valuable insights and information.

In our case, a news data analyzer is a system that combines transformer-based text classification (Cunha et al. 2022) and dynamic topic models (Grootendorst 2022) and classifies news articles into different business categories. Business categories refer to different types of businesses (e.g., healthcare, transportation, agriculture, construction, and real estate) that exist in the market. For this study, the domains were shortlisted based on the company's interest. After categorizing the business news, it identifies the main topics in each text. In the context of business, many different topics can be covered in news articles. For instance, company news and performance, which deals with updates on the performance and financial results of individual companies, as well as news on mergers and acquisitions, partnerships, and other business developments. The insights generated by understanding the spatio-temporal evolution of these topics will help to make informed decisions.

This paper is organized as follows. Section 2 reviews the background of transformer-based text classification and dynamic topic models. Section 3 introduces the proposed method. The discussion is mentioned in Section 4 and the conclusion is described in Section 5.

2. Background

Text classification is a task in natural language processing (NLP) where the goal is to assign predefined labels to a given text based on its content (Kowsari et al. 2019). The traditional approach to text classification was based on hand-engineered features, such as the presence of certain keywords, and simple machine learning algorithms like Naive Bayes (Webb et al. 2010). With the advent of deep learning, the focus shifted towards end-to-end models that could automatically extract useful features from the raw text. One of the most successful such models is the Transformer, introduced in 2017 (Rothman 2021).

Transformers are a type of neural network architecture that use self-attention mechanisms to process input sequences in a parallel manner (Rothman 2021). In the context of NLP, they have proven to be highly effective at capturing long-range dependencies and semantic relationships between words in a sentence, which is essential for accurate text classification. Transformers have been applied to various NLP tasks, including text classification (Bhardwaj et al. 2021). In text classification, the input text can be assigned one or multiple labels (Liu et al. 2021). If a text can be assigned multiple labels, the task is known as multi-label text classification. Transformers can be applied to both single-label and multi-label text classification tasks. In the case of multi-label text classification, the output of the model is a binary vector, with each dimension corresponding to a label (Liu et al. 2021). The value of each dimension is either 0 or 1, indicating whether the text belongs to that label. To train a Transformer for multi-label text classification, a binary cross-entropy loss is used (Hande et al. 2021). The model is trained to predict the probability of each label, and the binary cross-entropy loss measures the difference between the true label and the predicted probability.

Dynamic topic models (DTMs) are a type of generative probabilistic model that can be used to analyze and track the evolution of topics in large collections of text documents over time (Grootendorst 2022). In traditional topic models, such as Latent Dirichlet Allocation (LDA), topics are treated as fixed latent variables that generate the words in the documents (Blei et al. 2003). However, in dynamic topic models (e.g., BERTopic), topics are assumed to evolve over time, reflecting the changing interests and themes of the underlying text corpus. DTMs typically represent each document as a mixture of topics, and each topic is modeled as a probability distribution over words (Grootendorst 2022). The model then assigns topics to the words in each document, and the topic distributions for each document are used to infer the topic evolution over time. DTMs are a flexible and powerful tool for analyzing text data over time and can be used to identify trends, track the spread of ideas and information, and understand how topics change and evolve (Yao and Wang 2022). They have been applied to a wide range of text-based applications (Savin and Konop 2022), including social media analysis, news analysis, and scientific literature analysis (Sestino and De Mauro 2022).

In the existing literature, the intersection of transformer-based models and dynamic topic models (DTMs) has not been thoroughly investigated in the domain of business analysis. This research aims to fill this gap by exploring the benefits of

combining these two approaches for business analysis. The objective of this study is to show the efficacy of combining transformer-based models and DTMs for multi-label text classification and topic extraction, and to demonstrate the potential applications of this combination in the realm of business analysis.

3. Proposed Method

The proposed news data analyzer consists of four key stages:

- 1) Text preprocessing involves cleaning, transforming, and preparing text data for spatio-temporal analysis.
- 2) Text classification involves categorizing text data into predefined categories based on their content and characteristics.
- 3) Dynamic topic modeling involves identifying and tracking the evolution of topics over time within a text corpus.
- 4) Data visualization is the final stage, where the results of the previous stages are presented in an intuitive and graphical manner to facilitate understanding and interpretation of the business trajectories and their evolution over time and geographic regions.

3.1. Text preprocessing

It is a crucial step in the NLP pipeline that involves cleaning and transforming raw data into a format suitable for analysis. The goal of preprocessing is to improve the quality and reliability of the data and to facilitate the analysis by reducing the amount of noise and irrelevant information. The choice of preprocessing techniques depends on the specific dataset and the problem being solved, and it can have a significant impact on the performance of the final model. For our case, a dataset of 26,509 news articles is used. This dataset is manually labelled using 1,800 different categories based on the company's taxonomy. An example is shown in Fig. 1. The text is classified with labels. These labels correspond to the labels of the taxonomy. The news articles were preprocessed, and a corpus of around 8,128,981 words is obtained after preprocessing the news articles. The system implements several preprocessing functions for text data in the English language. It starts by importing the required libraries including NLTK, RE, and String. It first removes punctuation and digits from the text using regular expressions. Then, it converts the text to lowercase, tokenizes the text into words, removes stop words and duplicates, and removes small words. The final step is to join the preprocessed words into a single text to feed the classification model.

	LabelID	LabelDefinition
company building ventelle transport area	25276	transport and logistics
specialize corbehem consider logistics employee	112	construction of premises
move want brebières second built site base	4716	moving house
['112', '4716', '25276']		

FIGURE 1. Text labelled using the taxonomy of the company.

3.2. Multi-label text classification

As each text is labelled with multiple labels (see Fig. 1) in the dataset, the case of multi-label classification is followed. After preprocessing the text, a transformer-based multi-label text classification model is executed. The model uses a multi-layer neural network architecture and is trained using the binary cross-entropy loss function and the Adam optimizer. The code imports several libraries, including Numpy and Pandas for data processing and manipulation, and Keras, a high-level deep learning library, for building and training the model. The model starts by loading the preprocessed text data using Numpy. This data is then tokenized using the Tokenizer class from Keras. Later, the tokenized text data is converted into numerical sequences. The sequences are then padded to make all sequences the same length. This is done so that all the input data has the same shape, which is required for training deep learning models. The data is then split into training and validation sets using Numpy slicing. The split index is calculated as 80% of the total number of sequences.

Next, the deep learning model is defined using the Keras functional API. The model consists of several layers, including an input layer, an embedding layer, a global max pooling layer, a dense layer with a Rectified Linear Unit (ReLU) activation function, and a dropout layer. The dropout layer helps to prevent overfitting of the model by randomly dropping out a fraction of the neurons during each training step. The final layer is a dense layer with a sigmoid activation function, which produces binary output values. The model is then compiled, where the loss function, optimizer, and evaluation metrics are specified. The loss function used is binary cross-entropy, which is suitable for binary classification problems. The optimizer used is the Adam optimizer with a learning rate of 0.001. The evaluation metric used is accuracy and precision. Finally, the model is trained using the *fit* method. The training data is used along with several other parameters such as batch size and number of epochs. The validation data is also used to evaluate the performance of the model.

3.3. Dynamic topic modeling

To use BERTopic, a transformer-based dynamic topic modeling approach, we first convert the preprocessed news text into numerical data using the "sentence-

transformers" python package. This package converts text data into 512-dimensional vectors, providing document-level embeddings. We set the language to English since our corpus is in English. During training of the BERTopic model, topic probabilities are calculated. Once training is completed, the model outputs 66 business topics, each represented by the 10 most probable words (see Fig. 2).

To visualize the evolution of topics in news articles over time, we use the "*topic_model.topics_over_time*" function, which requires the preprocessed news text corpus (docs), acquired topics (topics), timestamps for each article, global tuning for averaging topic representation with global topic representation, evolution tuning for averaging topic representation with the previous time's representation, and the number of bins for timestamps (nr_bins). The input values result in a plot (Fig. 3) showing the evolution of business topics over the past 5 years.

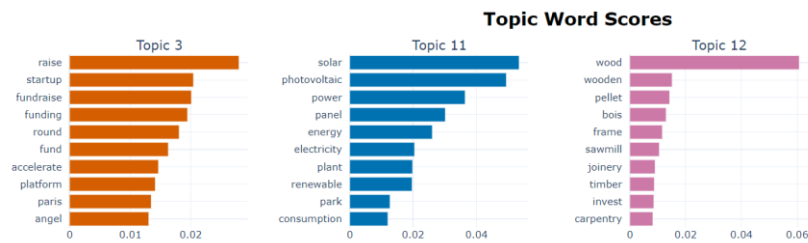


FIGURE 2. Different words representing different topics.

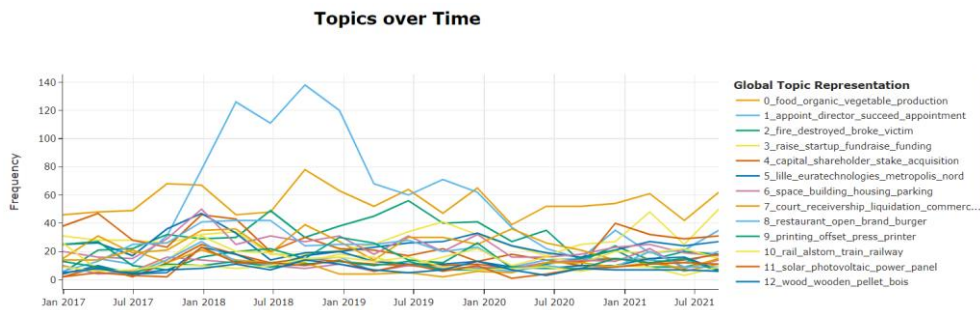


FIGURE 3. The evolution of different topics over time.

3.4. Visualizations of business trajectories

The process of constructing business trajectory visualizations involved first performing the multi-label classification of news text. Once the text is categorized and labelled, it is important to extract the topics for understanding different types of business events. For instance, the launch of the solar energy industry represents a significant development in the field of renewable energy. This business event signifies the introduction of a new player in the market, offering innovative

solutions for harnessing the power of the sun. The event is typically categorized under topic 11 of renewable energy and can be identified using keywords such as "solar," "power," "energy," "park," and others that are relevant to the industry (see Fig. 2).

It is crucial to analyze the volume (i.e., number) of news articles related to the launch of the solar energy industry. This information can provide valuable insights into the origin and evolution of the business event. By examining the number of articles published over different time periods, we can determine the level of public interest and media coverage surrounding the event. Additionally, tracking the geographic regions that the articles pertain to can give us an idea of the location and spread of the industry's development. This information can help us to identify the key players and driving forces behind the growth of the solar energy industry, providing valuable insights into the market and its potential for future growth. This information regarding the volume of news articles (see Fig. 4) was then utilized to create a business trajectory visualization (see Fig. 5) to show how the business topic evolved over time across different geographic regions for a better understanding of the overall business event.

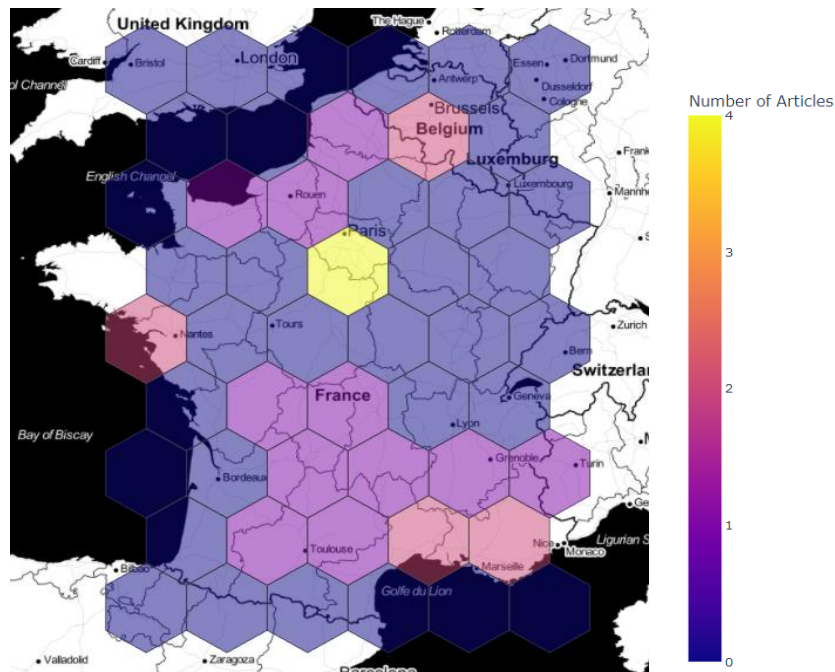


FIGURE 4. Number of articles published on a topic.

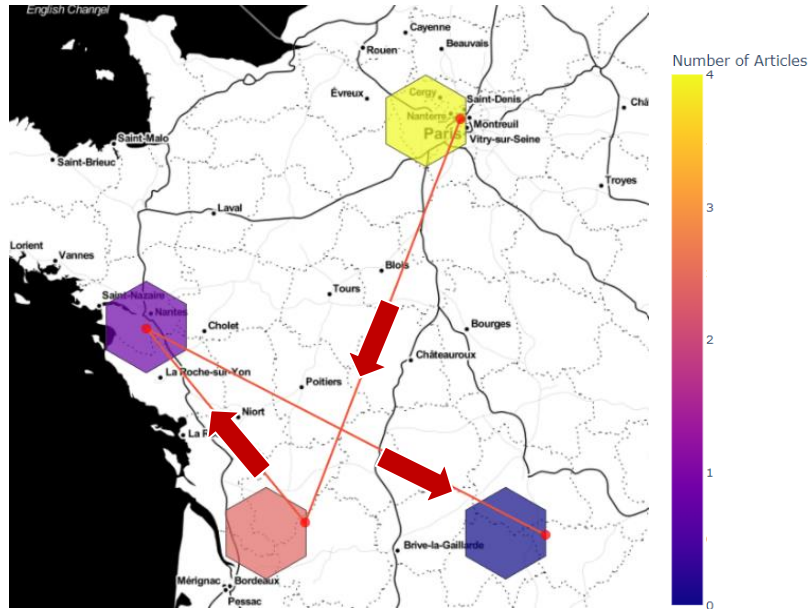


FIGURE 5. Business trajectory across geographic regions

4. Discussion

When using Transformers for multi-label text classification, there are some important considerations observed during this study. Most importantly, it is about class imbalance and label correlations. It is raised because of a huge number of labels (1800 in our case) and a small number of text examples of each label in a dataset. Class imbalance refers to the situation where some labels are much more frequent than others and can lead to a model that is biased towards the more frequent labels. To address the class imbalance, various techniques can be used, such as oversampling the minority class or using weighted loss functions. Label correlations refer to the situation where the presence of one label is often predictive of the presence of another label. This can be addressed by using techniques such as label powerset or label embedding, which represent the labels as a combination or embedding of multiple binary dimensions, respectively. However, these aspects concerning class imbalance and label correlations are not dealt with in this study.

Apart from that, visualizing business trajectories over time and geographic regions can provide several benefits (Arslan and Cruz 2023), which are:

- 1) *Understanding trends*: Visualizing the trajectory of a business over time helps to identify trends and patterns in its development, which can be useful for decision-making and planning.

- 2) *Spotting opportunities and challenges*: By seeing the business trajectory over time and space, it becomes easier to spot opportunities and challenges that the business may face. This information can be used to make informed decisions.
- 3) *Improved communication*: Visualizing business trajectories can help improve communication and collaboration among stakeholders, as it provides a clear and concise way to present information.
- 4) *Increased awareness*: By creating visual representations of business trajectories, it can increase awareness and understanding of the business among stakeholders.
- 5) *Enhanced analysis*: Visualizing the business trajectory over time and space allows for a more in-depth analysis of the business, as it provides a comprehensive view of its development and performance.

5. Conclusion

The application of Transformer-based spatio-temporal data analysis to business-related text data has proven to be an effective way to gain valuable insights into the health and trajectory of businesses. This method processes news articles and categorizes them into different business categories based on a company's interest. Later, the topics are extracted to understand different types of businesses over time. Visualizing hot and cold topics (referring to the current popularity and level of interest in a particular category) as business trajectories can greatly enhance and streamline business processes. By keeping track of these hot and cold topics, businesses can stay informed about the current state of the market and make informed decisions that align with current trends and consumer interests. This not only helps to identify potential opportunities for growth and innovation but also helps to mitigate any potential risks or challenges that may arise. Additionally, having a clear understanding of the current trends and topics can also inform and shape the direction of a company's marketing and communication strategies, allowing them to better connect with their target audience and stay relevant in the market.

Acknowledgements

The authors thank the French company FirstECO for providing the taxonomy and the French government for the plan France Relance funding.

References

- Alawadh, H. M., Alabrah, A., Meraj, T., & Rauf, H. T. (2023). Attention-Enriched Mini-BERT Fake News Analyzer Using the Arabic Language. *Future Internet*, 15(2), 44.

- Ashtiani, M. N., & Raahmei, B. (2023). News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Systems with Applications*, 119509.
- Bhardwaj, R., Majumder, N., Poria, S., & Hovy, E. (2021). More identifiable yet equally performant transformers for text classification. *arXiv preprint arXiv:2106.01269*.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Braşoveanu, A. M., & Andonie, R. (2020, September). Visualizing transformers for nlp: a brief survey. In *2020 24th International Conference Information Visualisation (IV)* (pp. 270-279). IEEE.
- Cunha, W., Viegas, F., França, C., Rosa, T., Rocha, L., & Gonçalves, M. A. (2022). A Comparative Survey of Instance Selection Methods applied to NonNeural and Transformer-Based Text Classification. *ACM Computing Surveys*.
- Fan, R., Li, J., Song, W., Han, W., Yan, J., & Wang, L. (2022). Urban informal settlements classification via a transformer-based spatial-temporal fusion network using multimodal remote sensing and time-series human activity data. *International Journal of Applied Earth Observation and Geoinformation*, 111, 102831.
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
- Hande, A., Puranik, K., Priyadarshini, R., Thavareesan, S., & Chakravarthi, B. R. (2021, April). Evaluating pretrained transformer-based models for COVID-19 fake news detection. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 766-772). IEEE.
- Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4), 150.
- Lau, C. Y., Yuen, M. C., Yiu, K. S., Lau, H. S., & Lo, C. T. (2021, February). Cat's Eye: Media Insights Analyzer. In *2021 4th International Conference on Big Data and Education* (pp. 80-87).
- Liu, S., Zhang, L., Yang, X., Su, H., & Zhu, J. (2021). Query2label: A simple transformer way to multi-label classification. *arXiv preprint arXiv:2107.10834*.
- Arslan, M. & Cruz, C. (2023). Modeling semantic business trajectories of territories for multidisciplinary studies through controlled vocabularies. *Data-driven Smart Cities (DASC)*. 39th IEEE International Conference on Data Engineering (ICDE 2023) Anaheim, California, USA.
- Rothman, D. (2021). *Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more*. Packt Publishing Ltd.
- Savin, I., Ott, I., & Konop, C. (2022). Tracing the evolution of service robotics: Insights from a topic modeling approach. *Technological Forecasting and Social Change*, 174, 121280.
- Sestino, A., & De Mauro, A. (2022). Leveraging artificial intelligence in business: Implications, applications and methods. *Technology Analysis & Strategic Management*, 34(1), 16-29.

- Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve Bayes. *Encyclopedia of machine learning*, 15, 713-714.
- Yao, F., & Wang, Y. (2020). Tracking urban geo-topics based on dynamic topic model. *Computers, Environment and Urban Systems*, 79, 101419.

Estimer un potentiel de regroupement de mobilité à partir de données spatio-temporelles

Louissette Garcin¹, Didier Josselin¹, Sonia Chardonnel²,
Mathilde Vernet³, Jean-Baptiste Chesneau⁴

1. UMR ESPACE, CNRS, Avignon Université

louisette.garcin@almuni.univ-avignon.fr; didier.josselin@univ-avignon.fr

2. UMR PACTE, CNRS, Université Grenoble Alpes

sonia.chardonnel@univ-grenoble-alpes.fr

3. UPR LIA, Avignon Université

mathilde.vernet@univ-avignon.fr

RÉSUMÉ. Cette communication présente une méthodologie d'estimation du potentiel de regroupement d'autosolistes au cours de leurs pérégrinations quotidiennes, basée sur des données spatio-temporelles issues des Enquêtes Ménages Déplacements. Dans un premier temps, il s'agit de regrouper des individus, en fonction de leurs trajectoires spatio-temporelles précises et d'ainsi obtenir un premier potentiel de regroupement théorique. Puis, dans un second temps, il est question d'associer l'analyse des trajectoires de pérégrinations aux caractéristiques sociales individuelles des autosolistes, dans le but d'obtenir un potentiel de transfert modal plus proche du réel. Des indicateurs, notamment environnementaux, sont proposés afin de mesurer les impacts du regroupement d'autosolistes dans le cadre de pérégrinations journalières.

ABSTRACT. This communication presents a methodology estimating the potential for regrouping solo drivers during their daily travels, based on spatiotemporal data from Household Travel Surveys. In a first step, the goal is to group individuals, based on their precise spatiotemporal trajectories and thus to obtain a first theoretical potential regrouping. Then, in a second step, the aim is to associate the analysis of travel trajectories with the individual social characteristics of solo drivers, in order to obtain a more realistic modal transfer potential. Indicators, including environmental ones, are proposed in order to measure the impacts of solo drivers regrouping in the context of daily travels.

MOTS-CLÉS : mobilité, regroupement, trajectoires spatio-temporelles

KEYWORDS: mobility, regrouping, spatiotemporal trajectories

1. Introduction

L'impact des transports dans la lutte contre le réchauffement climatique est à considérer avec acuité : il constitue un levier essentiel pour réduire les émissions de polluants atmosphériques. En effet, les mobilités sont à elles seules responsables de plus de 30% des émissions de gaz à effet de serre en France en 2020 (CITEPA, 2022). Processus complexe du quotidien, la mobilité individuelle demeure la principale contributrice à l'émission de polluants atmosphériques et constitue un enjeu majeur dans la lutte contre le réchauffement climatique.

La mobilité quotidienne relève de l'ensemble des enchaînements temporels et spatiaux nécessaires à la pratique d'une activité habituelle. Elle est façonnée par de multiples critères tels que le motif de déplacement, les horaires dédiés, la localisation des aménités, l'itinéraire emprunté ou encore le mode de transport choisi. La réduction des impacts environnementaux liés à la mobilité quotidienne relève alors de la prise en compte de l'ensemble des chaînes de déplacements effectuées au cours d'une journée, autrement dit, de l'étude des pérégrinations (Wiel, 2002).

En effet, l'organisation d'une mobilité durable se doit de prendre en compte l'ensemble des pérégrinations individuelles, dans le but d'aller au-delà de préconisations générales, ne prenant pas en compte la complexité des déplacements quotidiens (comme c'est le cas dans la plupart des études concernant les migrations domicile-travail). Ainsi, cette prise en compte de la mobilité comme un ensemble complexe, permet d'œuvrer directement à l'échelle individuelle, accroissant le potentiel d'adhésion à un changement des pratiques.

Cette communication a pour principale ambition d'identifier des potentiels d'amélioration des mobilités au sein des pérégrinations journalières, notamment par le regroupement spatio-temporel d'individus. Nous avons choisi de nous intéresser spécifiquement aux individus autosolistes (désignant les individus utilisant principalement leur voiture personnelle comme mode de transport), puisqu'étant particulièrement impliqués dans les émissions de gaz à effet de serre. En effet, dans le cadre des trajets quotidiens, l'autosolisme reste le mode de transport le plus populaire, atteignant plus de 70% de choix modal dans le cas des trajets domicile-travail (INSEE, 2021).

2. Méthodologie

2.1. *Limiter les émissions de gaz à effet de serre par le regroupement spatio-temporel d'individus*

2.1.1. *Le transfert modal comme outil de lutte contre l'autosolisme*

Lutter contre l'autosolisme nécessite de développer des alternatives à l'utilisation de véhicules particuliers. Le transfert modal envisagé, qui consiste en un report d'une partie des flux d'un mode de transport vers un autre (Géoconfluences, 2020), s'inscrit dans trois principales catégories de transports : les modes collectifs (bus, métros, tramways, trains, navettes fluviales, etc.), les modes doux (marche, vélo, trottinettes, objets de glisse urbaine, etc.) ainsi que les modes partagés (covoiturage, autopartage, transports à la demande, etc.). L'usage de chacun de ces modes par rapport à la voiture personnelle est généralement bénéfique à la fois pour l'environnement (notamment la réduction de l'émission de GES), l'économie (la voiture demeure l'un des modes les plus onéreux) et possède une dimension sociale avérée (l'inclusion de tous).

Ainsi, de nombreux territoires, politiques publiques et entreprises (à travers leur Responsabilité Sociétale) ont investi dans l'amélioration de l'offre de transports publics, et incitent au développement des modes décarbonés, afin de déclencher un report modal des individus mobiles (Kaufmann, 2003 ; Ministère de la transition écologique, 2021). En dépit de ces efforts, le trafic automobile ne diminue pas de manière significative (Guidez, 2002 ; Kaufmann, 2003 ; INSEE, 2021 ; Salomon *et al.*, 1993). Les démarches incitatives actuelles des politiques publiques concernant les mobilités alternatives à la voiture personnelle n'étant pas efficaces, le report modal constitue aujourd'hui un véritable enjeu, notamment dans le cadre de la réduction des émissions de gaz à effet de serre.

2.1.2. *Le regroupement d'individus mobiles*

L'autosolisme permettant une incontestable liberté spatio-temporelle, nous avons fait le choix, dans cette étude, de considérer le report modal vers un mode partagé relativement similaire, le covoiturage. Le covoiturage correspond au déplacement simultané d'individus à bord d'un même véhicule privé (Garcin, Josselin, 2021). De surcroît, les autosolistes disposant inévitablement d'une voiture personnelle, le transfert modal vers le covoiturage peut facilement être mis en place.

Regrouper des individus mobiles revient à rassembler des individus aux pratiques de mobilités semblables. Une fois identifiés, ces individus peuvent alors se réunir à bord d'un même véhicule, augmentant son taux d'occupation et réduisant mécaniquement les émissions de gaz à effet de serre rapportées à l'individu.

Concrètement, le regroupement d'individus consiste en l'identification de fe-

nêtres spatio-temporelles au sein desquelles plusieurs individus ont des comportements de mobilité similaires. De plus, les freins principaux au covoiturage relèvent de la complexité de se coordonner, du sentiment de perte de temps lié à de potentiels détours, mais aussi de divergences concernant les horaires de rencontre (Godillon, 2021). Pour pallier à ce problème, cette étude tend à regrouper uniquement des individus ayant des lieux d'origine et de destination strictement identiques, ainsi qu'une variabilité temporelle minimale, de l'ordre de 5 min, quant aux horaires de déplacements.

Enfin, la "partageabilité" de trajets en covoiturage est soumise à plusieurs critères comme le présentent Colin et Branguier ou encore Brahmi dans leurs travaux respectifs (Colin, Brangier, 2021 ; Brahmi *et al.*, 2022). En effet, certaines variables comme l'âge du covoitureur, son sexe, son niveau de vie ou encore le motif¹ du déplacement ont un impact direct sur le transfert effectif de l'autosolisme vers le covoiturage. Les individus ont tendance à covoiturer particulièrement avec des individus qui s'avèrent être leurs semblables. Aujourd'hui, devant l'urgence climatique, il faut passer outre les freins sociologiques et activer tous les leviers possibles pour faire basculer la plus grande partie de la population vers les transports publics et partagés. D'où notre estimation d'un potentiel théorique qu'il convient de fixer comme un objectif atteignable.

Cette communication a également pour objectif d'étudier les variabilités de regroupement suivant différents critères décrivant les individus.

Le regroupement spatio-temporel d'individus, prenant en compte des variabilités individuelles précises, nécessite alors de disposer de données précises.

2.2. Données traitées

L'étude des pérégrinations individuelles nécessite de disposer d'une enquête centrée sur le déplacement comme le décrit Commenges (Commenges, 2014), dans le but d'étudier rigoureusement les potentiels de regroupements d'individus ayant des comportements de mobilité similaires.

Nous choisissons de mobiliser les Enquêtes Ménages Déplacements (EMD) qui ont pour ambition de comprendre les comportements de mobilité des ménages au sein d'un territoire donné en France. Ces enquêtes sont pilotées par le CEREMA (anciennement CERTU) et permettent de recueillir plusieurs types d'informations liées à chaque déplacement dans une journée. Elles sont collectées par entretiens en "face à face" et par voie téléphonique. Les EMD recensent tous les déplacements des individus de ménages sélectionnés à l'échelle de la

1. Nous considérons huit motifs principaux de déplacement dans le cadre de notre étude : domicile, travail, études, achats, loisirs, démarches administratives ou de santé, accompagnement. Les motifs restants sont regroupés dans la catégorie "autres motifs".

"Zone Fine", un découpage spatial homogène en termes de taille de population et de forme d'habitat (semblable à la logique de découpage en IRIS). De plus, des informations comme le lieu d'origine, celui de destination, le motif, le mode de transport utilisé y sont présentées, ainsi que les caractéristiques sociales de l'individu concerné et de son ménage.

L'enquête repose sur une méthodologie d'échantillonnage standardisé au niveau national, pour assurer la comparabilité des résultats à la fois d'un point de vue spatial, mais également dans le cadre d'études diachroniques (Les partenaires de l'EMD, 2010 ; SMTTC, 2010). Les Enquêtes Ménages Déplacements portent sur des échantillons tirés aléatoirement, représentatifs des ménages des aires enquêtées.

Dans le cadre de notre étude, nous avons mobilisé les Enquêtes Ménages Déplacements de la ville de Grenoble en 2010 et de Marseille en 2009, recensant respectivement 7 600 et 9 750 ménages interrogés.

2.3. Traitement des données

2.3.1. Regroupement d'individus partageant des trajectoires spatio-temporelles similaires (RST)

L'identification de groupes d'individus partageant les mêmes trajectoires spatio-temporelles s'est effectuée par agrégation de leur localisation d'origine et de destination, ainsi que leurs horaires de départ et d'arrivée. Cela nous a alors permis de générer des regroupements basés sur le seul critère de trajectoires spatio-temporelles identiques.

Cependant, comme présenté précédemment, le potentiel réel de report modal s'avère plus faible que ce modèle théorique, puisque ne prenant pas en compte les variables à la fois liées au déplacement ainsi qu'aux individus les pratiquant. Il s'avère donc nécessaire d'étudier les variations au modèle théorique que pourrait entraîner la prise en compte des caractéristiques complémentaires individuelles. Nous considérons *in fine* un double potentiel : minimum (sous contraintes sociales de pratiques de mobilités) et maximum (potentiel théorique spatial de regroupement).

2.3.2. Regroupement d'individus partageant des trajectoires spatio-temporelles et de profils similaires (RSTP)

Les variables retenues dans l'analyse complémentaire des potentiels de regroupement sont liées au motif du déplacement, à l'âge de l'individu, à son sexe

ainsi qu'au niveau de vie². Les variables retenues sont alors à la fois qualitatives et quantitatives.

Plusieurs techniques statistiques s'offrent à nous, dans le but de prendre en compte des caractéristiques spécifiques dans le regroupement d'individus mobiles. Nous avons choisi de développer une Classification Ascendante Hiérarchique (CAH) afin de créer des regroupements d'individus en fonction de leur trajectoire spatio-temporelle ainsi que de leurs variables complémentaires. La CAH nous permet alors de créer des classes d'individus, en maximisant l'homogénéité intra-classe tout en maximisant l'hétérogénéité inter-classe. Les individus semblables seront plus enclins à se regrouper avec d'autres individus proches de leur profil.

Basée sur un critère de ressemblance, la CAH nécessite le calcul de la distance existante entre chaque individu pris deux à deux, aboutissant à la détermination de profils semblables dans le but de regrouper les individus concernés. Agissant sur des données à la fois de nature qualitatives et quantitatives, nous avons fait le choix de mesurer la distance de similarité entre chaque individu à l'aide de l'indice de Gower (Gower, 1971).

L'indice de Gower varie entre 0 (représentant deux individus parfaitement dissemblables) et 1 (représentant deux individus parfaitement identiques). Dans le cas particulier de l'étude de variables qualitatives, l'indice est binaire : il prend la valeur de 0 pour les individus n'ayant pas la même modalité et inversement.

À la suite du calcul de la distance de dissimilarité pratiqué grâce à la méthode de Gower, nous avons effectué la CAH en nous basant sur la méthode de Ward. Cette méthode permet de déterminer la distance entre les groupes au sein de la CAH, tout en créant des groupes d'individus homogènes, les plus différents possibles les uns des autres (Dumolard, 2011).

Désormais, il est nécessaire de se focaliser sur les paramètres liés aux regroupements des individus, disposant à la fois d'une trajectoire spatio-temporelle identique, mais aussi des caractéristiques communes. Ces groupes constituent alors des potentiels de regroupements de covoitureurs bien plus proches du réel que le simple regroupement théorique spatio-temporel.

3. Résultats

À partir des données issues des EMD de l'aire urbaine de Grenoble et de celle de Marseille, nous présentons dans cette partie une représentation synthétique des indicateurs liés au potentiel de regroupement d'individus mobiles. Quelques

2. Le niveau de vie a été reconstruit artificiellement en croisant la Catégorie Socio-Professionnelle (CSP) des individus avec le niveau de vie médian par CSP proposé par l'INSEE (*Niveau de vie selon la catégorie socioprofessionnelle*, s. d. ; INSEE, 2021)

précisions sont à apporter quant aux paramètres retenus pour la calibration de nos modèles.

Pour les CAH, nous avons fait le choix de pondérer les différents critères sélectionnés. En effet, lors du calcul de la distance de dissimilarité de Gower, nous avons respectivement attribué les pondérations suivantes aux différentes variables :

TABLE 1. *Pondération des variables dans le calcul de la distance de Gower*

Variables retenues	Poids dans le calcul de la distance de Gower
Trajectoire spatio-temporelle	0,75
Distance effective parcourue	0,5
Motif du déplacement	0,25
Niveau de vie	0,25
Âge	0,25
Sexe	0,25

Cette pondération tirée de notre revue de littérature (Godillon, 2021 ; Colin, Brangier, 2021 ; Brahmi *et al.*, 2022), présente l'importance de chaque variable dans la "partageabilité" des déplacements quotidiens, notamment pour le transfert modal de l'autosolisme vers le covoiturage.

De plus, afin d'obtenir des groupes d'individus partageant des trajectoires spatio-temporelles et des profils similaires (RSTP), nous nous sommes basés sur les nombres de groupes issus de l'étude du potentiel de regroupement spatio-temporel (RST), présentant un nombre de classes idéal dans le regroupement des individus de notre étude. C'est alors que nous avons augmenté le nombre de classes issues de la CAH, par rapport au nombre de groupes du RST, afin de minimiser l'hétérogénéité intra-classes et l'homogénéité inter-classes. Nous obtenons alors des classes plus nombreuses, mais également plus précises, concernant les caractéristiques individuelles. Le choix du nombre de classes s'est alors construit sur la base du nombre minimal de classes calculé lors de l'étude du potentiel de regroupement spatio-temporel, auquel nous avons ajouté une variabilité de 5%. Cette variabilité de 5% est basée sur la part de covoiturage quotidien tous motifs confondus (de 5%), et donne un aperçu de la part potentielle des autosolistes pouvant se tourner vers le covoiturage (Brahmi *et al.*, 2022).

Certains indicateurs présentés dans cette partie portent sur les émissions de gaz à effet de serre et plus précisément du CO₂. Pour construire un indicateur estimant l'évolution des émissions, nous nous sommes basés sur l'âge moyen du parc automobile français, étant de 12 ans (UTAC, 2021). Nous nous sommes par la suite appuyés sur une étude de l'ADEME (ADEME, 2022), estimant les émissions de CO₂/ km de différents types de motorisations. C'est ainsi que nous avons pu construire les indicateurs souhaités, exposant notamment les gains environnementaux du regroupement de passagers au sein de véhicules particuliers.

Nous avons alors obtenu ce tableau, résumant les indicateurs du potentiel de regroupement spatio-temporel d'individus aux profils similaires .

TABLE 2. *Indicateurs du potentiel de regroupement spatio-temporel multicritères*

Indicateurs	EMD 38, Grenoble			EMD 13, Marseille		
	OT*	RST**	RSTP***	OT	RST	RSTP
Nombre de véhicules concernés	29960	26680	28026	38291	28241	30170
Taux de variation du nombre de véhicules		-10,95%	-6,46%		-26,25%	-21,21%
Taux de remplissage moyen des véhicules	1	1,11	1,07	1	1,24	1,24
Taux de variation du remplissage moyen des véhicules		10,95%	6,84%		23,94%	23,53%
Somme des émissions de CO ₂ des véhicules (kg de CO ₂)	30833	29442	29989	46626	40083	41450
Somme des émissions de CO ₂ des véhicules économisée (kg de CO ₂)		1391	844		6543	5176
Taux de variation de la somme des émissions de CO ₂ des véhicules		-4,51%	-2,74%		-14,03%	-11,07%
Moyenne des émissions de CO ₂ par individu (kg de CO ₂)	1,17	1,12	1,11	1,54	1,39	1,33
Taux de variation de la moyenne des émissions de CO ₂ par individu		-1,90%	-1,31%		-4,55%	-4,47%

* OT : Observations de Terrain

** RST : Regroupement Spatio-Temporel

*** RSTP : Regroupement Spatio-Temporel aux Profils similaires

4. Discussion

4.1. Analyse et interprétation des résultats

Les résultats de notre étude présentent les impacts du regroupement d'individus autosolistes.

Dans le cas des deux terrains d'études que sont les agglomérations de Grenoble (EMD 38) et de Marseille (EMD 13), nous pouvons observer des variations du nombre de véhicules de l'ordre de -6,45% pour le RSTP grenoblois à -26,25% pour le RST marseillais. Or, la diminution du nombre de véhicules a un effet direct sur l'environnement : on note alors une diminution de l'ordre de quelques tonnes de CO₂ concernant les mobilités regroupées. En effet, nous pouvons l'observer grâce à l'indicateur présentant les gains d'émissions de CO₂ : selon les modèles, environ 1 à 6 T de CO₂ pourraient être économisés dans le cas du covoiturage d'individus autrefois autosolistes, au cours de leurs pérégrinations journalières. De plus, la diminution du nombre de véhicules personnels peut réduire considérablement la congestion du trafic routier, améliorant alors la fluidité de la circulation et diminuant les temps de trajets liés aux déplacements quotidiens.

Le regroupement d'individus à bord de véhicules partagés permet également d'augmenter significativement le taux d'occupation de voitures personnelles. Associé au calcul des émissions de CO₂, nous obtenons une baisse de l'ordre de -1,31% pour le RSTP grenoblois à -4,55% pour le RST marseillais.

De plus, nous pouvons observer une réelle inégalité spatiale concernant les impacts liés au regroupement d'autosolistes. Cela pourrait être dû à une différenciation de répartition des individus semblables entre eux, au sein d'un territoire.

En effet, nous pouvons observer que Marseille obtient des taux de variation du nombre de véhicules respectivement 2,4 et 3,3 fois supérieur au RST et au RSTP de Grenoble. De plus, Marseille connaît un taux de variation de la somme des émissions de CO₂ respectivement 3,1 fois et 4 fois supérieur au RST et au RSTP de Grenoble. Nous pouvons émettre l'hypothèse que cette variation serait liée à la densité de population sur les terrains d'études, façonnant le découpage géographique en Zones Fines : pour une même origine et une même destination, les zones fines grenobloises étant plus étendues spatialement que les marseillaises, les distances prises en compte dans le calcul du regroupement seraient alors moins semblables, limitant le regroupement d'autosolistes.

Enfin, il convient de souligner que les résultats que nous avons obtenus concernent le partage de véhicules sur une journée. Il est donc important de noter que l'utilisation d'un tel système de partage entraînerait des économies significatives en termes d'émissions de CO₂, allant de 2400 tonnes (RST) à 1900 tonnes (RSTP) pour Marseille et environ 56 tonnes (RST) à 35 tonnes

(RSTP) pour Grenoble. Ces résultats mettent en évidence que le partage de véhicules est un levier essentiel pour la réduction des émissions de gaz à effet de serre.

4.2. Limites

Notre étude nous a permis de développer des indicateurs illustrant le potentiel de regroupement de mobilité à partir de données spatio-temporelles et ses impacts.

Cependant, quelques limites peuvent être soulevées, comme par exemple la prise en compte des critères permettant de passer de la "partageabilité" difficile à mesurer (Colin, Brangier, 2021). Il serait alors pertinent de proposer une analyse des variables explicatives du partage effectif des transports, dans le but de réaliser des études plus précises et mieux calibrées.

De surcroît, d'autres thèmes seraient à développer, comme la pondération des critères de partageabilité ou encore le nombre de classes retenues après les CAH.

Enfin, pour obtenir une meilleure généralisation des regroupements de pégrinations, il est nécessaire d'assouplir non seulement les fenêtres de temps, mais aussi les fenêtres spatiales, à travers la proximité des zones fines et la compatibilité des profils sociaux pouvant potentiellement covoiturer. Il serait alors particulièrement pertinent de s'intéresser au calcul de la distance de Fréchet optimisée (Devogele *et al.*, 2016) ou encore la méthode de Clustering Trajectory (Lee *et al.*, 2007).

4.3. Apports de notre méthode dans le regroupement d'individus mobiles

Cette étude a permis d'aller au-delà du simple calcul du potentiel de regroupement d'individus basé sur les couples origine/destination ou l'identification de corridors, comme le font les méthodes k-means ou dbscan. Notre objectif consiste en le regroupement des individus susceptibles de covoiturer en utilisant des critères spatiaux, temporels ainsi que des critères liés à leur profil, incluant leur profil social et leurs habitudes de mobilité, comme les motifs de déplacements.

Bien que le regroupement d'individus limite la liberté associée à l'utilisation individuelle d'un véhicule, nous considérons qu'il est crucial d'exploiter toutes les opportunités de covoiturage disponibles. Cela est d'autant plus important dans le contexte actuel où il est primordial de changer nos habitudes de mobilité pour réduire notre impact environnemental.

Il est crucial de rappeler que le covoiturage fait partie intégrante des systèmes

de transport multimodaux proposés sur les territoires, en particulier dans les grandes agglomérations comme Marseille et Grenoble. Cette constatation devrait encourager la promotion du covoiturage en raison de ses multiples avantages par rapport à la voiture personnelle. Il est donc primordial de promouvoir des alternatives de mobilité plus durables pour améliorer la qualité de vie en ville et répondre aux enjeux environnementaux actuels.

5. Conclusion

Cette communication présente une estimation du potentiel de regroupement de mobilité à partir de données spatio-temporelles ainsi que de ses impacts.

Au travers de méthodes statistiques, nous avons observé les variabilités causées par la mise en place de covoiturage au sein de pérégrinations quotidiennes d'individus autosolistes socialement proches. L'aspect environnemental comme l'aspect lié à l'aménagement constituent un levier majeur nécessaire au déclenchement du report modal.

L'estimation du potentiel de regroupement d'individus par le biais d'une Classification Ascendante Hiérarchique permet de déterminer des groupes d'individus particulièrement semblables à la fois dans leurs pratiques de mobilité, mais aussi dans leur profil social.

Notre méthodologie met alors en lumière un potentiel de regroupement supérieur à 20% sur le territoire marseillais et de près de 10% pour le territoire grenoblois. De plus, ces regroupements permettent de limiter les émissions de gaz à effet de serre liées aux pérégrinations d'autosolistes de l'ordre de 3% à plus de 10%.

Le développement de méthodes d'analyse précises d'estimation de potentiel de transfert modal constitue alors un enjeu majeur, dans l'optique de diminuer sensiblement les émissions de gaz à effet de serre liées aux mobilités.

Bibliographie

- ADEME. (2022). Évolution du taux moyen d'émissions de CO2 en France.
- Brahmi L., Fumat V., Trevien C. (2022, juillet). *Se déplacer en voiture : seul, à plusieurs ou en covoiturage ?*. Rapport technique. Commissariat général au développement durable.
- CITEPA. (2022). *Gaz à effet de serre et polluants atmosphériques : Bilan des émissions en France de 1990 à 2021* Rapport national d'inventaire n° 2071sec.
- Colin C., Brangier E. (2021, octobre). Définition et validité de critères de « partageabilité » : premiers résultats. *Proceedings of the 17th "Ergonomie et Informatique Avancée" Conference*, p. 1-9.
- Commenges H. (2014). La mobilité comme variabilité temporelle de la présence spatiale. *Flux*, vol. 95, n° 1, p. 41-55.

- Devogele T., Esnault M., Etienne L. (2016). Distance discrète de fréchet optimisée. *Spatial Analysis and Geomatics (SAGEO)*, p. 15.
- Dumolard P. (2011). *Analyse Multivariée de Données Géographiques*. Éditions Lavoisier, Publications Hermès Sciences.
- Garcin L., Josselin D. (2021, septembre). *L'autocopartage, un nouveau mode de transport innovant et hybride*. Communication session accessibilité, mobilité, transport. Avignon.
- Godillon S. (2021). Le covoiturage quotidien en tension avec les emplois du temps chargés des actifs. Analyse des non-pratiques du covoiturage domicile-travail. In Y. Demoli (Ed.), *Peut-on se passer de la voiture hors des centres urbains ?*, p. 157 – 172. Gif-sur-Yvette, Paris-Saclay Éditions.
- Gower J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, vol. 27, n° 4, p. 857–871. Consulté sur <https://www.jstor.org/stable/2528823> (Publisher: [Wiley, International Biometric Society])
- Guidez J.-M. (2002). *Dix ans de mobilité urbaine en France : les années 1990*. Rapport technique. Lyon, Certu.
- Géoconfluences. (2020, juin). *Transfert modal, report modal*.
- INSEE. (2021). *En 2019, le niveau de vie médian augmente nettement et le taux de pauvreté diminue - Insee Première - 1875*. Rapport technique n° 1875.
- INSEE. (2021). *La voiture reste majoritaire pour les déplacements domicile-travail, même pour de courtes distances*.
- Kaufmann V. (2003). Pratiques modales des déplacements de personnes en milieu urbain : des rationalités d'usage à la cohérence de l'action publique. *Revue d'Economie Regionale Urbaine*, vol. février, n° 1, p. 39–58. (Publisher: Armand Colin)
- Lee J.-G., Han J., Whang K.-Y. (2007). Trajectory clustering: A partition-and-group framework. , p. 12.
- Les partenaires de l'EMD. (2010). *Enquête globale de déplacements des Bouches-du-Rhône : rapport de présentation des principaux résultats sur l'ensemble du territoire*. Rapport technique.
- Ministère de la transition écologique. (2021, mars). *Le vélo, un mode de déplacement vertueux et avantageux*.
- Niveau de vie selon la catégorie socioprofessionnelle*. Rapport technique. (s. d.).
- Salomon I., Bovy P., Orfeuil J.-P. (Eds.). (1993). *A Billion Trips a Day: Tradition and Transition in European Travel Patterns*. Springer Netherlands.
- SMTC. (2010). Enquête ménages-déplacements 2010 : Grande Région Grenobloise. , p. 56.
- UTAC. (2021). Contrôle technique périodique des véhicules légers - rapport d'activité annuel.
- Wiel M. (2002). *Ville et automobile*. Descartes et Cie.

Cartographeur pour le web avec la bibliothèque JavaScript `bertin`

Nicolas Lambert¹, Timothée Giraud¹, Matthieu Viry¹, Ronan Ysebaert²

1. UAR RIATE, CNRS

Université Paris Cité. Place Paul Ricœur. 75013 Paris

nicolas.lambert@cnrs.fr, timothee.giraud@cnrs.fr, matthieu.viry@cnrs.fr

2. UAR RIATE, Université Paris Cité

Université Paris Cité. Place Paul Ricœur. 75013 Paris

ronan.ysebaert@cnrs.fr

RÉSUMÉ. *bertin* est une bibliothèque JavaScript pour la réalisation de cartes thématiques pour le Web. La bibliothèque est bâtie principalement à partir de la bibliothèque JavaScript D3.js. Sa conception vise à permettre aux utilisateurs de créer rapidement des cartes thématiques interactives sans forcément connaître le langage JavaScript ni le framework de D3.js. Un "wrapper" de cette bibliothèque est également disponible dans les langages R et Python.

ABSTRACT. *bertin* is a JavaScript library for designing thematic maps for the web. This library is mainly based on the D3.js JavaScript library. It aims at providing to users an easy solution to design interactive thematic maps without necessarily knowing the JavaScript language or the D3.js framework. A wrapper of this library is also available in R and Python.

MOTS-CLÉS : JAVASCRIPT, R, CARTOGRAPHIE, OBSERVABLE, VISUALISATION DE DONNÉES, GEOWEB, NOTEBOOKS

KEYWORDS: JAVASCRIPT, R, CARTOGRAPHY, OBSERVABLE, DATAVIZ, GEOWEB, NOTEBOOKS

1. Introduction

Des pionniers de la dataviz (Bertin, 1967 ; Tukey, 1977 ; Tufte *et al.*, 1983) à la révolution du Web, couplée à l'essor des démarches quantitatives (Tobler, 1970 notamment), la visualisation de données s'est peu à peu structurée en tant que pratique interdisciplinaire (Grandjean, 2022) qu'on regroupe généralement sous le terme de *visual studies*. Ces travaux anciens sont d'ailleurs très largement mobilisés aujourd'hui et nombre d'acteurs n'hésitent pas à remettre au goût du jour des instigateurs célèbres de la visualisation de données tels que Charles-Joseph Minard (1781-1870), Emile Cheysson (1836-1910) ou W.E.B Du Bois (1868-1963) à l'aide de nouvelles technologies informatiques.

L'essor récent de la visualisation de données sur le Web a notamment été porté par les travaux de Mike Bostock en JavaScript et le développement de la bibliothèque D3.js (Bostock *et al.*, 2011). Cependant, contrairement à d'autres projets dans d'autres langages, le niveau d'abstraction relativement bas de cette bibliothèque ne vise pas à permettre de réaliser des cartes et des graphiques très rapidement mais plutôt de fournir un cadre souple et ouvert pour la visualisation de données. Or cette capacité à produire des visualisations en quelques lignes de codes est très important pour articuler des analyses de données et des visualisations dans le cadre de *notebooks*. En proposant une syntaxe concise et des comportements par défaut, la bibliothèque *bertin*, présentée ici, entend répondre à cet enjeu.

Cette contribution rappelle tout d'abord les enjeux qui ont guidé à la création bibliothèque *bertin* (partie 2). Elle décrit ensuite ses fondements techniques et ses fonctionnalités essentielles (partie 3). Les passerelles créées entre la bibliothèque *bertin* et d'autres langages de programmation (R et Python), ainsi que l'univers des *notebooks* (Quarto) sont enfin présentées (partie 4). Ces implémentations rendent ainsi possible l'usage de cette bibliothèque dans des environnement socio-techniques diversifiés.

2. Contexte et enjeux

2.1. Les notebooks

Les *notebooks* sont un exemple de programmation lettrée fournissant une manière de combiner langage naturel et langage informatique au sein d'un même document. Les *notebooks* sont fréquemment utilisés en sciences des données afin d'explorer ou d'analyser leur contenu et permettent de faciliter le partage mais également la reproductibilité des résultats et sorties graphiques qu'ils décrivent. Leur aspect interactif leur permet généralement de régénérer "à la volée" les différents calculs et les différentes sorties graphiques qu'ils contiennent, offrant ainsi un environnement particulièrement adapté à l'exploration de données. Dans une démarche de reproductibilité, ils sont aussi un moyen de documenter un travail scientifique (Shen, 2014).

Ce type de programmation n'est pas totalement nouveau, il a notamment été formalisé par Donald Knuth dès 1984 et des fonctionnalités *notebook* se développent dès la fin des années 80 dans plusieurs systèmes de calcul formel (*MathCAD* en 1987, *Mathematica* en 1988, *Maple* en 1989). Il connaît un essor important au cours des années 2000 puis 2010 dans le domaine de la science des données, en particulier dans les solutions open sources. Les fonctionnalités *notebook* sont ainsi introduites dans *IPython* (Pérez *et al.*, 2007) v0.12 en décembre 2011 et la première version de *knitr* (Xie, 2014) est publiée en 2012.

Le *notebook* se situe ainsi au centre d'un système socio-technique, comme un objet porteur de différents objectifs et de différentes fonctions : tout d'abord, et pour prendre le cas de *bertin*, par l'intégration de la réalisation cartographique au sein d'un *workflow* reproductible. Il se matérialise en amont par l'import des données et leur mise en forme, puis par l'explicitation des intentions de l'auteur et du code qui en résulte et conduisent à une représentation graphique originale.

2.2. Pourquoi JavaScript ?

De nombreux langages paraissent naturels pour l'analyse et la visualisation de données. C'est notamment le cas de Python et R, aussi largement utilisés à ces fins. Mais au regard de ces deux langages, le JavaScript n'est pas dénué d'atouts (Bostock, 2021). Tout d'abord, JavaScript est le langage du Web. Il tourne dans tous les navigateurs Web et ne nécessite aucune installation. Et si ce langage n'a pas été conçu au départ pour cela, de nombreuses bibliothèques ont été développées pour manipuler les données (*Arquero*, *tidy.js*, etc.), démontrant ainsi que ce langage peut être étendu pour mieux satisfaire les besoins inhérents aux *visual studies*. Par ailleurs, avec des bibliothèques comme *D3.js*¹, qui permet de manipuler *canvas* et *SVG*, JavaScript est probablement déjà aujourd'hui le langage le plus adapté pour la visualisation de données 2D. Le choix du Web est également particulièrement intéressant pour des raisons d'interopérabilité, permettant au développeur de concevoir des applications qui ne dépendent d'aucun système d'exploitation mais nécessitent uniquement un navigateur Web. Enfin, avec *WebGPU*² et *WebAssembly*³, l'avenir de JavaScript semble prometteur.

Nous nous inscrivons donc ici dans une transition déjà identifiée par des auteurs du champ tels que Roth (2015): "*a broad transition in client-side web mapping away from standalone, proprietary technologies (e.g., Adobe Flash) and towards open technologies that leverage the HTML, CSS, SVG, and XML web standards (the Open Web Platform) and the JavaScript programming language*". Cette tendance, qui n'est pas spécifique à la visualisation de données, s'est largement accélérée depuis les années 2010 en raison des améliorations de performances apportées par les navigateurs Web et leurs moteurs JavaScript (en effet la compilation *Just In Time* a été introduite dès 2008 dans les moteurs JavaScript *V8* et *SpiderMonkey*, témoignant des efforts mis en œuvre pour améliorer leurs performances). Cette tendance s'inscrit

¹ <https://github.com/d3/d3>

² <https://www.w3.org/TR/webgpu>

³ <https://www.w3.org/wasm/>

dans un processus amorcé dès les débuts du Web et déjà identifié comme pouvant améliorer l'acceptation des utilisateurs tout en évitant les écueils classiques de la distribution de logiciels (Rice et al., 1996).

2.3. Travaux connexes

Il existe de nombreuses bibliothèques JavaScript permettant de visualiser des données géographiques (figure 1). On peut néanmoins différencier les bibliothèques plutôt dédiées à la visualisation interactive des données géographiques, comme *MapLibre*, *Leaflet* ou encore *OpenLayers* qui contiennent plusieurs fonctionnalités permettant de zoomer, afficher/masquer des couches, créer des animations ; aux bibliothèques davantage orientées vers la visualisation des données statistiques suivant l'usage d'une grammaire graphique comme *Observable Plot*⁴ ou *Vega-Lite* (Satyanarayan, 2016).

Fonctionnalités	OpenLayers	MapLibre GL	Leaflet	Deck.gl	Observable Plot	Vega-Lite	bertin
Première <i>release</i>	2006	2020	2011	2015	2018	2013	2021
Famille	Carto interactive	Carto interactive	Carto interactive	Exploration de données	Visu de données	Graphique interactifs « ggplot »	Carto thématique pour le Web
Interactivité	Zoom	Oui	Oui	Oui	Oui	Non	Non
	Tooltips	Oui	Oui	Oui	Oui	Non	Oui
	Animations	Oui	Oui	Plugin	Oui	Non	Non
	Élément HTML	Canvas (WebGL optionnellement)	Canvas (WebGL)	Canvas	Canvas (WebGL)	SVG	SVG
Symbologie ⁵	+	++ (3D)	+	++ (3D)	++	+	+++
Habillage	Labels	Oui	Oui	Plugin	Oui	Oui	Oui
	Projections	++	Mercator	Mercator	Mercator	+++	++
	Minimap	Oui	Oui	Plugin	Oui	Non	Non
	Tuiles raster	Oui	Oui	Oui	Oui	Non	Non
	Graticules	Oui	Plugin	Plugin	Non	Non	Non
	Légende	Plugin	Plugin	Oui	Non	Oui	Oui

FIGURE 1. tableau comparatif des bibliothèques JavaScript de cartographie

La bibliothèque *bertin* est davantage à rapprocher de ces deux dernières, qui partagent un peu la même philosophie avec des références claires à l'écosystème

⁴ <https://github.com/observablehq/plot>

⁵ Variété des options de symbologie proposées

D3.js. *bertin* se démarque néanmoins en étant spécifiquement dédié à la cartographie thématique. Bien plus que les autres bibliothèques précédemment citées, l'effort est apporté à la multiplicité des fonctions de représentations proposées, suivant les préceptes de sémiologie graphique initiés par Jacques Bertin et la mise en scène de la carte (habillage de la carte, gestion des projections, paramétrage avancé de la légende).

3. La bibliothèque *bertin*

bertin est une bibliothèque JavaScript consacrée à la réalisation de cartes statistiques vectorielles dont le développement a débuté en novembre 2021. La bibliothèque a été nommée ainsi en hommage au géographe français Jacques Bertin (1918 - 2010) et ses travaux fondateurs sur la sémiologie graphique (Bertin, 1967 ; Bertin, 1973) dont l'influence est aujourd'hui encore majeure dans le domaine de la visualisation de données. Néanmoins, la bibliothèque ne propose en aucun cas un décalque de la sémiologie graphique de Jacques Bertin, mais plutôt des méthodes de représentations classiques telles que présentées dans les différents manuels de cartographie (Lambert et Zanin, 2020).

3.1. Description technique

La bibliothèque repose sur de nombreuses dépendances open source, principalement liées à l'écosystème D3.js mais également sur des bibliothèques spatiales telles que *jsts*⁶, *turf*⁷ et *proj4js*⁸. Elle est publiée sous licence MIT et est en développement actif. 164 versions ont été publiées entre novembre 2021 et mai 2023 (version 1.7.3). La bibliothèque *bertin* est maintenue par une personne, mais compte à ce jour 7 contributeurs externes. Le volume du code représente 336 Ko dans la dernière version. Le code source et la documentation sont accessibles sur GitHub : <https://github.com/neocarto/bertin>

3.2. Principes

Le principe de la bibliothèque *bertin* est de proposer un outil permettant de réaliser rapidement des cartes thématiques variées sans faire appel à la programmation en JavaScript ni directement à la bibliothèque D3.js.

⁶ <https://github.com/bjornharrtell/jsts>

⁷ <https://github.com/Turfjs/turf>

⁸ <https://github.com/proj4js/proj4js>



FIGURE 2. niveau d'abstraction de la bibliothèque bertin

Le but est d'avoir un outil de cartographie dans l'écosystème JavaScript directement appréhendable par des étudiants en géographie sans compétences préalables en développement Web, mobilisable dans des *notebooks*. Le niveau d'abstraction de la bibliothèque (figure 2) est donc volontairement très élevé, et mobilise une grammaire qui fait directement référence au vocabulaire métier de la cartographie thématique.

La principale fonction de la bibliothèque est *draw()* qui permet de dessiner des cartes thématiques (figure 3). Le fonction prend comme seul argument un objet JavaScript contenant l'ensemble des informations nécessaires au dessin de la carte. L'objet se divise en 2 parties : des paramètres généraux (propriété "params" de l'objet donné en argument, optionnel) définissant l'emprise de la carte, sa projection, sa couleur de fond, ses marges et sa taille ; et un *array* (propriété "layers" de l'objet donné en argument) contenant l'ensemble des couches à afficher (Annexe 1). Chaque couche est définie par un type spécifique (bubble, spikes, header, minimap...) qui détermine le type de rendu. L'ordre d'écriture des couches détermine l'ordre d'affichage.

```

bertin.draw( ← draw() est la fonction principale du package
  { ← Tous les paramètres de la cartes sont définis par un objet json {...}
    params: {...}, ← Paramètres généraux (projection, couleur de fond, taille, marges)
    layers: [ ← Les couches sont définies par un tableau [...].
      {type: "bubble", ...}, ←
      {type: "simple", ...}, ←
      {type: "outline", ...}, ←
    ]
  }
)

```

Toutes les couches sont définies par des objets {}

↑
Chaque couche est définie par un type qui détermine son comportement

FIGURE 3. la fonction draw()

Un type *custom* permet également d'intégrer à la carte tout type de contenu via des fonctions spécifiques construites par l'utilisateur, permettant ainsi pleinement l'interopérabilité de la bibliothèque avec D3.js notamment. La fonction renvoie un élément HTML de type SVG et éventuellement un générateur qui produit une nouvelle valeur chaque fois que l'utilisateur interagit avec la carte. Plusieurs paramètres par défaut sont définis de façon empirique pour déterminer un comportement par défaut optimal et ainsi alléger la syntaxe minimale nécessaire à la génération d'une carte.

Au final, la bibliothèque *bertin* permet de construire des visualisations cartographiques finalisées directement publiables sur des sites Web ou téléchargeables au format SVG. Ci-dessous, un exemple de carte réalisée avec *bertin* pour le journal l'Humanité.fr⁹ (figure 4).

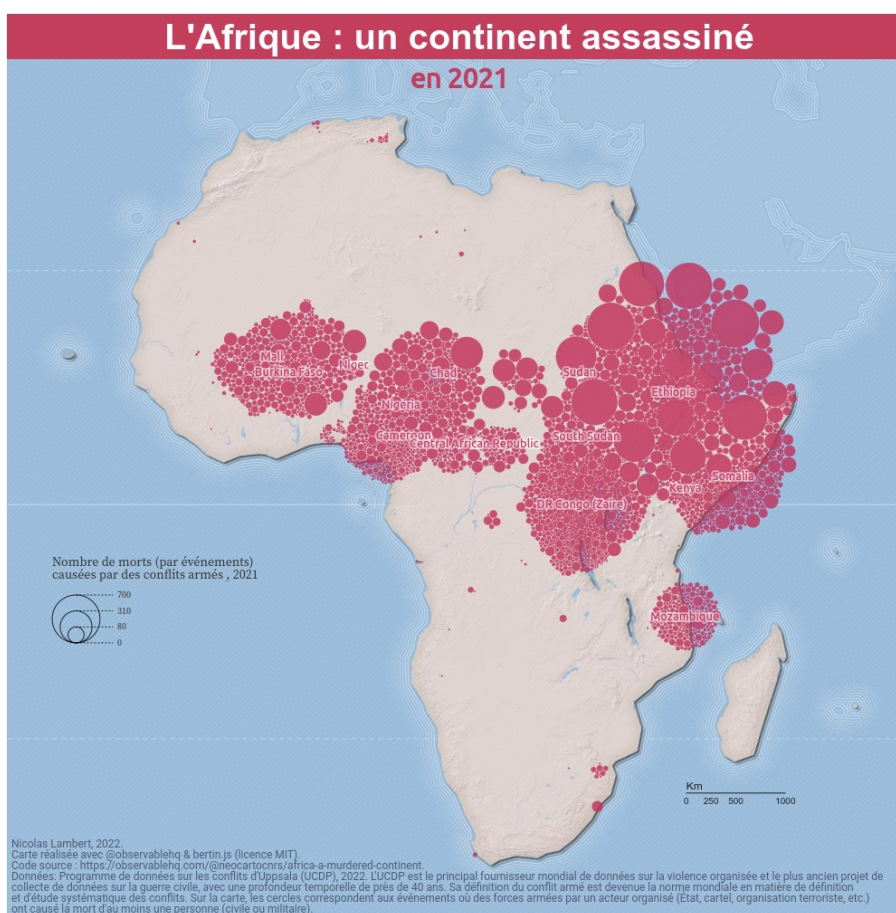


FIGURE 4. carte « finalisée » réalisée avec *bertin*

⁹ <https://www.humanite.fr/en-debat/regard-de-cartographe/l-afrique-un-continent-assassine-775086>

3.3. Types de cartes

Les types de couches définies dans la bibliothèque *bertin* peuvent être séparés en 2 catégories : les couches de symbologie d'une part et les couches d'habillage d'autre part (figure 5). Les couches de symbologie prennent systématiquement en entrée un fichier GeoJSON contenant les données statistiques et les géométries. Les autres paramètres permettent de définir la personnalisation de la couche à afficher (couleur, taille du symbole, épaisseur des traits, etc.). Les couches d'habillage ont principalement un rôle esthétique : ombrage, graticule, hachures, tuiles raster... Chaque fonction proposée par la bibliothèque dispose de paramètres cohérents par défaut (discrétisation, palettes de couleur, etc.), rendant possible la réalisation de cartes thématiques avec un code condensé (Annexe 2), mais aussi hautement paramétrable.

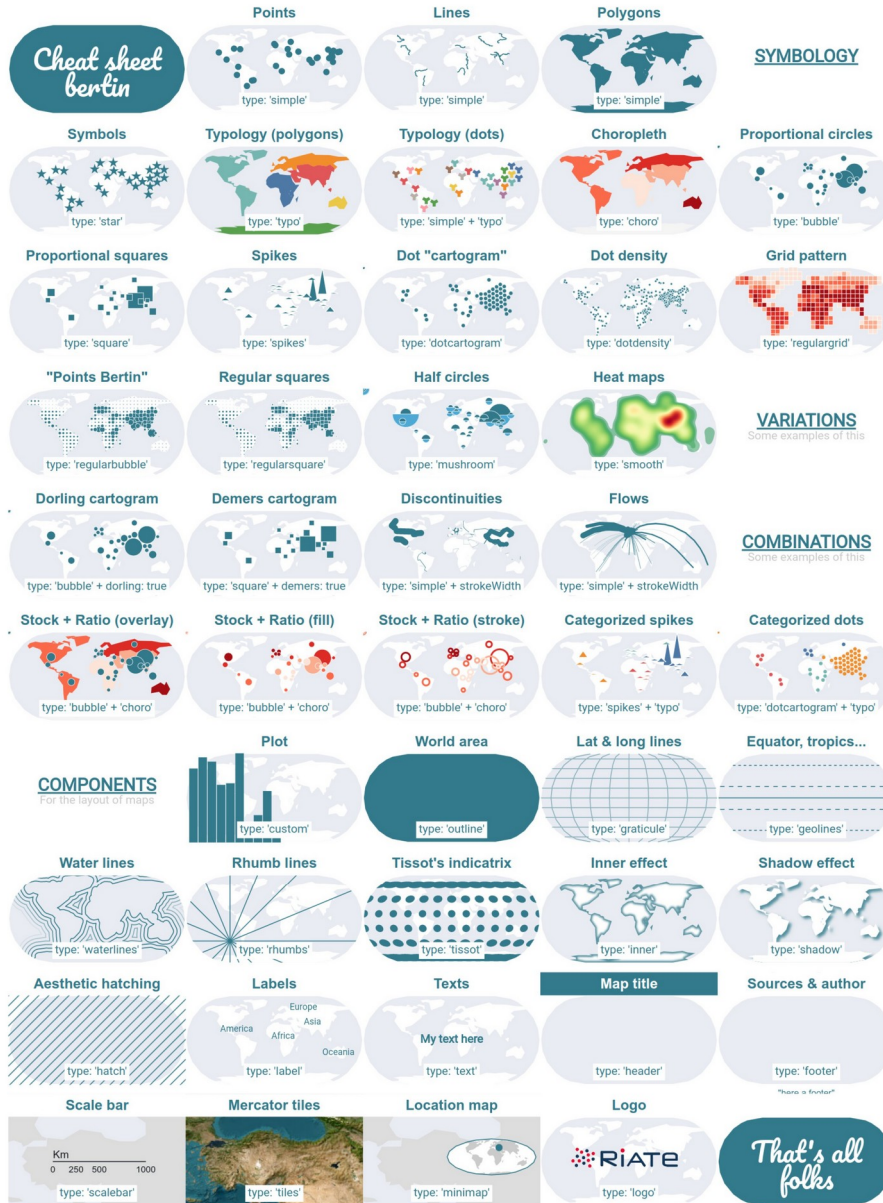


FIGURE 5. les types de couches

3.4. Une documentation abondante pour une diffusion large

En sus de la documentation sur GitHub et npm, *bertin* est agrémenté d'une collection Observable¹⁰ constituée d'une trentaine de *notebooks* qui permettent d'introduire la bibliothèque par une série de cas d'usage. Ils présentent la philosophie et les principales fonctionnalités de la bibliothèque, en jouant de façon didactique sur les différents arguments des fonctions proposées. Ces *notebooks* rappellent à la fois la façon d'initier une représentation (import des données, jointures attributaires), de jouer sur la symbologie, en combinant parfois plusieurs propositions ("stock + typo") et de mettre en œuvre des éléments avancés de mise en page (affichage de tuiles raster, *waterlines*, *minimap*, *tooltips*, *labels*).

Au final, cet important travail de documentation vise à favoriser l'appropriation de *bertin* par un public pas nécessairement familier du langage JavaScript et de la bibliothèque D3.js ; et rendre ainsi possible son usage par une multiplicité d'utilisateurs dans des contextes de création très variés.

4. Contexte d'utilisation

4.1. Notebooks (*Observable* et *Quarto*)

Comme n'importe quelle bibliothèque JavaScript *front-end*, *bertin* peut être directement importé et utilisé dans une page Web pour y insérer une carte interactive. Mais son utilisation fait particulièrement sens dans le cadre de *notebooks Observable* (Perkel, 2021), la bibliothèque étant pensée pour s'intégrer parfaitement avec cet écosystème, et notamment son aspect réactif. Les cartes dessinées avec *bertin* peuvent fonctionner comme des générateurs. Avec l'ajout du paramètre *viewof*, chaque couche de la carte peut servir d'*input* pour communiquer avec d'autres éléments du *notebook*. Cette fonctionnalité est particulièrement adaptée à la réalisation de tableaux de bord (ou *dashboards*) et propose au lecteur de la carte interactive résultante d'être acteur de son exploration et son interprétation.

Nous retrouvons la même logique dans Quarto (Allaire, J. *et al.*, 2022), avec la possibilité supplémentaire d'articuler des traitements et des analyses en R ou en Python avec des visualisations en JavaScript. Par ailleurs, l'aspect réactif d'Observable JavaScript et le système de *panels* proposé par Quarto, permet également de réaliser rapidement et avec très peu de code, des *single page applications* basées sur *bertin*.

4.2. R

Nous travaillons également à rendre la bibliothèque *bertin* disponible dans le logiciel libre R (R Core Team, 2022). Le logiciel R est un logiciel et un langage de programmation particulièrement adapté à la création de chaînes de traitements reproductibles. Son écosystème spatial est à la fois riche et robuste (Bivand, 2021) et

¹⁰ <https://observablehq.com/collection/@neocartocnrs/bertin>

il est déjà assez facile d'y manipuler, traiter et représenter des données géographiques. Le système d'extensions du logiciel R (les *packages*) permet d'enrichir les fonctionnalités du logiciel. Ces extensions peuvent notamment être des *wrappers* autour de bibliothèques écrites dans d'autres langages, par exemple en C, C++ ou JavaScript. Le *package sf* (Pebesma, 2018) est le *package* de référence pour réaliser les opérations d'import, de manipulation et de géotraitement. Ce *package* sert de fondation à plusieurs *packages* de cartographie thématique tels que *tmap* (Tennekes, 2018) et *mapsf* (Giraud, 2022) mais également à d'autres *packages* (*leaflet*, *mapview*, *mapdeck*) qui sont eux dédiés à la création de cartes interactives (*slippy maps*) et s'appuient sur des bibliothèques JavaScript pour fonctionner.

Nous avons donc créé le *package bertin* (Giraud et Lambert, 2023) qui sert d'interface entre R et la bibliothèque JavaScript. Ce faisant nous rendons disponible directement dans R les représentations offertes par *bertin* (JavaScript). Ce portage en R permet de créer les cartes de qualité en SVG ou en HTML directement dans une chaîne de traitement reproductible (Annexe 3). D'un point de vue plus technique, le *package bertin* s'appuie sur le *package htmlwidgets* (Vaidyanathan *et al.*, 2023) dont le but est de créer des *widgets* en HTML et sur *sf* pour s'intégrer facilement dans l'écosystème des *packages* pré-existants. Nous avons choisi de proposer chaque type de représentation dans des fonctions distinctes, cette méthode permet d'exposer la documentation du *package* de manière plus détaillée, plus intégrée et plus idiomatique. Par ailleurs ce *package* nous permet d'intégrer facilement *bertin* dans des applications *shiny* (Chang *et al.*, 2022) et de proposer des applications interactives de cartographie thématique sans maîtriser le JavaScript. Des tests sont actuellement en cours pour valider l'utilisation de *bertin* avec *shiny*.

4.3. Python / Jupyter

Enfin, des utilisateurs de *bertin* contribuent également à rendre cette bibliothèque disponible dans un environnement Python, et plus particulièrement dans l'environnement interactif Jupyter (Kluyver *et al.*, 2016) qui permet notamment de réaliser des *notebooks* en Python. Ainsi l'extension *ipybertin*¹¹ permet d'appeler facilement la bibliothèque *bertin* depuis des *notebooks* interactifs écrits en Python. *ipybertin* utilise le mécanisme d'extension proposé par Jupyter Notebook et le widget *DOMWidget* proposé par *ipywidget*, la bibliothèque de références proposant des *widgets* HTML interactifs pour les *notebooks* Jupyter.

Comme dans le cas du *package R bertin*, *ipybertin* permet de créer des cartes thématiques mobilisant toutes les fonctionnalités de *bertin* et ce sans saisir le moindre code JavaScript (Annexe 4). Toutefois, contrairement au *package R bertin*, l'API de *ipybertin* suit l'API proposée par *bertin* en exposant un simple objet *Map* encapsulant la fonction *draw* et acceptant un dictionnaire Python (un type de données similaire aux objets JavaScript attendus par *bertin*) contenant les propriétés *params* et *layers*.

¹¹ <https://github.com/davidbrochart/ipybertin>

5. Conclusion

Alors que la philosophie de D3.js est de mettre à disposition un écosystème complet permettant de développer des visualisations de données sur mesure pour le Web, la bibliothèque *bertin* permet de produire rapidement de nombreuses cartes thématiques interactives dans une syntaxe concise. Sans chercher à remplacer d'autres projets JavaScript similaires comme *Observable Plot* ou *Vega-Lite*, la bibliothèque *bertin* vise à contribuer, en complémentarité, à la constitution d'un écosystème spatial en JavaScript particulièrement pertinent pour les *visual studies* via des *notebooks*. Cet écosystème semble se constituer autour de bibliothèques de plutôt haut niveau permettant d'effectuer des tâches plus ou moins complexes en une ou quelques lignes de code. D'autres travaux vont dans ce sens (*geotoolbox*, *statsbreaks*, *go-cart-wasm...*).

Bibliographie non numérotée et références

- Allaire J., Teague C., Scheidegger C., Xie Y., Dervieux C., (2022). Quarto (Version 1.2) [Computer software]. <https://doi.org/10.5281/zenodo.5960048>
- Bertin J. (1967), Sémiologie graphique. Les diagrammes. Les réseaux. Les cartes, Paris/La Haye, Mouton ; Paris, Gauthier-Villars.
- Bertin J. (1973), Sémiologie graphique. Les diagrammes, les réseaux, les cartes, Paris/La Haye, Mouton ; Paris, Gauthier-Villars.
- Bivand R.S. (2021). Progress in the R ecosystem for representing and handling spatial data. *Journal of Geographical Systems*, 23(4), 515-546. <https://doi.org/10.1007/s10109-020-00336-0>
- Bostock M. (2021). Script for Data Analysis, <https://towardsdatascience.com/javascript-for-data-analysis-2e8e7dbf63a7>
- Bostock M., Ogievetsky V., Heer J. (2011). D³ Data-Driven Documents, IEEE, Volume: 17, <https://ieeexplore.ieee.org/abstract/document/6064996>
- Chang W., Cheng J., Allaire J., Sievert C., Schloerke B., Xie Y., Allen J., McPherson J., Dipert A., Borges B., (2022). *_shiny*: Web Application Framework for R_. R package version 1.7.4, <<https://CRAN.R-project.org/package=shiny>>.
- Giraud T. (2022). *mapsf*: Thematic Cartography. <https://CRAN.R-project.org/package=mapsf>
- Giraud T., Lambert N., (2023). *bertin*: Thematic Cartography with 'bertin.js'_. <https://github.com/riatelab/bertin/>, <https://riatelab.github.io/bertin/>.
- Grandjean M. (2022). La visualisation de données, entre usages démonstratifs et heuristiques.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90).
- Lambert N., Zanin C., (2020). *Practical handbook of thematic cartography: principles, methods, and applications*. CRC Press

- Pebesma E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10 (1), 439-446, <https://doi.org/10.32614/RJ-2018-009>
- Pérez F., Granger B.E., (2007). IPython: A System for Interactive Scientific Computing, *Computing in Science and Engineering*, vol. 9, no. 3, pp. 21-29, doi:10.1109/MCSE.2007.53. URL: <https://ipython.org>
- Perkel J.M. (2021). Reactive, reproducible, collaborative: computational notebooks evolve, <https://www.nature.com/articles/d41586-021-01174-w>
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rice J., Farquhar A., Piernot P., Gruber T., (1996). "Using the Web instead of a window system". In: Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96. the SIGCHI conference. Vancouver, British Columbia, Canada: ACM Press, pp. 103-110. isbn: 978-0-89791-777-3. doi: 10.1145/238386.238442.
- Roth R.E., Donohue R.G., Sack C.M., Wallace T.R., Buckingham T.M.A., (2015). "A Process for Keeping Pace with Evolving Web Mapping Technologies". In: *Cartographic Perspectives* 78, pp. 25-52. issn: 1048-9053. doi: 10.14714/CP78.1273. url: <https://cartographicperspectives.org/index.php/journal/article/view/cp78-roth-et-al>
- Satyanarayan A., Moritz D., Wongsuphasawat K., Heer J. (2016). Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, 23(1), 341-350. <https://idl.cs.washington.edu/files/2017-VegaLite-InfoVis.pdf>
<https://vega.github.io/vega/about/vega-and-d3/>
- Shen H. (2014). Interactive notebooks: Sharing the code. *Nature*, 515(7525), 151-152.
- Tennekes M. (2018). "tmap: Thematic Maps in R." *Journal of Statistical Software*, *84*(6), 1-39. doi:10.18637/jss.v084.i06 <<https://doi.org/10.18637/jss.v084.i06>>.
- Tobler W.R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1), 234-240.
- Tufte E., Graves-Morris P., (1983). The visual display of quantitative information. *Diagrammatik-Reader. Grundlegende Texte aus Theorie und Geschichte*. Berlin: De Gruyter, 219-230.
- Tukey J.W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160).
- Vaidyanathan R., Xie Y., Allaire J., Cheng J., Sievert C., Russell K., (2023). `_htmlwidgets: HTML Widgets for R`. R package version 1.6.1, <<https://CRAN.R-project.org/package=htmlwidgets>>.
- Xie Y. (2014). "knitr: A Comprehensive Tool for Reproducible Research in R." In Stodden V, Leisch F, Peng RD (eds.), *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. ISBN 978-1466561595, <http://www.crcpress.com/product/isbn/9781466561595>.

Annexes

Annexe 1. Exemple de syntaxe utilisée par la bibliothèque bertin

```
bertin.draw({
  params: { width: 500, projection: "Bertin1953" },
  layers: [
    { geojson: world, fill: "#ffc94d" },
    { type: "graticule" },
    { type: "outline" },
    { type: "header", text: "Carte du Monde" }
  ]
})
```

Annexe 2. Concision du code grâce aux paramètres par défaut

Carte SVG



PIB par habitant en 2020

Code long

```
bertin.draw({
  layers: [
    {
      geojson: world,
      display: true,
      fill: {
        type: "choro",
        values: "gdpcc",
        colors: "#5555",
        nbreaks: 5,
        method: "quantile",
        col_missing: "#f5f5f5"
      },
      fillOpacity: 1,
      stroke: "white",
      strokeOpacity: 1,
      strokeWidth: 0.5,
      strokeLinecap: "round",
      strokeLinejoin: "round",
      strokeDasharray: "none",
      tooltip: false
    },
    {
      type: "graticule",
      display: true,
      stroke: "white",
      strokeWidth: 0.5,
      strokeOpacity: 0.5,
      strokeDasharray: 2,
      strokeLinejoin: "round",
      strokeLinecap: "round",
      step: [10, 10]
    },
    {
      type: "outline",
      display: true,
      fill: "#d9d9d9",
      stroke: "none",
      fillOpacity: 1,
      strokeWidth: 1
    },
    {
      type: "header",
      display: true,
      text: "PIB par habitant en 2020",
      fontSize: 25,
      fill: "#999999",
      background: "white",
      backgroundOpacity: 1,
      anchor: "middle"
    }
  ]
})
```

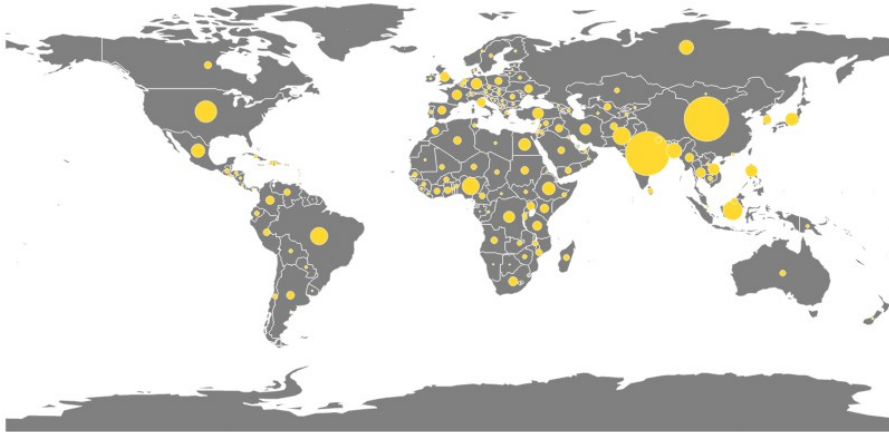
Code condensé grâce aux valeurs par défaut

```
bertin.draw({
  layers: [
    {
      geojson: world,
      fill: { type: "choro", values: "gdpcc" }
    },
    { type: "graticule" },
    { type: "outline" },
    { type: "header", text: "PIB par habitant en 2020" }
  ]
})
```

Annexe 3. Exemple de syntaxe dans R

Exemple

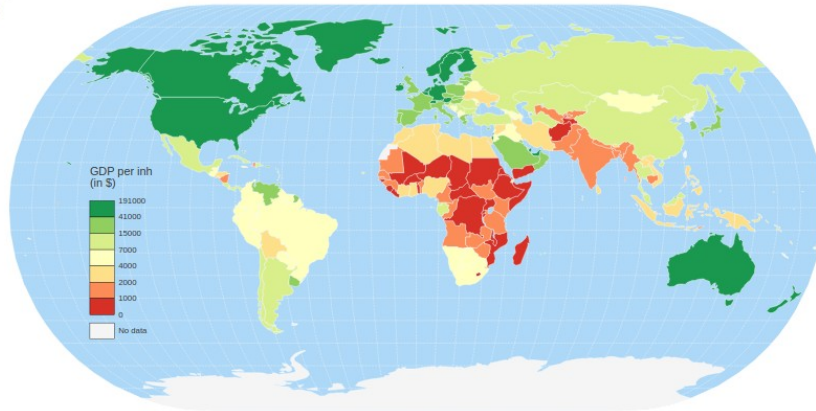
```
library(bertin)
library(sf)
world <- st_read(system.file("gpkg/world.gpkg", package = "bertin"),
                 layer = "world", quiet = TRUE)
bt_param(width = 800)|>
bt_layer(data = world, fill = "#808080") |>
bt_bubble(data = world, values = "pop", k = 20) |>
bt_draw() |>
bt_save("map.svg")
```



Annexe 4. Exemple de syntaxe dans l'environnement Python

```
In [9]: options = {
  "params": {"projection": "d3-geo-projection.geoEckert3()"},
  "layers": [
    {
      "type": "layer",
      "geojson": f'bertin.merge({world}, "ISO3", {data}, "id")',
      "fill": {
        "type": "choro",
        "values": "gdppc",
        "nbreaks": 7,
        "method": "quantile",
        "colors": "RdYlGn",
        "leg_round": -3,
        "leg_title": "GDP per inh\n(in $)",
        "leg_x": 100,
        "leg_y": 200,
      },
      "tooltip": ["$name", "sgdppc", "(current US$)"],
    },
    {"type": "graticule"},
    {"type": "outline"},
  ]
}
Map(options=options)
```

Out[9]:



La partie immergée de l'information géographique : Analyse critique à travers le cas de la planification spatiale maritime

Juliette DAVRET¹, Brice TROUILLET²

1. Nantes Université, CNRS, UMR LETG, 44000 Nantes, France
Juliette.Davret@etu.univ-nantes.fr

2. Nantes Université, CNRS, UMR LETG, 44000 Nantes, France
Brice.Trouillet@univ-nantes.fr

RESUME. *Ce travail consiste à étudier les apports potentiels d'une analyse critique de l'ensemble du cycle de vie de l'information géographique décomposé en quatre étapes : la fabrication des données, les traitements des informations géographiques, leur représentation et enfin leur circulation à travers le cas d'étude de la Planification Spatiale Maritime (PSM). La PSM vise à organiser le développement de l'économie maritime et à atteindre des objectifs sociaux tout en respectant le bon état écologique des mers à partir « meilleures données disponibles ». Cela se base sur des informations géographiques et des géotechnologies. Cependant, alors qu'une partie de la « réalité » n'est ni décrite par les informations existantes, ni mobilisée dans le cadre de la planification et qu'il demeure des limites techniques intrinsèques aux technologies utilisées, le rôle des informations géographiques et des géotechnologies reste peu étudié. Cela aboutit à une représentation partielle, voire partielle, des enjeux que la cartographie critique ne cesse de questionner. Ces recherches contribuent à expliciter les enjeux informationnels de la planification, à l'ère du numérique, et à alimenter le tournant critique à l'œuvre dans ce domaine par une démarche exploratoire et qualitative. Ce travail s'intéresse aux choix faits lors des processus de planification et les influences qu'ils exercent sur les décisions politiques. En abordant le sujet par le volet sociotechnique des processus de planification, les relations de domination entre les différents acteurs et actrices sont mis en lumière. Ces travaux démontrent que l'information géographique légitime et est façonnée par les décisions de planification.*

MOTS-CLES : *Information géographique, Cartographie critique, Technologies, Planification Spatiale Maritime.*

ABSTRACT. *This work consists in studying the life cycle of geographical information as a whole divided into four main stages: data manufacturing, processing of geographical information, its representation and finally its circulation through the case of the Maritime Spatial Planning (MSP). MSP aims to organize the development of the maritime economy and to achieve social objectives while respecting the good ecological status of the seas, based on*

the "best available data". It is therefore based on geographic information and geotechnologies. However, while part of the "reality" is neither described by existing information nor mobilized for planning purposes, and while there are still technical limitations intrinsic to the technologies used, the role of geographic information and geotechnologies remain still little studied. This leads to a lack or even partial representation of space, which critical cartography is constantly questioning. This research contributes to clarifying the informational stakes of planning, in the digital era, and to feed the critical turn at work in this field by an exploratory and qualitative approach. This work focuses on the technical choices made during planning processes and the influences they have on political decisions. By approaching the subject through the socio-technical aspect of planning processes, the relations of domination between the different actors, notably the invisible powers, are brought to light. This work demonstrates that geographic information legitimizes and is shaped by planning decisions.

KEYWORDS: *Geographic Information, Critical Cartography, Technologies, Marine Spatial Planning*

1. La place de l'information géographique dans la planification

Le recours aux informations géographiques peut être quotidien souvent même sans que l'on ne s'en rende compte. C'est d'autant plus le cas dans la planification puisque les informations géographiques et les géotechnologies sont la matière première qui sert à connaître l'espace, à l'analyser, et le moyen technique pour mettre en œuvre une planification. Alors même que la carte n'est plus systématiquement un produit fini ou un produit en soi (Gautreau & Noucher, 2022), dans le contexte de la planification, elle reste l'outil privilégié pour la prise de décision. Ainsi l'examen de la place de l'information géographique et des géotechnologies dans la planification correspond à l'effacement de la carte comme produit fini, mais interroge dans le même temps la réalité de cet effacement. En effet, le tournant numérique de ces dernières années rend encore plus centrales les interrogations qui portent sur les enjeux informationnels dans la gouvernance. Cette communication propose d'étudier la partie immergée des informations géographiques en référence à la métaphore de l'iceberg de Schuurman (2000). Les approches critiques développées au sein des sciences de l'information géographique pointent comment les paramètres sociaux et numériques définissent la technologie et qu'il est contre-productif de ne se concentrer que sur l'un ou sur l'autre. Ainsi, Schuurman fait la distinction entre deux choses : ce qui est émergé, autrement dit, accessible à tous et toutes (souvent, il s'agit de la cartographie finale), et ce qui est immergé, c'est-à-dire, la fabrique des données, les traitements géomatiques des informations, les choix des représentations et la façon de faire circuler les informations. Alors même que les SIG et les informations géographiques associées apparaissent d'autant plus pertinentes dans l'aménagement et la gestion des espaces puisque l'on considère généralement qu'environ 80 % des données sont rattachées à une localisation ou un objet géographique (Longley et al., 1999), une partie du processus de choix des informations géographiques reste floue, voire inaccessible pour la plupart des publics. Les choix qui émanent du champ de l'information géographique sont pourtant déterminants dans la façon dont les décisions sont prises et dans la façon dont l'espace est perçu et représenté. Ceux-ci sont à l'origine des jeux de pouvoir et de relation de domination (Monmonier & Grataloup, 2019) qui structurent les relations sociales. Il est donc nécessaire de les mettre en lumière et renverser l'iceberg pour donner à voir la partie immergée de l'information géographique. Les informations géographiques constituent la matière première pour décrire les activités et planifier les enjeux. Cependant, alors qu'une partie de la réalité n'est ni décrite par les informations existantes ni mobilisée dans le cadre de la planification (St. Martin & Hall-Arber, 2008) et qu'il demeure des limites techniques dans la représentation « des réalités » (Leroy, 2018; Noucher, 2017), le rôle des géotechnologies et des informations géographiques reste paradoxalement peu étudié dans son ensemble. Cela aboutit à une certaine représentation de « la réalité », partielle, voire partielle, que la cartographie critique ne cesse de questionner (Crampton & Krygier, 2006). Ainsi l'approche critique est mobilisée dans ces travaux afin de documenter les relations pouvoir-savoir et de domination dans la planification et notamment les enjeux informationnels que ces relations arbitrent.

Aussi, la planification se tourne vers une véritable numérisation. Mais, alors que l'influence des outils classiques de la planification, tels que les cartes, sur les acteurs et actrices a été mise à jour (Harley, 1988, 1989; Monmonier, 1991; Monmonier & Grataloup, 2019; Wood & Fels, 1992), le rôle des nouvelles technologies dans les relations sociales et politiques commence seulement à être explicité. Les enjeux informationnels de la gouvernance nécessitent d'interroger la façon dont les informations sont questionnées et, au-delà même, si elles le sont. Cela entraîne par extension des réflexions sur la notion de pouvoir et notamment comment il est distribué dans un contexte de gouvernance informationnelle. Par extension, cela questionne la place et le positionnement de l'experte ou de l'expert dans cette gouvernance qui se trouve sollicitée à différentes étapes du processus de planification. Il convient donc d'étudier comme un tout les relations entre le tournant numérique et la planification à travers les apports des études critiques. L'objectif est ainsi de documenter les relations de pouvoir et savoir qui s'entremêlent dans les projets de planification, à travers l'information géographique, et dans le contexte du tournant numérique de la gouvernance maritime.

2. Etude à travers le cas de la planification spatiale marine (PSM)

Cet objectif appliqué à la Planification Spatiale Marine (PSM), qui est un dispositif permettant aux États d'organiser leur espace maritime, nécessite d'autant plus de regarder ce qui se passe dans la partie immergée, au sens propre comme figuré. En effet, la PSM a pour but d'organiser le développement de l'économie maritime et de garantir la réalisation d'objectifs sociaux dans le respect du bon état écologique des mers (Ehler & Douvère, 2009) en se basant sur les « meilleures connaissances scientifiques disponibles ». Ainsi, la planification (spatiale marine) repose essentiellement sur des informations géographiques et des géotechnologies faisant apparaître des enjeux tout au long du cycle de vie de l'information géographique, schématiquement résumé autour de quatre grandes étapes : la fabrication, le traitement, la représentation et la circulation. Ainsi la partie immergée des informations dans la PSM se réfère à la fois aux processus sociotechniques qui ont trait aux informations géographiques et qui ne sont pas encore complètement documentés, mais aussi aux informations qui concernent les dimensions humaines (notamment les activités, mais pas uniquement) et non-humaines (écosystèmes) de l'espace pour lesquelles la disponibilité d'informations est parfois partielle, voire absente. Aussi, l'océan est à la fois un espace naturel tout autant qu'un espace politique (Trouillet, 2004), ce qui en fait un cas d'étude particulièrement intéressant à analyser au crible de l'information géographique, mobilisant à la fois les sciences naturelles ou physiques et les sciences humaines et sociales. Cela est d'autant plus pertinent que les controverses maritimes sont nombreuses (Benvegna et al., 2022) et les recherches étudiant l'impact des décisions politiques sur l'espace maritime, notamment par des approches anthropologiques, ont mis en évidence le caractère très réducteur des approches institutionnelles particulièrement dans la gestion des ressources (e.g., Berkes, 1985). Les résultats de ces études ont considérablement encouragé la prise en compte d'une diversité de pratiques telles que les savoirs locaux ou la gestion communautaire des écosystèmes. Dans ce registre-là, la gestion

des espaces maritimes s'est ouverte à des expérimentations plus intégrées et moins sectorielles en favorisant les articulations entre les savoirs dits « scientifiques », les savoirs pratiques et les décisions (Bremer & Glavovic, 2013; van der Molen, 2018). Pourtant, la PSM reste profondément ancrée dans une démarche positiviste et/ou rationaliste pour laquelle la dimension politique apparaît a priori reléguée au second plan. Ainsi, les géotechnologies occupent différentes fonctions dans la démarche de planification. De la description de l'état initial à la création de scénarios, elles permettent finalement d'aboutir à des mesures de gestion (Fig. 1).

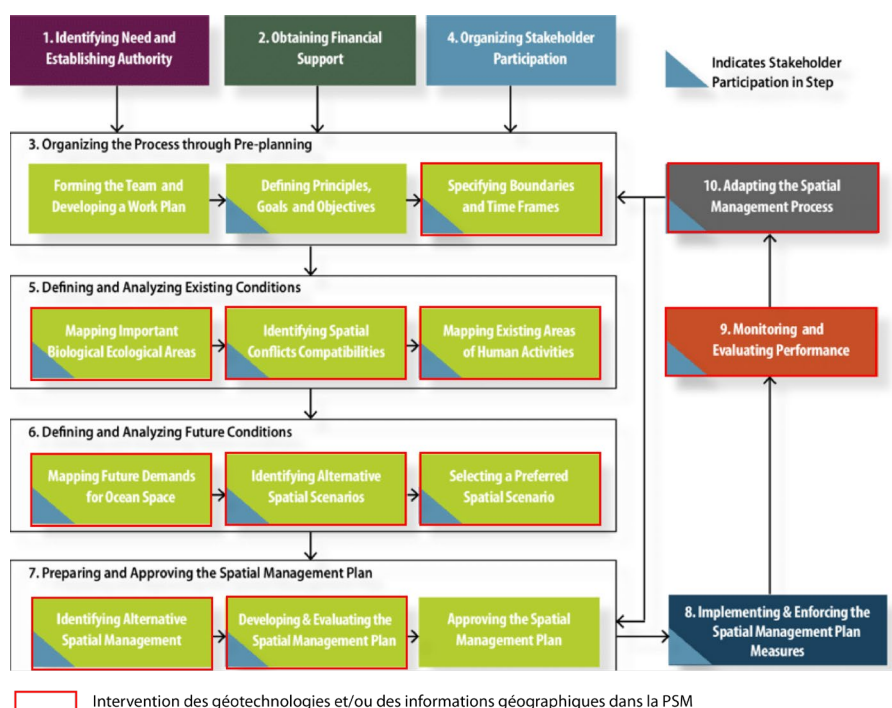


Figure 1. Étapes pas-à-pas pour mettre en place la PSM (source : Trouillet, 2019 d'après Elher & Douvere, 2009)

Alors que les mesures de gestion spatiale et sectorielle sont déjà relativement anciennes en mer, la PSM a émergé comme paradigme intégrateur dominant ces dix ou vingt dernières années (Zaucha & Gee, 2019). En dépit d'une démarche souvent présentée comme une solution à des problèmes divers (e.g., conflits entre usages, protection de l'environnement, développement de l'économie bleue), la PSM n'a fait jusque-là que peu l'objet de débats théoriques, notamment dans le champ social et politique (Jones et al., 2016; Kidd & Shaw, 2014). Des chercheurs et chercheuses ont fait émerger un courant critique dans l'abord de ces nouvelles politiques à l'œuvre sur l'espace maritime. Tafon (2019) a montré que des jeux de pouvoir et de domination sont à l'œuvre dans la mise en place de ces pratiques de planification. Il encourage ainsi à porter un regard critique sur les

intérêts que la PSM permet de favoriser au détriment des autres. L'approche critique revendiquée dans ces travaux a trait à la démarche « émancipatrice et de transformation de la société » (Gintrac, 2012) qui vise à remettre en question des approches analytiques classiques pour étudier les pratiques de planification et la volonté de peser sur les évolutions politiques et sociales des sociétés étudiées (Morange & Calbérac, 2012).

3. Une approche en quatre étapes

La singularité de ce travail tient au fait de passer en revue l'ensemble du cycle de vie de l'information géographique dans la planification. L'approche transversale au cycle de vie de l'information géographique est encore peu documentée dans le champ de la planification et permet de baliser l'ensemble des enjeux relatifs à l'utilisation d'information géographique. De plus, ce travail s'ancre dans la cartographie critique, mais ne s'y limite pas puisqu'il ne se restreint pas exclusivement aux rapports de domination. Ce travail relève donc d'un double objectif, à la fois de pointer un ensemble de transformations techniques et sociales liées à la production de cartes nécessitant de documenter les relations de domination entre la production cartographique et ses usages, mais aussi de l'appliquer sur un cas d'étude encore peu investi par les questions de cartographie critique, pourtant au cœur des logiques de savoir-pouvoir. L'analyse du rôle joué par les informations géographiques et les géotechnologies dans la planification nous amène à définir les différentes phases du processus de planification où elles sont mobilisées. À la croisée entre les différentes fonctions des systèmes d'informations géographiques (Jégou, 2022) et par extension des géotechnologies mobilisant par définition les informations géographiques et des différentes phases d'un processus de planification (voir Fig. 1) (Ehler & Douvère, 2009) qui fait usage de l'information géographique et des géotechnologies (Trouillet, 2019), nous avons dégagé quatre étapes principales mobilisant l'information géographique dans la planification. La « fabrique » regroupe les opérations de création, d'acquisition et de structuration qui permettent de créer et de convertir les données (puis les informations) dans des géotechnologies afin de les exploiter. Le « traitement » correspond à la principale fonctionnalité des SIG et recoupe toutes les opérations effectuées sur les informations géographiques en vue de répondre à des questions comme le croisement de différentes informations. La « représentation » correspond aux choix qui sont opérés sur les outils pour afficher les informations géographiques ; cette étape s'apparente souvent à des représentations sous forme de cartes. Enfin, la « circulation » correspond aux modes de diffusion des informations géographiques. Il est toutefois entendu que ces étapes sont à la fois schématiques et de plus en plus artificielles. En effet, le tournant numérique « a pour effet le brouillage de la limite entre la notion de production et celle de circulation de l'information (...) [laquelle] devient une des modalités de production de l'information » (Gautreau, 2021). C'est pourquoi il est devenu essentiel d'étudier l'ensemble du processus et non seulement les étapes indépendamment les unes des autres. Ainsi, la figure 2 fait référence au « cycle de l'information géographique » afin de documenter à la fois l'existence

d'un continuum (au détriment de ruptures marquées entre les étapes) et de sa non-linéarité.

Pour contribuer à alimenter les débats théoriques qui ont trait à la planification, cette communication propose la présentation d'analyses menées afin de comprendre le rôle joué par l'information géographique en matière de planification, notamment spatiale et maritime, et la manière avec laquelle ce rôle pèse sur les choix d'aménagement. Ces recherches concernent plus spécifiquement les aspects sociotechniques de la planification, c'est-à-dire l'imbrication entre l'élaboration des connaissances et la dimension sociale (Callon, 1986). Autrement dit, nous présenterons l'analyse des opérations par lesquelles l'information géographique submerge la planification et irrigue les enjeux pouvoir-savoir. En effet, le mode de gouvernance des espaces maritimes soulève des enjeux majeurs pour les usagers ou usagères de la mer et, plus globalement, pour les sociétés. Aussi, il convient d'étudier précisément les logiques sous-jacentes à la mise en place de cette politique. Ces recherches visent à répondre à la question suivante : dans quelle mesure l'utilisation des informations géographiques et des géotechnologies affecte-t-elle la gouvernance maritime et impacte les relations de domination présentes dans une démarche de planification multi-acteurs et actrices ?

Afin de répondre à cette problématique, une démarche méthodologique a été mise en place. Chaque étape du cycle de vie de l'information géographique a fait l'objet d'une étude de cas qui découle sur la mise en place d'une expérimentation et d'une méthode d'analyse particulière. Cette démarche a été mise en œuvre dans le but de vérifier le rôle de l'information géographique dans la planification transversalement à son cycle de vie. Une synthèse de la démarche méthodologique est proposée ci-après (Fig. 2).

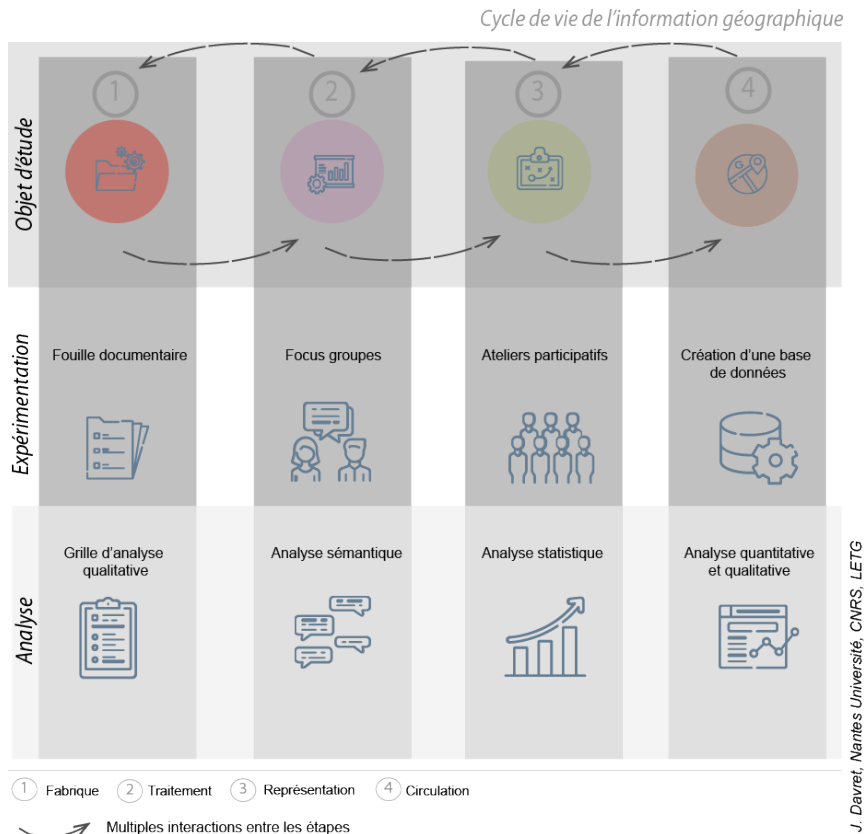


Figure 2. Démarche méthodologique transversale au cycle de vie de l'information géographique

Ces quatre expérimentations ont pour objectif de montrer l'importance des dynamiques sociotechniques qui influent sur les processus spatiaux, mais aussi la difficulté que présentent ces dynamiques pour y établir des conclusions définitives sur les effets sociotechniques engendrés. L'originalité de ce travail tient au fait d'étudier le cycle de l'information transversalement en mobilisant une large diversité d'expérimentations et de méthodes d'analyse adaptées au cas d'étude choisi. Une approche qui fait appel à différentes échelles d'analyse a été privilégiée afin de rendre compte des multiples formes de planification. La diversité des quatre expérimentations peut être appréciée dans le tableau 1 suivant :

Tableau 1. Contenu des expérimentations

Étape	Objet d'étude	Collecte de données	Méthode d'analyse	Échelle
Fabrique	Données environnementales de la PSM	38 plans marins	Quantitative et qualitative	Union européenne
Traitement	Caractéristiques de l'activité de pêche et de plaisance	5 focus groupes réunissant 30 acteurs et actrices	Analyse sémantique et qualitative de discours	France
Représentation	Cartographies de la PSM	5 ateliers participatifs réunissant 30 acteurs et actrices	Analyse statistique de données qualitatives	France
Circulation	Géoportails	Construction d'une base de données de 110 géoportails	Quantitative et qualitative	Monde

3.1. La fabrique des données dans la planification : gouvernance informationnelle de l'environnement

La première expérimentation se focalise sur la fabrique de l'information géographique ou plutôt témoigne de la difficulté d'étudier la fabrique de l'information géographique. Elle démontre ainsi que les « meilleures connaissances scientifiques disponibles » ne sont souvent pas des connaissances adaptées à l'objet même de la planification. Les informations géographiques sont finalement un des instruments mobilisés par les décideurs ou décideuses pour imposer une certaine vision des espaces, souvent une vision économique dominante. De plus, l'hégémonie des informations dites « officielles » est peu, voire pas, remise en cause ou même simplement questionnée. Ainsi les apports du champ de la *political ecology* mettent à jour les mécanismes qui structurent l'utilisation des informations géographiques et notamment la façon dont cette utilisation est guidée par des relations de domination entre acteurs et actrices. Cette expérimentation permet aussi de montrer les antagonismes entre les approches conceptuelles pour mettre en place une PSM et la difficile réalité des informations géographiques disponibles au regard des besoins théoriques. De fait, ce volet de l'étude permet de montrer que la PSM ne se décide pas sur la base des « meilleures informations disponibles », mais plutôt, que les décisions politiques sont justifiées à l'aide des données. Or la sélection des données et les traitements opérés reposent sur des choix partiellement subjectifs, qu'ils soient conscients ou inconscients, mais surtout non explicités et débattus. Cela change

considérablement le rapport de force entre les décisions politiques et les preuves scientifiques. Nous démontrons avec cette expérimentation comment les documents de planification révèlent en réalité les intentions politiques et la flexibilité des directives qui guident ces démarches.

3.2. Ouvrir la boîte noire : analyse de la compréhension des traitements cartographiques par les acteurs et actrices du processus de planification

Une deuxième expérimentation a été mise en place en utilisant le cas de la pêche et de la plaisance pour investiguer la question des traitements, autrement dit ouvrir la « boîte noire » de l'information géographique. Pour cela, le dispositif méthodologique propose de déconstruire les traitements opérés sur les données afin d'analyser les différents résultats pouvant être obtenus. À l'aide de la théorie de l'agencement, nous identifions différentes relations pouvoir-savoir dans la planification, mais également l'existence d'une approche de cartographie critique déployée par les acteurs et actrices à propos des traitements visibilisés sur les cartes dites officielles. Cette expérimentation révèle la manière dont les réalités complexes des activités de pêche et de plaisance deviennent une réalité singulière retranscrite par des informations géographiques souvent standardisées et réduites à des intérêts par exemple marchands laissant de côté les aspects socioculturels des pratiques, notamment pour le cas des pêches artisanales. Cela entraîne donc une homogénéisation des activités lors de leur cartographie, et traduit une certaine vision utilitariste de la planification. Les apports de ce volet de l'étude illustrent également les relations de domination entre acteurs et actrices en fonction notamment des connaissances cartographiques qui ont structuré certains débats lors de l'expérimentation.

3.3. Représentation des informations géographiques à travers les cartes de la planification (spatiale marine)

Une troisième expérimentation permet de démontrer comment et pourquoi les cartes sont utilisées en tant qu'instruments de l'action publique pour des projets de planification et dans quelle mesure cela contribue à façonner les relations de domination entre acteurs et actrices et surtout à légitimer les décisions en cristallisant à la marge les débats et en laissant aux cartes le lourd rôle d'arbitrer les décisions. La carte joue donc un double jeu en étant un outil d'aide à la décision en même temps qu'elle agglomère les débats autour des représentations qu'elle propose. Son rôle est d'autant plus disproportionné que les acteurs et actrices ont une capacité à s'approprier cet instrument très hétérogène. Ce volet de l'étude a permis d'exposer de quelle manière la construction, l'utilisation et la réception des cartographies interfèrent avec des réseaux sociotechniques et sociopolitiques.

3.4. Circulation des informations géographiques à travers les géoportails

La constitution d'une base de données des géoportails en lien avec la planification maritime dans le monde a permis d'étudier le tournant numérique qu'a

pris la PSM. Mais surtout, cela a permis d'informer sur la diversité des fonctionnalités présentes dans les géoportails et la façon dont elles interfèrent avec la prise de décision politique. En effet, ce travail a montré que, contrairement à ce qui est avancé par les ouvrages théoriques sur la PSM (Ehler & Douvère, 2009), les informations qui circulent dans les géoportails permettent de justifier les décisions politiques. Ainsi, l'enjeu informationnel se retrouve relégué après la phase de prise de décision. Ainsi, avant l'approbation du plan, la circulation des informations géographiques est opaque. Il s'est avéré que les géoportails donnent à voir la planification, mais ne permettent pas de la faire, ce que confirme également l'illusion participative qui émane des accès aux fonctionnalités interactives en ligne pour les acteurs et actrices. Ainsi, le géoportail est davantage un produit du plan que le plan n'est un produit du géoportail.

4. Conclusion : Donner à voir les enjeux informationnels de la PSM

À l'aide de la cartographie critique et des *Science and Technology Studies* (STS), ce travail a cherché à donner une représentation du réseau sociotechnique de l'information géographique en abordant de front les paramètres sociaux et techniques qui interfèrent et influent sur l'utilisation des géotechnologies et des informations géographiques. Il a permis de rendre compte du volet socio-politique des technologies et du volet technologique qui influe sur la composante socio-politique et notamment sur les prises de décision. Il a également montré que les facteurs sociotechniques sont une part importante et inhérente au processus de décision. Ils constituent souvent la partie immergée de l'information géographique, elle-même enfermée dans une boîte noire, qu'il convient de décrire et d'analyser pour mettre en lumière les biais et instrumentalisation humains et technologiques qui en émanent. De plus, le tournant numérique de l'information géographique rend l'étude de ce système sociotechnique d'autant plus difficile que la numérisation multiplie les acteurs et actrices potentiels, mais aussi démultiplie les fonctionnalités qui leur sont accessibles.

Le cycle de vie de l'information géographique passé au crible de la cartographie critique et des STS dans ce travail confirme les relations de domination qui émanent de la partie immergée de l'information géographique. Si l'information géographique peut être garante de justice socio-spatiale, elle encourage aussi à l'illusion participative qui découle de ses usages dans les politiques de planification et contribue dans ce sens à appauvrir le rapport que les groupes humains entretiennent avec l'espace sur lequel ils vivent. En effet, l'information géographique transmet une forme rigide de la représentation de l'espace qui ne permet pas d'utiliser des représentations alternatives (Gautreau, 2021). Les solutions pour que la planification soit plus démocratique, nécessitent non seulement une ouverture à chaque étape du cycle de vie de l'information géographique aux parties prenantes et au grand public, mais surtout une modification de la gouvernance informationnelle telle qu'elle est conçue aujourd'hui afin d'adapter la fabrique de l'information, d'ouvrir la boîte noire des traitements géographiques, de mettre en lumière les limites des représentations cartographiques et leur chaîne de circulation.

Références :

- Bremer, S., & Glavovic, B. (2013).** Mobilizing Knowledge for Coastal Governance : Re-Framing the Science–Policy Interface for Integrated Coastal Management. *Coastal Management*, 41(1), 39-56. <https://doi.org/10.1080/08920753.2012.749751>
- Callon, M. (1986).** Éléments pour une sociologie de la traduction : La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc. *L'Année sociologique (1940/1948-)*, 36, 169-208.
- Crampton, J. W., & Krygier, J. (2006).** An Introduction to Critical Cartography. *ACME: An International Journal for Critical Geographies*, 4(1), 11-33.
- Ehler, C., & Douvère, F. (2009).** *Marine spatial planning : A step-by-step approach toward ecosystem-based management*. Commission océanographique intergouvernementale. <https://unesdoc.unesco.org/ark:/48223/pf0000186559>
- Gautreau, P. (2021).** *La Pachamama en bases de données : Géographie politique de l'information environnementale contemporaine*. Éditions de l'IHEAL. <https://doi.org/10.4000/books.iheal.9362>
- Gautreau, P., & Noucher, M. (2022).** L'adieu aux cartes. Refaire de la cartographie politique aux temps du numérique. In B. Debarbieux & I. Hirt (Éds.), *Politiques de la carte* (p. 77-99). ISTE-Wiley.
- Gintrac, C. (2012).** Géographie critique, géographie radicale : Comment nommer la géographie engagée ? *Carnets de géographes*, 13.
- Harley, B. (1988).** Silences and secrecy : The hidden agenda of cartography in early modern Europe. *Imago Mundi*, 40(1), 57-76. <https://doi.org/10.1080/03085698808592639>
- Harley, B. (1989).** Deconstructing the map. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 26(2), 1-20. <https://doi.org/10.3138/E635-7827-1757-9T53>

- Jégou, L. (2022).** *Structurer, analyser et représenter des données spatialisées.* Introduction à QGIS. <https://geotests.net/cours/qgis/fr>
- Jones, P. J. S., Lieberknecht, L. M., & Qiu, W. (2016).** Marine spatial planning in reality : Introduction to case studies and discussion of findings. *Marine Policy*, 71, 256-264. <https://doi.org/10.1016/j.marpol.2016.04.026>
- Kidd, S., & Shaw, D. (2014).** The social and political realities of marine spatial planning : Some land-based reflections. *ICES Journal of Marine Science*, 71(7), 1535-1541. <https://doi.org/10.1093/icesjms/fsu006>
- Leroy, Y. (2018).** *Cartographie critique de réalités géographiques* [Thèse de doctorat]. Université de Nantes.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (1999).** *Geographical information systems* (2nd ed). J. Wiley.
- Monmonier, M. (1991).** Ethics and Map Design : Six Strategies for Confronting the Traditional One-Map Solution. *Cartographic Perspectives*, 10, Art. 10. <https://doi.org/10.14714/CP10.1052>
- Monmonier, M., & Grataloup, C. (2019).** *Comment faire mentir les cartes* (Autrement).
- Morange, M., & Calbérac, Y. (2012).** Géographies critiques » à la française » ? *Carnets de géographes*, 16.
- Noucher, M. (2017).** *Les petites cartes du web : Analyse critique des nouvelles fabriques cartographiques.* Éditions Rue d'Ulm.
- Schuurman, N. (2000).** Trouble in the heartland : GIS and its critics in the 1990s. *Progress in Human Geography*, 24(4), 569-590. <https://doi.org/10.1191/030913200100189111>
- St. Martin, K., & Hall-Arber, M. (2008).** The missing layer : Geo-technologies, communities, and implications for marine spatial planning. *Marine Policy*, 32(5), 779-786. <http://dx.doi.org/10.1016/j.marpol.2008.03.015>

- Tafon, R. (2019).** *The “Dark Side” of Marine Spatial Planning A study of domination, empowerment and freedom through theories of discourse and power.* <http://urn.kb.se/resolve?urn=urn:nbn:se:sh:diva-38051>
- Trouillet, B. (2004).** *La « mer côtière » d'Iroise à Finistère : Étude géographique d'ensembles territoriaux en construction* [Thèse de doctorat]. Université de Nantes. <https://www.theses.fr/2004NANT3031>
- Trouillet, B. (2019).** Aligning with dominant interests : The role played by geotechnologies in the place given to fisheries in marine spatial planning. *Geoforum*, S0016718519303008. <https://doi.org/10.1016/j.geoforum.2019.10.012>
- van der Molen, F. (2018).** How knowledge enables governance : The coproduction of environmental governance capacity. *Environmental Science & Policy*, 87, 18-25. <https://doi.org/10.1016/j.envsci.2018.05.016>
- Wood, D., & Fels, J. (1992).** *The power of maps.* Guilford Press.
- Zaucha, J., & Gee, K. (Éds.). (2019).** *Maritime Spatial Planning : Past, present, future.* Springer International Publishing. <https://doi.org/10.1007/978-3-319-98696-8>

Pour un dialogue entre la géomatique et les humanités numériques : propositions pour contribuer à des pratiques géo-numériques critiques

Juliette Morel¹, Baptiste Hautdidier²

1. Lab'Urba, Université Paris-Est Créteil,
61 av. du Général de Gaulle, 94000 Créteil
Juliette.morel@u-pec.fr

2. Inrae, UR ETTIS
50, Avenue de Verdun, Gazinet F-33612 Cestas Cedex, France
baptiste.hautdidier@inrae.fr

RÉSUMÉ. Comme en témoignent plusieurs articles récents (Gautreau et Noucher 2022; Mahmoudi et Shelton 2022), l'approche critique des sciences de l'information géographique ne consiste plus seulement à identifier et documenter les « limites » des SIG (démarche maintenant ancienne car initiée dès les années 1990), mais également à faire des propositions concrètes pour « Refaire de la cartographie critique aux temps du numérique » (Noucher et Gautreau). Cet article entend s'inscrire dans ce sillage en faisant le point sur ce que la géomatique peut concrètement tirer des expériences d'humanités numériques (HN) afin de rafraichir certaines méthodes et pratiques, et de participer à les rendre davantage « critiques » ou simplement réflexives. Nous proposons un état de l'art qui dégagerait l'intérêt d'un dialogue interdisciplinaire entre la géomatique et les humanités numériques et proposerait des processus méthodologiques et techniques concrets développés du côté des HN et qu'il pourrait être intéressant d'importer ou de développer davantage du côté de la géomatique.

ABSTRACT. As shown in several recent articles (Gautreau and Noucher 2022; Mahmoudi and Shelton 2022), the field of critical geographic information sciences has gone well beyond its initial goal of identifying and documenting the "limits" of GIS (an approach initiated as early as the 1990s), extending its interests to concrete proposals for "redoing critical cartography in the digital age" (Noucher and Gautreau). We intend to follow this lead by taking stock of advances in digital humanities (DH) that may be actionable for the refreshment of methods and practices in GIScience, helping in making them more "critical" or simply reflexive. We propose a state of the art on the potential for interdisciplinary

2 SAGEO'2023 – Québec juin 2023

dialogue between GIS and DH, highlighting methodological and technical processes developed on the side of HN that should be of interest for GIS practitioners.

MOTS-CLÉS : approche critique, géomatique, humanités numérique, interdisciplinarité.

KEYWORDS: critical approach, GIScience, digital humanities, interdisciplinarity.

1. Introduction

“[The archeological project] Pleiades is concerned about places as humans have understood them in time, it is not concerned about points, lines and polygons in some database”, Tom Elliott, conférence « The Pleiades gazetteer of ancient places: inventing an “Un-GIS” and its uses »

1.1. Les limites des Sciences de l’Information géographique identifiées par les approches critiques depuis 30 ans

La réflexivité critique vis-à-vis de la cartographie numérique et de la géomatique est déjà ancienne. Elle fut entamée dans les années 1990 par deux événements majeurs : la rencontre en 1993 de Friday Harbor (États-Unis) entre défenseurs et détracteurs des SIG autour de la question « *GIS and society* » ; la parution en 1995 de *Ground Truth*, un ouvrage dirigé par John Pickles. Cette approche reste aujourd’hui marginale mais s’est depuis constituée en tant que communauté épistémique à part entière. Les diverses études se fondent sur des constats communs et régulièrement réitérés (Desbois 2015; Goodchild 2015; Wilson 2015; Joliveau 2004; Schuurman 2000; Wilson 2017; Gautreau et Noucher 2022). Dans des années 80, le contexte d’une remise en cause de l’objectivité et du positivisme dans les SHS a pu se concrétiser par une attaque de la géomatique et de la cartographie numérique. Portées à la fois par la fascination technologique et des intérêts politico-économique, elles étaient alors en plein essor, estampillées d’un « brevet de scientificité » par ses défenseurs (Desbois 2015). La critique se fonde également largement sur le fait que la cartographie est un domaine de très fortes relations entre pouvoir et savoir, ce que le passage au numérique a exacerbé du fait de son coût technologique. Si le développement des SIG s’est à l’origine fait dans une grande porosité avec la sphère militaire, l’instrumentalisation académique de méthodes supposément valorisables et transférables a été manifeste dans le contexte des années Reagan et Thatcher (Desbois 2015). On peut considérer que la première guerre du Golfe a constitué un autre jalon d’une utilisation marquée de manière croissante par des objectifs de sécurité et de surveillance. Si ces technologies sont passées rapidement au civil, elles sont longtemps restées, du fait de leur coût à la fois financier et technique très élevé, confinées à un usage élitare par les politiques, militaires et universitaires (Goodchild 2015). Ce contexte de développement des SIG a des conséquences sur les processus méthodologiques proposés par ces outils et les visions du monde et de l’espace qu’ils véhiculent. Henri Desbois par exemple dénonce l’« hypertrophie » de la partie méthodologique au détriment de l’objet étudié, qu’il explique par le coût d’entrée technique dans les logiciels. De fait, il a longtemps subsisté un certain tropisme dans le développement des SIG vers les informaticiens et les applications commerciales, au détriment de géographes aux

visées plus académiques. Il en découle une prégnance de la technique, des normes imposées par le secteur commercial et de nouvelles conditions de propriété des données, notamment spatiales (Desbois 2015; Goodchild 2015). La critique féministe, quant à elle, a pu dénoncer l'hégémonie de ce regard surplombant, omniscient, principalement masculin – et donc dominant – de la cartographie, ainsi que son européocentrisme et sa proximité avec l'économie capitaliste (Pavlovskaya et Martin 2007). Avec les bouleversements des années 2000, dont l'avènement du géoweb, les questions se compliquent encore car s'ajoute à tout cela une couche de flou concernant l'identité de qui produit, utilise ou modifie les informations géographiques, ce qui vient compliquer la traçabilité de la production cartographique et donc l'identification des visions qu'elle véhicule. Dans ces conditions, comment faire pour contextualiser les cartes et informations géographiques, comment savoir d'où viennent les données, comment elles ont été construites et cartographiées ? Et donc, *in fine*, comment pouvoir continuer à les critiquer ? (Noucher, Gautreau 2022) On peut enfin mentionner la critique écologique que l'on peut faire des technologies géomatiques : la Terre a-t-elle les capacités suffisantes pour supporter l'augmentation technologique qu'elles induisent (*data centers*, calculs, réseaux de capteurs diverses, etc.) ?

D'un point de vue méthodologique, la géomatique critique dénonce la simplicité à la fois géométrique et sémantique des formalisations de l'espace permises par les systèmes d'information géographiques : « Comment la science de l'information géographique peut-elle prétendre représenter l'humanité quand ses bases de données ne contiennent que des caricatures ? » s'interroge par exemple Goodchild (2015, traduction personnelle). En effet, le primat technique d'une information géographique précise et figée, hermétiquement organisée en points, lignes ou polygones, ne laisse pas beaucoup de possibilités pour intégrer les localisations floues, incertaines ou changeantes.

1.2. Comment faire de la géomatique critique aujourd'hui ?

Dans le sillage de l'identification et de la documentation de ces limites et en parallèle de la promulgation de socio-histoires des SIG (Wilson 2015) et de méthodes pour réfléchir à la géomatique « en train de se faire » (Kelly et Bosse 2022), des pistes ont été ouvertes pour *faire* de la géomatique critique (Mahmoudi et Shelton 2022; Gautreau et Noucher 2022; Wilson 2015). Nous pouvons en citer quelques-unes, sans prétendre que cette liste soit exhaustive. La première réside dans la diversification des sujets de cartes ou des systèmes d'informations géographiques, pour la représentation et le traitement d'informations concernant les dominés : pauvres, femmes, minorités. C'est la direction ouverte notamment par les expérimentations radicales de la *Detroit Geographical Expedition* (Bunge 1971) et par une lignée de cartographes féministes (Pavlovskaya et Martin 2007; Kwan 2002). Une autre des pistes principales tient à la démocratisation de la cartographie numérique par le web et la généralisation des équipements informatiques, par exemple à travers la diffusion très large de l'utilisation de Google Earth et Google Maps, à travers la création hétéroclites de « petites cartes du web » (Noucher 2017),

ou encore à travers des processus de création de l'information géographique contributifs (ou *Crowdsourcing*), comme l'initiative OpenStreetMap. Ces pratiques ont tendance à remplacer le surplomb dominant d'une production cartographique centralisée par des démarches *bottom-up*. Des attentes optimistes ont sur ce plan été exprimées envers les projets de cartographie participative, qui peuvent ou non passer par les technologies numériques, notamment avec des peuples autochtones. S'il s'agit alors de remettre sur les cartes des toponymes, des lieux, des occupations, des significations de l'espace alternatifs à ceux que l'on trouve dans les représentations cartographiques classiques, cet objectif ambitieux peut peiner à se concrétiser dans des réalisations convaincantes (Hirt 2009; Gautreau 2021).

Une autre large piste de pratique critique des SIG est le travail à l'intégration en leur sein de données alternatives : soit du point de vue de leurs sources – comme c'est le cas pour les informations géographiques participatives ou contributives déjà citées ; soit du point de vue leur formalisation géométrique ou sémantique pour intégrer davantage de flou, de souplesse dans les classifications ou d'incertitude ; ou encore du point de vue de la nature de la connaissance impliquée, pour intégrer à ces systèmes largement quantitatifs davantage d'informations d'ordre qualitatives, multimédias, issues de méthodes mixtes de recherche (Cope et Elwood 2009). De cette manière, une multitude de « x-GIS », prônant des méthodes et des points de vue différents a été proposée (SIG queer, féministes ou encore qualitatifs) (Wilson 2017). Dans tous les cas, ces initiatives éprouvent jusqu'où on peut élargir et rendre inclusive l'une des prérogatives centrales de tous les SIG : à savoir la capacité de croisement de données de types, de sources, de thématiques hétérogènes grâce à une dimension spatiale commune.

1.3. Une proposition : aller voir ce que l'on peut retirer des HN

En complément de ces expérimentations, nous proposons un regard interdisciplinaire pour renouveler nos pratiques géomatiques, nous tournant pour ce faire vers les techniques et méthodes développées dans les humanités numériques (HN). Dans la droite ligne du *spatial turn* advenant dans les sciences humaines à partir des années des années 1980, les humanités numériques spatialisées (HN spatialisées) – ou *Geohumanities* en anglais – se sont constituées en un champ scientifique spécifique à partir des années 2010 (Bodenhamer, Corrigan, et Harris 2010; Murrieta-Flores et Martins 2019; Brando et al. 2021). Les HN spatialisées entendent faire des liens et structurer entre elles des initiatives parfois déjà anciennes, qui font rencontrer outils numériques et sciences humaines (principalement histoire, archéologie, linguistique, littérature, narratologie, sciences de la documentation, et plus marginalement design, sociologie, etc.) avec la dimension spatiale. Ce dialogue se déploie dans un champ des humanités numériques lui-même en voie de stabilisation et d'institutionnalisation, avec ses propres réseaux et institutions, rencontres, colloques et publications¹. On comprend que dans ce cadre, les circulations entre humanités

¹ On peut citer par exemple les institutions européennes DARIAH et CLARIN, l'association française Humanistica ou la TGIR Huma-Num ; les colloques DH

numériques et géomatique aient été permanentes mais largement asymétriques. Or, de la même manière que les HN spatiales s'interrogent sur les usages et apports des outils numériques de l'analyse spatiale et la visualisation spatio-temporelle dans les sciences humaines et sociales (Brando et al. 2021), nous voudrions poser la question inverse des apports – ou retours – potentiels qu'il est possible de dégager des expérimentations en humanités numériques, afin de complexifier, nuancer, enrichir les processus et outils en géomatique.

Nous organisons ces apports en quatre grandes catégories. La première catégorie regroupe les leçons qui peuvent être tirées en matière de modélisation et formalisation, à partir des expérimentations en archéologie et en histoire. La deuxième catégorie regroupe les leçons qui peuvent être tirées en termes de contextualisation des informations géographiques à partir de projets dans les domaines des archives et des bibliothèques. La troisième catégorie regroupe les leçons que la géomatique peut tirer d'expérimentations linguistiques et narratologiques afin d'intégrer les informations issues des dimensions textuelles et narratives de nombreuses sources de données géographiques. Enfin la quatrième catégorie que nous dégagons concerne les inspirations que l'on peut dégager de gros projets transversaux et interdisciplinaires en humanités numériques spatialisées qui encouragent l'intégration de vision alternatives et assume un rapport sensible au monde. Nous avons choisi de mettre en valeur des projets principalement francophones qui sont exemplaires de chacune de ces catégories – ce qui ne veut pas dire que la liste soit exhaustive, ni que de tels questionnements n'aient pas été soulevés, voire résolus, par d'autres, y compris au sein de la géographie et de la géomatique plus classiques.

2. Remettre en question le modèle SIG, vers d'autres modélisations spatio-temporelles : leçons venues de l'archéologie et de l'histoire.

2.1. Le défi de l'imperfection ontologique des données en histoire et archéologie

En archéologie, les outils de la géomatique ont été utilisés dès les années 1980 d'abord pour localiser artefacts et fouilles, puis rapidement pour « poser des questions croisées complexes » en utilisant les méthodes d'analyse spatiales, enfin pour explorer les terrains de manière préventive grâce à la télédétection ou les relevés 3D par lidar (Brando et al. 2021). Au-delà des défis technologiques que ces utilisations impliquent, les besoins très spécifiques de l'archéologie et de l'histoire en termes de modélisation spatio-temporelle ont provoqué une réflexion intense sur la formalisation de leurs données spatio-temporelles. Les archéologues et historiens ont besoin de référencer des traces matérielles ou des mentions documentaires, c'est-à-dire des matériaux concrets – ruines de bâtiments, vestiges, artefacts, livres, textes

de l'ADHO chaque année, les THATCamp ; ou encore la revue française *Humanités Numériques*.

– posant de véritables défis pour être formalisés numériquement. Ces matériaux sont multidimensionnels : ils ont des caractéristiques temporelles, géographiques, géométriques et sémantiques complexes. Mais leurs traces et mentions sont souvent très lacunaires et une ou plusieurs de ces dimensions manquent la plupart du temps : l'emprise des vestiges excavés peut ne pas correspondre à celle du bâtiment en entier, la date de construction peut ne pas être connue précisément, les usages non plus. En même temps, ce même bâtiment peut avoir plusieurs phases de construction, donc plusieurs géométries correspondant à plusieurs datations (elles-mêmes inconnues précisément) et peut changer de fonctionnalités dans le temps. Par ailleurs, le propre de l'archéologie et l'histoire est qu'on ne peut connaître une thématique de manière absolument exhaustive : l'archéologie ne mettra jamais à jour tous les temples du monde ancien et ne connaîtra jamais que ceux dont la trace a été conservée et découverte ; l'histoire n'aura jamais accès à des sources textuelles décrivant l'entièreté du monde médiéval et est soumise aux sources conservées. D'un point de vue la localisation, il existe beaucoup de lieux et espaces du monde médiéval et ancien dont les archéologues ou historien.nes attestent l'existence, qui ont des noms et des caractéristiques bien connues (par des sources qui les mentionnent) mais qu'on ne peut pas localiser sur le terrain aujourd'hui (car aucun vestige n'a été découvert). Et il est fréquent que persistent l'incertitude sur la correspondance entre les fouilles et les mentions documentaires et donc sur d'identification d'artefacts ou de vestiges excavés (Elliott 2018; du Mouza et Lamasse 2022). En outre, au-delà du caractère parcellaire de cette connaissance, il s'agit d'une connaissance en cours de constitution, évoluant sans cesse et qui sera sans doute modifiée dans le futur. Les archéologues sont très conscients qu'ils modélisent un « état de connaissance » [*state of the knowledge*] (Elliott 2018).

2.2. Vers des Un-GIS

On voit que les disciplines historiques et archéologiques sont ontologiquement marquées par d'importants défis en termes de numérisation... Elles y ont apporté des réponses intéressantes. Certains archéologues et historien.nes (principalement médiévistes) ont en effet proposé de remettre en question le modèle des systèmes d'information géographique (SIG) dans le sens spécifique – mais néanmoins hégémonique en géomatique – des logiciels de même nom (QGIS, ArcGIS). Ils proposent d'en sortir car ils estiment que ces systèmes imposent de trop grandes contraintes de modélisation pour être capables de répondre aux défis listés précédemment, notamment du fait de la différenciation des informations en couches hermétiques d'informations aux implantations (points, lignes ou polygones), granularités et thématiques distinctes. C'est une idée qu'on retrouve par exemple dans la proposition de *un-GIS* (traduisible par « non-SIG » en français) faite dans le cadre du projet de référence Pleiades, qui vise à rassembler la connaissance archéologique et historique sur les lieux anciens (Bagnall et al. 2006) ; dans la structure des données du projet ANR Col&Mon, qui visait à fusionner plusieurs bases de données existantes concernant des établissements religieux médiévaux et à en étudier les caractéristiques spatio-temporelles d'implantation à travers un modèle conceptuel en « facettes » (Deflou-leca et Massoni s. d.) ; ou encore dans le modèle

conceptuel de données OH_FET développés par les archéologues de l'Université de Tours (Rodier et Saligny 2010). Toutes ces modélisations sont fondées sur un double principe : d'une part l'atomisation de l'information et d'autre part sa description de manière tridimensionnelle – espace, temps, thématique (Peuquet 1994). Cela permet plusieurs choses : premièrement, l'atomisation permet la différenciation entre les traces spatio-temporelles ou mentions documentaires (éléments empiriques récoltés sur le terrain, lors des fouilles ou dans les textes) et les objets géo-historiques qu'elles renseignent (issus de l'interprétations de ces traces). Les sources concrètes de l'information, donc les éléments concrets dont on dispose, deviennent les entités centrales de la base de données – les plus petites unités atomisées – et l'interprétation, les objets reconstitués par l'analyse, apparaissent par des relations entre ces éléments. L'interprétation, la connaissance plus générale retrouvent ainsi leur juste place : celle d'abstractions reconstruites à partir d'observations concrètes. Deuxièmement, l'atomisation et la reconstruction des informations à différents niveaux par des relations entre les éléments atomisés permettent d'intégrer les données spatio-temporelles exactement telles qu'on les a, sans subir de généralisation. Ainsi est-il possible de mettre ensemble des données de granularités spatiales ou temporelles hétérogènes, mais également de reconstruire des spatialités ou des temporalités par voie d'héritage entre éléments mis en relation. Troisièmement, la tridimensionnalité des éléments stockés, qui a pour principe de mettre sur le même plan les trois dimensions (espace, temps, thématique), permet d'une part une liberté d'interrogation des données, par n'importe laquelle ou plusieurs de ces dimensions, et d'autre part l'intégration à part entière dans le modèle spatio-temporel, et donc leur prise en compte dans les traitements, des informations lacunaires (à partir du moment où elles sont au moins définies une des trois dimensions), même celles pour lesquelles il manque l'espace ce qui serait éliminatoire dans le cadres de SIG classiques. Enfin, les principes cités permettent un certain agnosticisme envers le travail concret de la structuration des données : ils peuvent aussi bien être traduits par des bases de données relationnelles (dans le cas de Col&Mon ou de l'*Atlas Historique du Limousin* par exemple (Morel, Crouzevialle, et Massoni 2020)) que par des graphes du web des données (dans le cas de Pleiades), selon les technologies maîtrisées et les usages visés.

Ces principes, car ils relâchent un peu la contrainte de la structuration SIG et permettent de réintroduire une conception constructiviste de la connaissance dans les modélisation spatio-temporelles, méritent d'être explorés en détails dans la perspective d'une pratique réflexive de la géomatique.

3. Référencer de manière complexe les sources et le chemin des données dans les systèmes d'informations : leçons venues des sciences de la documentation et des archives

Dans cette même veine constructiviste, et pour répondre à l'une des critiques récurrentes des SIG, notamment féministe, concernant le point de vue des SIG, une des pistes importantes de pratiques critiques en géomatique est la toujours plus grande documentation des contextes de production de la connaissance géomatique :

par qui la carte ou les informations géographiques ont-elles été faites ? Dans quel contexte ? Pour répondre à quel objectif ? Selon quel processus et avec quels outils ? En un mot : d'où viennent les informations géographiques et/ou la carte ? – d'autant plus que répondre à cette question revient à se lancer dans une enquête de plus en plus complexe avec l'augmentation de la circulation des informations géographiques, notamment via le web (Gautreau et Noucher 2022). Cette proposition n'est pas forcément nouvelle ni particulièrement radicale : elle rejoint la place de plus en plus centrale que prennent les métadonnées dans les systèmes d'informations géographiques et dans le travail des géomaticiens². Celles-ci, bien qu'elles soient encore sous-exploitées, offrent un potentiel important d'intégration d'informations qualitatives de contextualisation des données géographiques (Schuurman 2009). Il pourrait s'agir d'en élargir les prérogatives : elles pourraient ne plus servir seulement à l'échange des informations géographiques, mais également en devenir davantage une dimension d'interrogation et de traitements. Il importe pour autant de reconnaître les ambiguïtés et zones d'ombres inhérente à ce chantier. De même qu'il existe un travail largement invisibilisé de la brutification des données (Denis et Goëta, 2017), un travail de maintenance réside toujours dans les modalités de la mise à disposition des données géographiques et de leur enrichissement par la métadonnée (Denis et Marquet, 2019). Cette épaisseur sociotechnique permet de saisir en quoi l'incomplétude et l'inégal avancement d'un tel projet sont non seulement inéluctables mais souhaitables³.

Se tourner vers les humanités numériques, et en particulier des modèles de standardisation mis en place par les sciences documentaires, pourrait aider à complexifier et enrichir les métadonnées et les informations géographiques grâce à leur ouverture⁴. En effet, l'interopérabilité et donc la connexion potentielle des informations ou métadonnées géographiques à d'autres bases données documentaires permettent un enrichissement considérable de celles-ci. C'est ainsi que fonctionnent les grands projets de gazetiers géo-historiques comme Pleiades. Celui-ci est connecté à un autre thésaurus largement diffusé pour les toponymes, GéoNames, au thésaurus PeriodO pour les périodes historiques, ou encore au Getty Art and Architecture Thesaurus (Elliott 2018) pour la description des objets ou bâtiments. L'intérêt premier est de faciliter l'export des données, mais il est également intéressant de penser que ces référentiels sont connectés par les liaisons du web des données à une infinité d'entrepôts de données qui les utilisent également comme références, offrant autant de potentiels d'enrichissement sémantique. L'ouverture des données des projets d'humanités numériques de cet envergure est

² Voir à cet égard la directive INSPIRE qui visent dès 2007 à standardiser les métadonnées au niveau européen

³ Abordant la question du point de vue de l'enquête en sciences sociales, Johanna Siméant-Germanos (2022) discute des injonctions contradictoires entre ouverture des données et respect de l'intégrité et de la parole des individus.

⁴ INSPIRE est d'ailleurs, dans cette perspective, développé en lien avec le modèle de description documentaire Dublin Core venu du monde de la documentation.

également une des conditions de leur succès : face à l'incommensurable tâche de la numérisation des corpus en sciences humaines, l'accessibilité et donc la circulation permises par l'OpenData, l'application des principes FAIR et le web de données garantissent l'enrichissement des informations et la constitution de grandes bases de données transversales (en termes d'institutions, d'équipes de recherche, de disciplines). Sans la construction de ces bases de données, impossible de développer les méthodes numériques dans les humanités. De quoi faire réfléchir les ambitions parfois globalisantes de certains systèmes d'informations géographiques et autre « jumeaux numériques ».

Les campagnes de numérisation et de géoréférencement des archives cartographiques menées par les archives et les bibliothèques de par le monde contribuent quant à elles à l'historicisation des informations géographiques, démarche critique en soi. Autrement dit, la numérisation et la géolocalisation de représentations cartographiques anciennes, dans des formats (surtout raster dans un premier temps) qui permettent de les superposer et donc de les croiser au reste des informations géographiques numériques, contribuent à mettre en perspective à la fois les territoires, mais également les représentations cartographiques elles-mêmes. On peut ainsi d'autant mieux se rendre compte – ou faire se rendre compte dans une perspective pédagogique – que l'absence d'un élément sur une carte peut être due à son absence sur le terrain mais aussi à des choix de cartographie ou des contraintes de collecte. Dans cette perspective, se multiplient en France les projets visant à la superposition de représentations cartographiques diachroniques d'un même territoire (par exemple le projet Alpage sur Paris ou l'Atlas Historique du Limousin sur Limoges) ou au géoréférencement et à la vectorisation d'un ou plusieurs fonds documentaires (par exemple de la carte de Cassini dans le projet GeoHistoricalData, des cartes anciennes de l'IGN dans le Géoportail, ou des fonds cartographiques de bibliothèques et laboratoires de recherche via le projet Imagéo).

Il est enfin intéressant de regarder comment les bibliothèques intègrent et traitent les données « géospatiales » associées aux informations documentaires. Quentin Mercuriol, géographe et conservateur de bibliothèque, identifie bien le gisement d'informations géographiques constitué par les bibliothèques au fil de la numérisation de leurs collections et les potentiels qu'il offre, en termes de recherche documentaire complexe par lieu pour les chercheurs,ses notamment géographes, en termes de connaissance et de gestion des collections documentaires pour les bibliothécaires, ou en termes d'enrichissement des métadonnées géonumériques pour les géomaticien.nes (Mercuriol 2021). En conséquence, de nombreuses bibliothèques, au premier rang desquelles la BNF en France avec le projet Gallicarte, expérimentent les potentiels d'interrogation et de visualisation de leur collection via la géovisualisation.

Ces initiatives illustrent les potentiels d'enrichissement et d'historicisation des informations géographiques via les gisements de données documentaires numériques constitués par les services d'archives, les bibliothèques et les équipes de recherche. Leur prise en compte en géomatique permettrait non seulement d'avoir une meilleure connaissance des contextes de production des cartes et informations, mais également de les intégrer, sous forme de données, aux systèmes d'information en

question. Un potentiel immense de traitements numériques incluant des informations de contexte et de production géomatique s'ouvrirait alors.

4. Intégrer et modéliser les représentations spatiales textuelles : leçons de linguistique, littérature et narratologie.

4.1. Intégration massive d'informations géographiques issues de sources textuelles

Un autre grand domaine des humanités numériques spatialisées concerne le traitement automatique des langues naturelles (TALN) et se concentre autour des méthodes de reconnaissance, d'extraction, de formalisation et de localisation automatique d'informations géographiques à partir de sources textuelles. Les questionnements méthodologiques auxquels répondent les chercheurs et chercheuses engagés dans ce domaine sont nombreux : comment identifier de manière non-ambiguë des lieux dont le nom change à travers le temps ? Comment gérer les homonymes ? Comment localiser des espaces et lieux désignés par des relations (« proche de », « à 5 km de », « un peu plus loin », etc.) (Moncla et al. 2016; McDonough, Moncla, et van de Camp 2019; Brando et al. 2021) ? Ces questionnements et les méthodologies qui en découlent peuvent s'appliquer à une diversité de corpus textuels : guides de voyage ou description de randonnées (Gaio et Moncla 2017), relations de voyage (Alrahabi et al. 2021), encyclopédie (Vigier et al. 2020), romans réalistes du XIX^e siècle (Joliveau, Gaio, et Moncla 2018). En parallèle, se développent en Humanités Numériques littéraires les entreprises de numérisation massives de corpus de textes divers via notamment la *Text Encoding Initiative* (TEI) – autant de corpus sur lesquels les méthodes de reconnaissance automatique des entités de lieux pourraient être appliquées. Toutes ces expérimentations promettent donc la possibilité d'introduire massivement des informations géographiques que l'on pourrait qualifier d'alternatives par rapport aux données utilisées classiquement en géomatique. Elles ouvrent la voie à la diversification des sources et donc des types et surtout des points de vue des informations géocodées.

4.2. Modéliser les représentations spatiales plus que l'espace en lui même

Les promesses des HN spatiales linguistiques ne s'arrêtent pas là : elles cherchent également à extraire des informations sémantiques associées à ces lieux, spécifiques aux sources travaillées. Perceptions, émotions, connotations, significations, en un mot, ce qu'on tente de numériser ce sont les représentations associées aux lieux (Cooper et Gregory 2011; Alrahabi et al. 2021; Dominguez et al. 2019). Parallèlement, des expérimentations sont faites plutôt dans le domaine des études littéraires, cinématographiques ou de la narratologie pour modéliser le rapport entre récit et espace. Le précurseur de ce domaine est sans conteste Franco Moretti qui a développé une méthode de *distant reading* permettant de dégager de grandes tendances géo-historiques en littérature, grâce à la cartographie des lieux

dans lesquels se déroulent de grandes quantités de romans (Moretti 2005). D'autres expérimentations adoptent une perspective plus rapprochée et s'intéressent à la manière dont des espaces sont mis en relation au fil récit, ou comment le fait de raconter une histoire en passant d'un lieu à l'autre au fil des pages définit une géographie particulière. On peut citer par exemple un travail mené sur la géographie des œuvres fictionnelles de l'écrivain algérien de langue française Kateb Yacine (Morel 2019). Le principe de cette étude est de mettre en relations dans une base de données d'une part les informations concernant la narration, en suivant les formalisations très compatibles avec le numérique proposées en narratologie notamment par Gérard Genette, et d'autres part les informations géographiques extraites des romans de l'auteur. L'objectif est ensuite de les visualiser de manière croisée afin de comprendre l'influence de la narration complexe de Kateb Yacine sur la construction de l'espace littéraire. L'Atlasiné développé à l'Université de Concordia à Montréal propose un mode de mesure et de cartographie de l'importance effective des lieux à l'écran : ceux-ci sont pondérés par le temps filmique qui s'y déroule et liés entre eux suivant le déroulement du film (Caquard et Fiset 2016). Dans les deux exemples, il s'agit d'utiliser des modèles et des visualisations spatio-temporels pour comprendre l'influence des manières de raconter (quantité de temps, d'images, de pages et enchainements) sur les représentations des lieux et espaces. Comme le prouve la réutilisation du dispositif de l'Atlasiné pour cartographier les récits de vie de rescapés du génocide rwandais (Caquard 2014), on peut tout à fait imaginer exporter ces expérimentations et les reproduire dans d'autres contextes. Cela permettrait d'intégrer plus généralement dans les modélisations la place que les espaces et lieux occupent dans les sources d'informations utilisées pour les construire ; et par là-même rappeler que notre connaissance du monde – y compris de l'espace – est intimement liée aux récits et représentations qui en sont faits.

5. Intégrer les visions alternatives, assumer un rapport sensible au monde : apports épistémique et tactique des humanités

Face à ces velléités de mise en relation des informations hétérogènes diverses pour traduire une vision complexe du monde, une tentation assez naturelle pourrait être celle de la réserve. Avec ce que ces grands projets peuvent supposer de nouveaux et nombreux utilisateurs, de redites dans les attentes et de surenchère dans l'économie des promesses, un tel déferlement peut en effet susciter un sentiment de mise à distance critique de la part d'une partie des praticien·ne·s SIG. Il importe pour autant de se focaliser sur les plus-values liées aux formes de transversalité et d'interdisciplinarité inhérentes à un grand nombre de projets d'humanités numériques.

Un exemple de cette utilité peut être trouvé dans les évolutions du courant dit du *deep mapping*. Définies par Bodenhamer et al. (2015) comme « des représentations multimédia finement détaillées d'un lieu et des personnes, des bâtiments, des objets, de la flore et de la faune qui y vivent et qui sont indissociables des activités de la vie

quotidienne », les « cartes profondes » ont fait l'objet d'efforts récents de réflexions, largement attisées par les possibilités techniques offertes par les outils numériques.

On peut compter Denis Wood, figure par ailleurs emblématique de la cartographie critique, comme une des précurseurs de ce *deep mapping*. Il raconte dans un compte-rendu rétrospectif (2015) le cheminement ayant pu faire évoluer une tentative progressive de cartographie sensible du quartier résidentiel de Boylan Heights à Raleigh vers ce qui allait devenir les éléments d'un atlas narratif remarqué. Wood explique le caractère lent et processuel du projet, basé sur une démarche pédagogique auprès d'étudiants éloignés des techniques et attendus habituels de la carte. Si la visée a été de formaliser et de défendre progressivement une forme de poésie de la cartographie, elle n'a trouvé une valorisation que de manière tardive et quasi-fortuite.

Cette approche « par la bande » a pu faire école : devenant « à la fois une plateforme, un processus et un produit » (Bodenhamer et al, 2015 : 3) les *deep maps* ont depuis été défendues comme une production légitime des humanités numériques spatialisées, selon un mode considérablement plus institutionnalisé et programmatique que dans leurs premières versions. Devant être « plus comprises comme une conversation que sous la forme d'un énoncé, [les *deep maps*] sont par nature instables, elles se déploient et changent continuellement en fonction de nouvelles données, de nouvelles perspectives et de nouvelles idées. » (Bodenhamer et al, 2015 : 4) S'il est possible d'être dubitatif envers de telles promesses, il faut reconnaître le grand confort associé à une telle configuration : fermement ancré dans les HN, le dispositif est présenté comme au service d'une narration, voire d'une argumentation, qui est également spatiale.

6. Conclusion

Il serait possible de continuer cet inventaire. On pourrait par exemple présenter l'intérêt des outils numériques utilisés en sociologie pour étudier les réseaux de circulation des données spatialisées ou les réseaux professionnels des cartographes et géomaticien.nes afin de retracer les influences sur leurs choix techniques. Aussi, les potentiels dégagés des projets cités ne sont pas figés : d'autres pourront y trouver des inspirations différentes ; et de nombreux autres projets pourraient être cités... L'objectif n'était pas un inventaire exhaustif, mais bien de montrer que de nombreux travaux en cours dans les humanités numériques pouvaient converger et se combiner vers la complexification des approches et des modélisations géomatiques, afin de répondre aux enjeux de la géomatique critique : diversifier les points de vue, les types de données et les sujets ; rendre davantage manifeste d'où viennent et comment sont produites les connaissances et en particulier le sinueux chemin des informations géographiques ; être capable d'intégrer pleinement et de traiter l'hétérogénéité et la complexité conséquente.

Bibliographie

- Alrahabi, Motasem, Carmen Brando, Muhamed AlKhalil, et Joseph Dichy. 2021. « Paris dans les récits de voyage d'écrivains arabes : repérage, analyse sémantique et cartographie de toponymes ». *Humanités numériques*, n° 3 (mai). <https://doi.org/10.4000/revuehn.1079>.
- Bagnall, Roger, Richard J. A. Talbert, Sarah Bond, Jeffrey Becker, Tom Elliott, Sean Gillies, Ryan Horne, et al. 2006. « Pleiades: A Community-Built Gazetteer and Graph of Ancient Places ». Collection. Institute for the Study of the Ancient World, New York University. 2006. <http://pleiades.stoa.org>.
- Bodenhamer, David J, John Corrigan, et Trevor M Harris. 2010. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=642395>.
- Bodenhamer, David J., Corrigan, John, & Trevor M. Harris. 2015. *Deep maps and spatial narratives*. Bloomington & Indianapolis: Indiana University Press.
- Brando, Carmen, Francesca Frontini, Dominic Moreau, Mathieu Roche, et Éric Masson. 2021. « Introduction. Humanités numériques et analyses spatiales : enjeux et perspectives ». *Humanités numériques*, n° 3 (mai). <https://doi.org/10.4000/revuehn.2038>.
- Bunge, William W. 1971. *Fitzgerald; Geography of a Revolution*. Cambridge, Massachusetts: Schenkman.
- Caquard, Sébastien. 2014. « Atlas des récits de vie rwandais ». Atlas des récits de vie rwandais. 2022 2014. <https://rs-atlascine.concordia.ca/rwanda/index.html>.
- Cooper, David, et Ian N Gregory. 2011. « Mapping the English Lake District: a literary GIS ». *Transactions of the Institute of British Geographers* 36 (1): 89-108.
- Cope, Meghan, et Sarah Elwood. 2009. *Qualitative GIS: A Mixed Methods Approach*. London. <https://doi.org/10.4135/9780857024541>.
- Deflou-leca, Noelle, et Anne Massoni. s. d. « Collégiales et Monastères – de La France Médiévale ». Consulté le 3 février 2023. <https://colemmon.huma-num.fr/>.
- Denis, Jérôme & Samuel Goëta. 2017. Rawification and the careful generation of open government data. *Social Studies of Science*, 47(5), 604–629. <https://doi.org/10.1177/0306312717712473>
- Denis, Jérôme & Clément Marquet. 2019. De la production à la maintenance : faire exister collectivement des données géographiques ouvertes. *Questions de communication*, 36, 63-83. <https://doi.org/10.4000/questionsdecommunication.20914>
- Desbois, Henri. 2015. « La carte et le territoire à l'ère numérique ». *Socio. La nouvelle revue des sciences sociales*, n° 4 (avril): 39-60. <https://doi.org/10.4000/socio.1262>.

Dominguès, Catherine, Laurence Jolivet, Carmen Brando, et Marion Cargill. 2019. « Place and Sentiment-Based Life Story Analysis ». *Revue Française Des Sciences de l'information et de La Communication*, n° 17 (août). <https://doi.org/10.4000/rfsic.7228>.

Elliott Tom. 2018. « The Pleiades gazetteer of ancient places: inventing an “Un-GIS” and its uses ». In . Paris. <https://www.youtube.com/watch?v=MPv20lfPRuk>.

Gaio, Mauro, et Ludovic Moncla. 2017. « Extended Named Entity Recognition Using Finite-State Transducers: An Application To Place Names ». In *The Ninth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2017)*. Nice, France. <https://hal.archives-ouvertes.fr/hal-01492994>.

Gautreau, Pierre. 2021. *La Pachamama en bases de données : Géographie politique de l'information environnementale contemporaine. La Pachamama en bases de données : Géographie politique de l'information environnementale contemporaine*. Travaux et mémoires. Paris: Éditions de l'IHEAL. <http://books.openedition.org/iheal/9362>.

Gautreau, Pierre, et Matthieu Noucher. 2022. « L'adieu aux cartes. Refaire de la cartographie politique aux temps du numérique ». In *Politiques de la carte*, ISTE Editions, 19.

Goodchild, Michael F. 2015. « Two Decades on: Critical GIScience since 1993: Critical GIScience since 1993 ». *The Canadian Geographer / Le Géographe Canadien* 59 (1): 3-11. <https://doi.org/10.1111/cag.12117>.

Hirt, Irène. 2009. « Cartographies autochtones. Éléments pour une analyse critique ». *L'Espace géographique* 38 (2): 171-86. <https://doi.org/10.3917/eg.382.0171>.

Joliveau, Thierry. 2004. « Géomatique et gestion environnementale du territoire. Recherches sur un usage géographique des SIG ». <http://thierry.joliveau.pagesperso-orange.fr/HDR.html>.

Joliveau, Thierry, Mauro Gaio, et Ludovic Moncla. 2018. « Cartographier les ononymes de Paris cités dans les romans du XIXème siècle ». In *Humanités Numériques Spatialisées (HumaNS)*. Montpellier, France. <https://agritrop.cirad.fr/589755/>.

Kelly, Meghan, et Amber Bosse. 2022. « Pressing Pause, “Doing” Feminist Mapping ». *ACME: An International Journal for Critical Geographies* 21 (4): 399-415.

Kwan, Mei-Po. 2002. « Feminist Visualization: Re-Envisioning GIS as a Method in Feminist Geographic Research ». *Annals of the Association of American Geographers* 92 (4): 645-61. <https://doi.org/10.1111/1467-8306.00309>.

Mahmoudi, Dillon, et Taylor Shelton. 2022. « Doing Critical GIS ». *ACME: An International Journal for Critical Geographies* 21 (4): 327-36.

McDonough, Katherine, Ludovic Moncla, et Matje van de Camp. 2019. « Named entity recognition goes to old regime France: geographic text analysis for early modern French corpora ». *International Journal of Geographical Information Science* 33 (12): 2498-2522. <https://doi.org/10.1080/13658816.2019.1620235>.

Mercurol, Quentin. 2021. « Après la carte ? De nouvelles rencontres entre l'information géographique et les bibliothèques - Notice bibliographique | Enssib ». Mémoire d'étude de Diplôme de Conservateur des Bibliothèques (DCB), Lyon: Enssib. <https://www.enssib.fr/bibliotheque-numerique/notices/70139-apres-la-carte-de-nouvelles-rencontres-entre-l-information-geographique-et-les-bibliotheques>.

Moncla, Ludovic, Mauro Gaio, Javier Noguera-Iso, et Sébastien Mustière. 2016. « Reconstruction of itineraries from annotated text with an informed spanning tree algorithm ». *International Journal of Geographical Information Science* 30 (6): 1137-60. <https://doi.org/10.1080/13658816.2015.1108422>.

Morel, Juliette. 2019. « Modélisation spatiale du récit littéraire complexe de Kateb Yacine ». *Mappemonde*, n° 125 (janvier). <https://doi.org/10.4000/mappemonde.766>.

Morel, Juliette, Rémi Crouzevialle, et Anne Massoni. 2020. « Construction and Management of a Geo-Historical Information System for an Interdisciplinary and Contributory Atlas: The Historical Atlas of Limousin ». *International Journal of Humanities and Arts Computing* 14 (1-2): 27-45. <https://doi.org/10.3366/ijhac.2020.0243>.

Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.

Mouza, Cédric du, et Stéphane Lamasse. 2022. « Digital Academic History and the Challenge of Uncertainty: XIth Heloise Workshop ». In *Heloise*. Paris. <https://heloise.hypotheses.org/685>.

Murrieta-Flores, Patricia, et Bruno Martins. 2019. « The geospatial humanities: past, present and future ». *International Journal of Geographical Information Science* 33 (12): 2424-29. <https://doi.org/10.1080/13658816.2019.1645336>.

Noucher, Mathieu. 2017. *Les Petites Cartes du web. Analyse critique des nouvelles fabriques cartographiques*. Actes de la recherche à l'Ens 18. Paris: Éditions Rue d'Ulm.

Pavlovskaya, Marianna, et Kevin St. Martin. 2007. « Feminism and Geographic Information Systems: From a Missing Object to a Mapping Subject: Feminism and Geographic Information Systems ». *Geography Compass* 1 (3): 583-606. <https://doi.org/10.1111/j.1749-8198.2007.00028.x>.

Peuquet, Donna. 1994. « It's about Time: A Conceptual Framework for the Representation of Temporal Dynamics. Geographic Information Systems ». *Annals of the Association of the American Geographers* 3 (84): 441-61.

Rodier, Xavier, et Laure Saligny. 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 Geography*, juin. <https://doi.org/10.4000/cybergeo.23175>.

Schuurman, Nadine. 2000. « Trouble in the Heartland: GIS and Its Critics in the 1990s ». *Progress in Human Geography* 24 (4): 569-90. <https://doi.org/10.1191/030913200100189111>.

———. 2009. « Qualitative GIS: A Mixed Methods Approach ». In *Qualitative GIS: A Mixed Methods Approach*, 41-56. London: SAGE Publications Ltd. <https://doi.org/10.4135/9780857024541>.

Siméant-Germanos, Johanna. 2022. Qui protéger, consentir à quoi, enquêter comment ? Les sciences sociales face à la bureaucratisation de la vertu scientifique. *Genèses*, 129(4), 66-87. doi:10.3917/gen.129.0066

Vigier, Denis, Ludovic Moncla, Alice Brenon, Katherine McDonough, et Thierry Joliveau. 2020. « Named entity classification in the Encyclopédie ou dictionnaire raisonné des sciences et des métiers par une société de gens de lettres (1751-1772) ». In . <https://doi.org/10.1051/shsconf/20207811008>.

Wilson, Matthew W. 2015. « New Lines? Enacting a Social History of GIS ». *The Canadian Geographer / Le Géographe Canadien* 59 (1): 29-34. <https://doi.org/10.1111/cag.12118>.

———. 2017. *New Lines: Critical GIS and the Trouble of the Map*. 1st edition. Minneapolis: Univ Of Minnesota Press.

Wood, Denis. (2015). Mapping Deeply. *Humanities*, 4(3), 304-318. <https://doi.org/10.3390/h4030304>