



HAL
open science

Classification des tissus urbains à partir de bases de données vectorielles - application à Strasbourg

Anne Puissant, N. Lachiche, A. Braud, Grzegorz Skupinski, Julien Perret

► **To cite this version:**

Anne Puissant, N. Lachiche, A. Braud, Grzegorz Skupinski, Julien Perret. Classification des tissus urbains à partir de bases de données vectorielles - application à Strasbourg. Conférence Internationale de Géomatique et Analyse Spatiale SAGEO'10, Nov 2010, Toulouse, France. pp.198-211. halshs-00627704

HAL Id: halshs-00627704

<https://shs.hal.science/halshs-00627704>

Submitted on 16 Jun 2021

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.

Classification des tissus urbains à partir de données vectorielles – application à Strasbourg

Anne Puissant*, Nicolas Lachiche**, Agnès Braud**, Grégoire Skupinski*, Julien Perret***

* *Laboratoire Image, Ville, Environnement, ERL 7230 CNRS / Université de Strasbourg - 3, rue de l'argonne, F-67083 Strasbourg cedex, France.*
{anne.puissant; g.skupinski@live-cnrs.unistra.fr}

** *Laboratoire des Sciences de l'Images, de l'Informatique et de la Télédétection, UMR 7005 CNRS /Université de Strasbourg - Pôle API Bd Sébastien Brant 67412 Illkirch Cedex, France.*
{nicolas.lachiche; agne.braud@unistra}

*** *Laboratoire COGIT, Institut Géographique National 73 avenue de Paris, F-94165 Saint Mandé Cedex, France*
Julien.perret@ign.fr

RÉSUMÉ. L'analyse morphologique des tissus urbains est un préalable important afin de mieux comprendre les dynamiques urbaines et afin de les simuler. Dans ce contexte, l'objectif de cet article est de tester un algorithme d'apprentissage supervisé symbolique (TILDE) permettant d'obtenir un modèle de classification représenté sous forme de règles de décision. L'objectif est d'automatiser le processus d'identification des tissus urbains à partir de bases de données vectorielles et d'extraire des connaissances morphologiques sur les tissus urbains. Différents tests, appliqués à l'agglomération de Strasbourg, sur des données historiques reconstruites sur une période de 50 ans, met en évidence l'intérêt de ce type d'algorithme pour classer les tissus urbains. Toutefois, l'identification automatique de règles caractérisant la morphologie des tissus urbains de manière univoque n'est pas satisfaisante.

ABSTRACT. Morphological analysis of urban fabric is a relevant step in order to perform the analysis of urban dynamics and to simulate their evolutions. In this context, the objective of this paper is to test a classification method based on a supervised and symbolic algorithm (tilde) to label urban blocks and to produce rules on their morphology. The objective is to automate the identification of urban fabrics from vector and historical database and to extract knowledge on urban morphology. Several tests applied on the Strasbourg area on historical database, are encouraging but rules are not relevant to extract urban blocks.

MOTS-CLÉS : tissu urbain, morphologie, bases de données vectorielles, classification, arbre de décision.

KEYWORDS: urban fabric, morphology, vector database, classification, decision tree.

1. Introduction

La ville est aujourd'hui au cœur des grandes préoccupations contemporaines. L'étalement urbain et la maîtrise de ses conséquences restent un défi important de l'aménagement du territoire. L'analyse morphologique des tissus urbains est un préalable important afin de mieux comprendre les processus d'urbanisation de nos agglomérations. A partir de cette connaissance, on peut imaginer de proposer des scénarios de développement urbain ou de développer des concepts différents pour trouver des réponses réalistes au défi de l'étalement urbain.

Le projet de recherche GeOpenSim (ANR Blanc 2007-2010) a pour objectif de proposer une plate-forme Open Source pour l'analyse et la simulation des dynamiques urbaines à l'aide de bases de données topographiques historiques. Ces dernières sont construites à l'aide de bases de données topographiques récentes ainsi que de photographies aériennes et/ou de cartes historiques (Puissant *et al.*, 2010). Ces bases de données vectorielles sont analysées afin d'améliorer les connaissances sur les objets spécifiques comme les bâtiments et afin d'identifier des règles d'évolution à différents niveaux méso-géographiques (e.g. la ville ou l'îlot) (Curie *et al.* 2010).

L'extraction de connaissances au niveau de l'îlot est une étape préalable à l'analyse de leur évolution. Celle-ci passe par une classification et une analyse morphologique des tissus urbains qui peut être assimilée à un problème de fouille de données. Dans ce contexte, l'objectif de cet article est de tester un algorithme d'apprentissage supervisé symbolique (Tilde) permettant d'obtenir un modèle de classification représenté sous forme de règles de décision. L'objectif est d'automatiser le processus d'identification des tissus urbains à l'échelle d'une agglomération de la taille de Strasbourg et d'extraire des connaissances morphologiques sur les tissus urbains permettant de suivre leur évolution dans le temps.

Après un état de l'art (section 2) sur les méthodes de classification des tissus urbains et sur les méthodes d'extraction de connaissances à partir de telles données, la zone d'étude ainsi que les mesures et les paramètres statistiques utilisés sont détaillés à la section 3. Les différentes étapes de la classification sont ensuite présentées à la section 4. Dans la section 5, les tests effectués ainsi que les résultats sont détaillés et discutés. Les conclusions et perspectives sont finalement exposées à la section 6.

2. Classification des tissus urbains et extraction de connaissances : état de l'art

A l'heure actuelle, la classification automatique ou semi-automatique des tissus urbains est rarement effectuée à partir de données vectorielles. Classiquement, ce type de classification est réalisé à partir d'images satellites et/ou de photographies aériennes et passe par une phase de photo-interprétation manuelle combinée à des vérifications terrain. Le résultat se présente sous la forme d'une base de données sur

l'occupation et l'utilisation des sols. La BD-OCS (base de données 'occupation du sol') proposée par la région Alsace peut être citée comme exemple (Geoméditerranée, 2004). Celle-ci est réalisée par photo-interprétation à partir d'images satellites, proposant une cartographie des tissus urbains à l'échelle du 1/10 000^e.

L'automatisation de ce type de cartographie en milieu urbain, fait l'objet de nombreux travaux de recherche en particulier à partir d'images satellites haute résolution (Donnay *et al.*, 2001) mais également depuis l'avènement des images à très haute résolution (Blaschke, 2009). Des méthodes de traitements d'image fondées sur un apprentissage supervisé (classification supervisée) sont largement utilisées dans ce domaine (Bauer *et al.*, 2001). Ces techniques basées sur les valeurs radiométriques des pixels des images, permettent d'identifier des zones qui combinées avec un découpage vectoriel (IRIS de l'INSEE par exemple) permettent d'obtenir une cartographie des îlots urbains (Holland *et al.*, 2006). Toutefois, les images satellites ne permettent pas d'analyser la composition et l'organisation spatiale des tissus urbains.

Avec l'arrivée des nouvelles images à très haute résolution (résolution spatiale métrique) et le développement de nouvelles méthodes de traitement d'images orientées-objets (Blaschke *et al.*, 2004), c'est-à-dire permettant de travailler directement sur des régions homogènes (agrégats de pixels homogènes), de nombreux travaux utilisent une segmentation multi-échelle pour extraire et classer des objets micros (bâtiment, végétation, route) et des objets méso (tissus urbains) (Jacquin *et al.*, 2010, Bouziani *et al.*, 2010). D'autres méthodes tentent d'utiliser simultanément les informations issues des images à différentes résolutions spatiales afin de reconstruire les objets méso (Kurtz *et al.* 2010). Malgré des résultats prometteurs, ces méthodes sont toutefois encore exploratoires et pas directement opérationnelles. Elles ne permettent pas non plus d'analyser et de caractériser la composition morphologique des tissus urbains.

L'analyse morphologique d'un tissu urbain peut être assimilée à un problème de fouille de données. En effet, la fouille de données spatiales est née du besoin d'exploitation dans un but décisionnel de données à caractère spatial produites ou importées, susceptibles de délivrer des informations ou des connaissances par le moyen d'outils exploratoires (Zeitouni *et al.*, 2000). Dans ce domaine, des méthodes classiques comme les arbres de décision ont été adaptées. Les arbres de décision ont pour but de trouver les attributs explicatifs et les critères précis sur les attributs donnant le meilleur classement vis-à-vis d'un attribut à expliquer. Dans Chelghoum *et al.* (2002), les auteurs pré-calculent les relations spatiales entre objets spatiaux et les stockent. Les critères utilisés dans l'arbre de décision construit sont alors des tests d'existence d'un objet vérifiant une relation spatiale donnée. Cet algorithme a été appliqué à l'analyse du risque d'accidents routiers. A la suite de ce travail, les auteurs se sont intéressés à un système de construction d'arbre de décision à partir de données relationnelles appelé TILDE (Blockeel and De Raedt, 1998), et l'ont étendu aux données spatiales avec S-TILDE qui intègre en plus au niveau du critère de

division les propriétés d'objets voisins (Chelghoum *et al.*, 2006). D'autres travaux ont porté sur la fouille de données à partir de données vecteurs. Il existe notamment un ensemble de travaux sur les règles d'association, comme (Salleb et Vrain, 2000) qui introduisent dans les règles d'association des relations spatiales entre les objets qui peuvent provenir de différentes tables. Ce travail étend celui de (Koperski et Han, 1995). Dans ce contexte, nous avons testé l'algorithme relationnel TILDE qui a une expressivité supérieure à l'algorithme attribut-valeur C4.5 utilisé couramment (Quinlan, 1995). Il s'applique également à nos données de départ (cf. section 3) contenues dans une base de données relationnelles, sous la forme de deux tables (une table pour les îlots et une table pour les bâtiments).

3. Zone d'étude, données et mesures morphologiques

La Communauté Urbaine de Strasbourg (CUS) est une agglomération qui s'est développée, comme la plupart des grandes villes françaises, selon le schéma de l'étalement urbain (Bessy-Piétry, 2000). Depuis sa création (en 1967) jusqu'aux années 1982, la surface bâtie a augmenté de 71.6% (ADEUS, 2003), contre 20.9% entre l'année 1988 et 2000. Depuis on assiste à une forte concentration de l'emploi tandis que la population se diffuse largement dans l'ensemble du département (Bas-Rhin), voire dans l'espace transfrontalier (Saib, 2009).

Dans ce contexte, les objets géographiques de la base de données topographique de l'IGN sur la CUS (IGN, 2002) sont utilisés comme données de référence. Les bases de données historiques sont ensuite construites sur quatre zones tests (Figure 1). Elles ont été sélectionnées en raison de leur localisation dans l'agglomération (zones périurbaines) et des processus d'évolution représentatifs des grandes agglomérations françaises au cours des 50 dernières années. Ces données historiques ont été produites, dans le cadre du projet, pour 5 dates distinctes : 1956, 1966, 1976, 1989, 2002 (référence) et 2008.

La typologie des tissus urbains proposée (Tableau 1) est adaptée afin de cartographier l'espace urbain d'une grande agglomération française à une échelle du 1/10 000^e. Ces classes, compatibles avec les nomenclatures existantes (e.g. Corine Land Cover, BDOCS (Alsace), ne tiennent pas compte de la fonctionnalité des tissus urbains car cette information est impossible à reconstruire sur des données historiques. Elles permettent de se concentrer sur leur analyse morphologique. Parmi ces classes, trois d'entre elles ne sont pas prise en compte dans la suite de l'analyse en raison de leur faible potentiel d'évolution (classe n°1 – tableau 1) et d'interprétation (classe n°8 – tableau 1).

L'analyse morphologique des tissus urbains passe par une étape préalable de construction des îlots urbains. Ceux-ci sont construits automatiquement à partir des objets géographiques linéaires de la base de donnée topographique (objets de type 'réseaux de communication' regroupant le réseau routier, ferré et hydrographique).

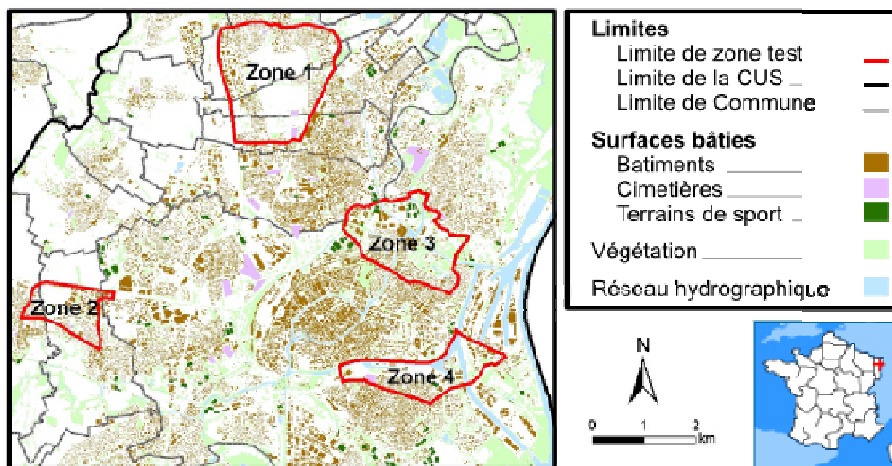


Figure 1. Localisation des quatre zones tests utilisées pour analyser l'évolution des tissus urbains (Source: BDTPOPO@2002).

Classes de tissu urbain
1. Habitat continu dense de type centre ville ou centre bourg
2. Habitat discontinu de type pavillonnaire ou individuel
3. Habitat discontinu de type collectif ou grand ensemble
4. Habitat discontinu mixte : <i>Mélange des classes 2 (pavillonnaire) et 3 (grand ensemble)</i>
5. Tissu urbain discontinu mixte <i>Mélanger de classes d'habitat (classe 1, 2, 3) et de la classe 6 (emprise spécialisées bâties)</i>
6. Emprises spécialisées bâties : <i>emprises militaires, zones industrielles, commerciales, emprises scolaires et universitaires, emprises hospitalières, emprise publique ou patrimoine</i>
7. Emprises spécialisées peu ou pas bâties : <i>emprises liées à l'activité d'extraction – carrières, surfaces associées aux réseaux : gare de triage, échangeur, emprises aériennes, ferroviaires et portuaires, équipements sportifs et de loisirs, espaces en mutation ou libres de construction (sols nus, artificiels ou végétalisés).</i>
8. Réseau de communication
9. Réseau hydrographique

Tableau 1. Liste des classes d'îlots urbains retenus

Des mesures (ou paramètres) sont ensuite calculées pour caractériser d'une part les objets géographiques surfaciques (e.g. l'aire, l'élongation et la convexité) et d'autre part l'organisation spatiale de ces objets dans les îlots (l'orientation des bâtiments et l'orientation relative des bâtiments par rapport à la route la plus proche

par exemple). D'autres mesures, telles que le nombre de bâtiments dans l'îlot, la surface construite (la somme des surfaces des bâtiments) et la densité (surface construite/surface de l'îlot) sont disponibles. Pour chacune d'entre elles, les valeurs du minimum, du maximum, de la moyenne, de la médiane et de l'écart type sont également calculées.

4. Etapes de classification et modules proposés

L'objectif de la classification est d'associer une des classes prédéfinies (Tableau 1) aux îlots urbains construits automatiquement à partir du réseau de communication. Cette étape de classification est une étape préliminaire nécessaire à l'analyse des évolutions proprement dite. De la qualité des résultats vont dépendre, la précision de l'analyse des évolutions. Cette classification effectuée par apprentissage supervisé se décompose en 3 étapes. Elle débute par la création d'une base d'exemples d'îlots urbains et par la construction d'un modèle de classification. Ce modèle est ensuite appliqué à l'ensemble de la base de données îlots. Le résultat est ensuite validé de manière quantitative et qualitative. La figure 2 schématise les différentes étapes ainsi que les 'modules' logiciels utilisés.

La base d'îlots urbains (destinés à l'apprentissage et à la validation) est créée à partir d'un module d'étiquetage open source fondé sur Géoxygène (Lesbegueries *et al.* 2009). Ce module permet à l'expert d'étiqueter manuellement des îlots urbains représentatifs d'une classe thématique (Tableau 1). L'analyse réalisée par l'expert est fondée sur les seuls critères morphologiques issus des bases de données topographiques et ne considère que les formes des objets topographiques.

Lors de la classification, comme les données sont contenues dans une base de données relationnelle sous la forme de deux tables (une table pour les îlots et une table pour les bâtiments), l'algorithme TILDE permet de construire un arbre de décision impliquant :

- des conditions sur l'îlot, par exemple sa superficie,
- des fonctions d'agrégation sur les propriétés des bâtiments contenus dans l'îlot, par exemple la superficie moyenne des bâtiments,
- des contraintes sur le nombre de bâtiments ayant telles propriétés, par exemple qu'il y ait au moins 5 bâtiments de superficie inférieure à 150 m².

Le modèle obtenu est ensuite appliqué à l'ensemble des îlots non-étiquetés afin de prédire leurs classes. Les résultats de la classification sont stockés dans la table attributaire et peuvent être visualisés dans un SIG.

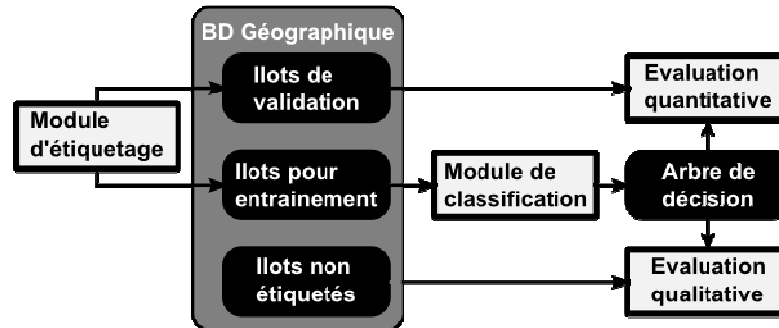


Figure 2. Etapes de la méthode de classification

Les résultats de la classification sont ensuite validés d'une part, qualitativement en analysant l'arbre de décision produit par TILDE et d'autre part quantitativement par validation croisée. Lorsque l'on dispose d'un seul jeu d'îlots étiquetés (apprentissage et entraînements confondus) (cf. test 1 – section 5), l'évaluation est faite par une validation croisée en 10 'paquets'. Cette procédure, standard en fouille de données, consiste à diviser le jeu de données en 10 sous-ensembles de données de taille semblable, en prenant soin que chaque groupe contienne un nombre d'îlot de chaque classe proportionnel au nombre présents dans le jeu complet (Witten et Frank, 2005). Le nombre de 'paquets', en général 5 ou 10, est un compromis entre le nombre d'exemples pour apprendre et le nombre de modèles à apprendre. Chaque paquet sert de jeu de validation, à son tour, et un modèle est construit à partir des 9 autres paquets. On construit donc 10 modèles, à partir d'un sous-ensemble des données, et on les évalue sur des données distinctes de celles ayant servi à construire chaque modèle. La moyenne des performances de ces modèles est un estimateur non-biaisé de la performance du modèle construit sur le jeu de données complet. Lorsque plusieurs jeux de données sont disponibles, par exemple les îlots des zones 1 à différentes dates (cf. tests 2 et 3 – section 5), un arbre de décision est construit à partir d'un jeu de données, par exemple les îlots de la zone 1 en 2002, puis ce modèle est évalué sur tous les autres jeux de données.

5. Tests de classification et analyse des résultats

La BDTopo@2002 est utilisée comme base de données de référence pour l'étape de création des îlots étiquetés (apprentissage et validation). Le premier objectif est d'identifier le type d'échantillonnage le plus approprié pour classer les îlots urbains sur l'ensemble de l'agglomération (section 5.1). Le second objectif est d'appliquer le modèle appris le plus pertinent, c'est-à-dire avec le taux de classement le plus élevé, à la base de données historique (section 5.2). Les résultats sont évalués quantitativement en calculant le taux de bon classement et qualitativement par l'interprétation de l'arbre de décision.

5.1. Résultats de la classification sur l'ensemble de l'agglomération

L'objectif est ici d'évaluer l'impact de la construction de la base d'exemples sur les résultats de classification. Deux tests de classifications sont appliqués à l'ensemble des îlots urbains de l'agglomération. La pertinence des règles produites sera évaluée uniquement sur le test fournissant le meilleur résultat de classement.

Le premier test de classification est effectué en créant une seule base d'exemple d'îlots urbains (Tableau 2). Ces îlots (apprentissage et validation confondus) sont sélectionnés par l'utilisateur comme morphologiquement représentatifs d'une classe d'îlots. Ils sont répartis sur l'ensemble de l'agglomération et représentent 20% du nombre d'îlots total. Le nombre d'exemples pour les six classes thématique comporte, pour ce test, au minimum 30 exemples par classes afin d'assurer la significativité statistique des mesures et des paramètres (Tableau 2). Pour les classes les plus représentées dans l'agglomération (habitat et tissu mixte), le nombre d'exemples est plus important.

Le second test de classification est effectué en utilisant une base d'exemples plus restreinte (10% du nombre total d'îlots) correspondant soit à la zone d'intérêt 1 (zone périurbaine - test 2a), soit à la zone d'intérêt 3 (zone en péricentre – test 2b). Le test 2c est effectué en tenant compte des exemples issus de ces deux zones d'intérêt (Tableau 3). L'objectif est également de vérifier si le nombre d'exemples influence le résultat de la classification.

L'analyse des résultats du taux de bon classement (Tableau 2) montre, comme on pouvait s'y attendre, que le nombre d'exemples influence le taux de bon classement puisque les tests 2a et 2b effectués à partir d'échantillons restreints sont plus faibles que les deux autres tests présentant un nombre d'échantillons plus important. Les tests effectués avec un plus grand échantillon obtiennent plus de 75% de bon classement. Ce résultat met en évidence que le choix et le nombre des échantillons doit être significativement représentatif des classes thématiques à identifier.

2002	Test 1	Test 2		
		Test 2a	Test 2b	Test 2c
Classes d'îlots urbains	Agglomération	Zone 1	Zone 3	Zone 1 et 3
Habitat pavillonnaire	40	56	22	78
Habitat collectif	39	21	21	42
Habitat discontinu mixte	71	15	16	31
Tissu discontinu mixte	60	14	7	21
Emprises spécialisées bâties	30	4	23	27
Emprises spécialisées peu ou pas bâties	34	11	25	36
<i>Nb total d'îlots*</i>	<i>5940</i>	<i>133</i>	<i>201</i>	<i>354</i>
<i>Taux global de bon classement</i>	<i>78%</i>	<i>71%</i>	<i>63%</i>	<i>77%</i>

Tableau 2. Récapitulatif du nombre d'îlots étiquetés (apprentissage et validation confondus) selon les tests effectués et les zones d'intérêts

* y compris les îlots correspondant à des réseaux de communication

L'analyse de la distribution des îlots urbains classés (Figure 3) combinée à l'interprétation visuelle (Figure 4), permettent toutefois de nuancer ce résultat. Sur l'ensemble de l'agglomération, le nombre d'îlots de type individuel et collectif est plus important à partir du test 2c. L'interprétation visuelle confirme ce résultat, plus proche. En effet, sur la figure 5, la majorité des îlots urbains de type individuel et collectif sont bien classés (par exemple au nord de la zone 1 – Figure 5). Le test 1 a plutôt tendance à surestimer les classes d'habitat et de tissu mixte. La surreprésentation de la classe d'emprise spécialisée peu ou pas bâtie était prévisible puisque aucun échantillon de cette classe n'a été pris en compte lors de la phase d'échantillonnage. Elle pourrait être corrigée en éliminant les îlots de cette classe selon un critère de forme combinée à un indice d'élongation, par exemple.

L'analyse des résultats de l'arbre de décision (obtenus à partir du modèle issu du test 2c) permet d'extraire 8 règles de caractérisation dont voici deux exemples :

- Règle 1 : Si densité îlot < 0,05 alors îlot urbain de type emprise spécialisée peu bâtie ;
- Règle 2 : Si densité îlot > 0,05 et nombre de bâtiments (dont la surface est inférieure à 105m²) est inférieur ou égal à 3 et que le nombre de bâtiments (dont la surface est inférieure à 185m²) est inférieur ou égal à 1 et la surface moyenne des bâtiments est inférieure à 1205 m², alors îlot urbain de type habitat collectif.

Deux règles pour les emprises spécialisées bâties ainsi que pour l'habitat mixte sont extraites. L'analyse détaillée de celles-ci si elle permet de repérer certains seuils (en terme de densité par exemple) mais n'a pas permis d'extraire une connaissance interprétable pour le géographe et susceptible d'être utilisée dans un modèle de simulation. D'autres tests sont en cours afin de vérifier si d'autres algorithmes tenant compte notamment du voisinage ou des relations spatiales permettraient d'extraire une connaissance réellement exploitable.

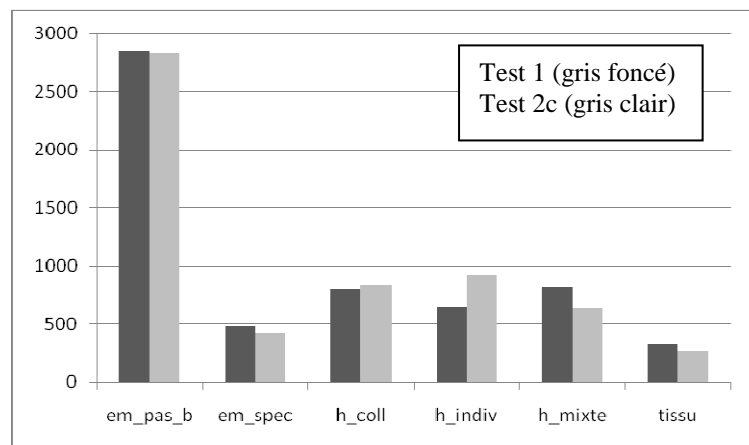


Figure 3. Répartition des classes thématiques sur l'ensemble de l'agglomération (5940 îlots urbains) selon le test 1 (gris foncé) et selon le test 2c (gris clair)

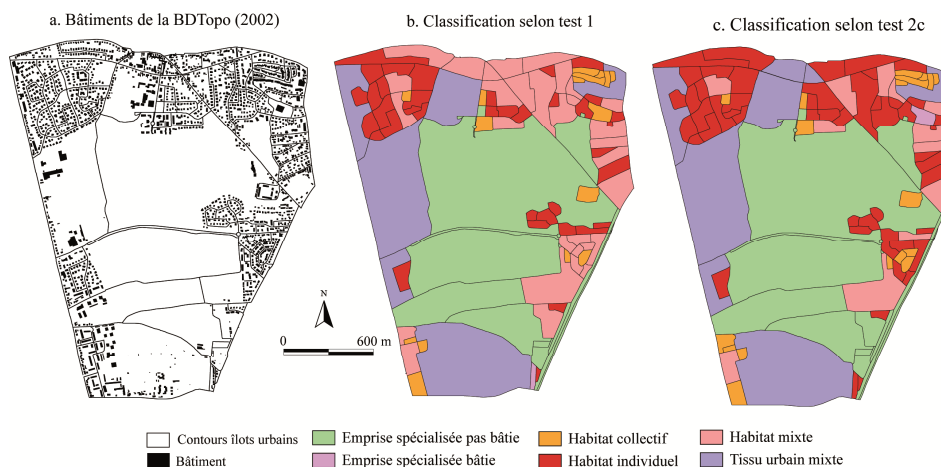


Figure 4. Comparaison des résultats de la classification issus des tests 1 et 2c (appliqué à l'ensemble de l'agglomération).

Le modèle construit à partir du test 2c le plus satisfaisant en terme quantitatif et qualitatif, a ensuite été appliqué à l'ensemble des données historiques sur les 4 zones d'intérêt (section 5.2). Les règles produites par l'arbre de décision sont également analysées.

5.2. Résultats de la classification sur la base de données historiques

Le modèle construit le plus satisfaisant (test 2c) est utilisé pour classer les îlots des bases de données historiques construites sur les zones 1 à 4 entre 1956 et 2008. Les résultats du taux de bon classement sont résumés au tableau 3. Les résultats de 2002 ne sont pas mentionnés pour les zones 1 et 3, puisque ces îlots ont contribué à construire le modèle appliqués aux autres dates.

Test 2c	Zone 1	Zone 2	Zone 3	Zone 4
2008	89	65	79	65
2002	/	63	/	58
1989	83	62	79	60
1976	84	71	76	56
1966	NA	NA	80	57
1956	79	93	83	53

NA – données non-acquises à cette date en raison de l'absence de photographies aériennes

Tableau 3. Taux de bon classement (%) par zone, selon les dates disponibles.

L'analyse ce tableau montre que le taux de bon classement aux dates antérieures à 2002 (et en 2008) est d'une manière générale proche de 80%. Ceci signifie que la composition d'un îlot urbain, par exemple, de type habitat individuel a peu varié dans le temps. Les différentes classes d'îlots définies initialement sont donc pertinentes pour l'analyse future de leur évolution. Deux exceptions cependant sont relevées : (1) un taux de bon classement plus faible (proche de 60%) sur les zones 2 et 4 et (2) un taux de bon particulièrement bon sur la zone 2 en 1956. La première anomalie s'explique par la spécificité de la morphologie urbaine de la zone 4, zone au contact du noyau urbain dense, qui a subi de nombreuses transformations durant les 50 dernières années en raison de la présence des instances européennes (Parlement européen, Conseil de l'Europe, etc.). L'interprétation visuelle de la morphologie de ces îlots dans le temps met en évidence de profondes restructurations (découpage, regroupement d'îlots). La seconde anomalie (93%) s'explique en partie par la taille de la zone (plus petite que les autres). En 1956, celle-ci est constituée d'un petit nombre d'îlots en raison de sa localisation en périphérie du noyau urbain (zone rurale en 1956).

La figure 5, qui présente les résultats de classification de la zone 1 dans le temps, permet de nuancer ces résultats d'un point de vue thématique. En effet, quelques anomalies de classement peuvent apparaître d'une date à l'autre, par exemple entre 2002 et 2008 (îlot central de grande dimension). Ce type de changement peut être dû à la manière dont les îlots sont découpés en fonction de l'interprétation du réseau de communication entre deux dates (chemin, par exemple).

6. Conclusion et perspectives

Dans cet article, une procédure de classification supervisée fondée sur un algorithme symbolique a été proposée afin de classer automatiquement les tissus urbains et d'extraire de la connaissance sur leur morphologie. Cette procédure utilise deux modules (étiquetage et classification) qui feront partie du module d'analyse spatiale proposé dans la plate-forme open-source GéopenSim. L'algorithme Tilde a permis de bien classer près de 80% des tissus urbains de l'ensemble de l'agglomération et des zones représentatives (en terme d'évolution) à différentes dates. Ceci signifie qu'il est pertinent d'analyser l'évolution tissus urbains selon typologie proposée. Toutefois, l'analyse des règles a mis en évidence la difficulté de caractériser de manière objective (mesures et paramètres statistiques) les tissus urbains. L'extraction d'une connaissance interprétable et exploitable dans un modèle de simulation n'est donc pas facilement automatisable. Ces règles permettent toutefois de relever des seuils pertinents, par exemple en termes de densité d'un îlot. Ces résultats pourraient être améliorés d'une part en créant des échantillons pour les classes plus spécifiques tels que le réseau de communication et d'autre part en utilisant d'autres algorithmes permettant de tenir compte des relations spatiales entre les bâtiments par exemple (Chelgoum *et al.*, 2002).

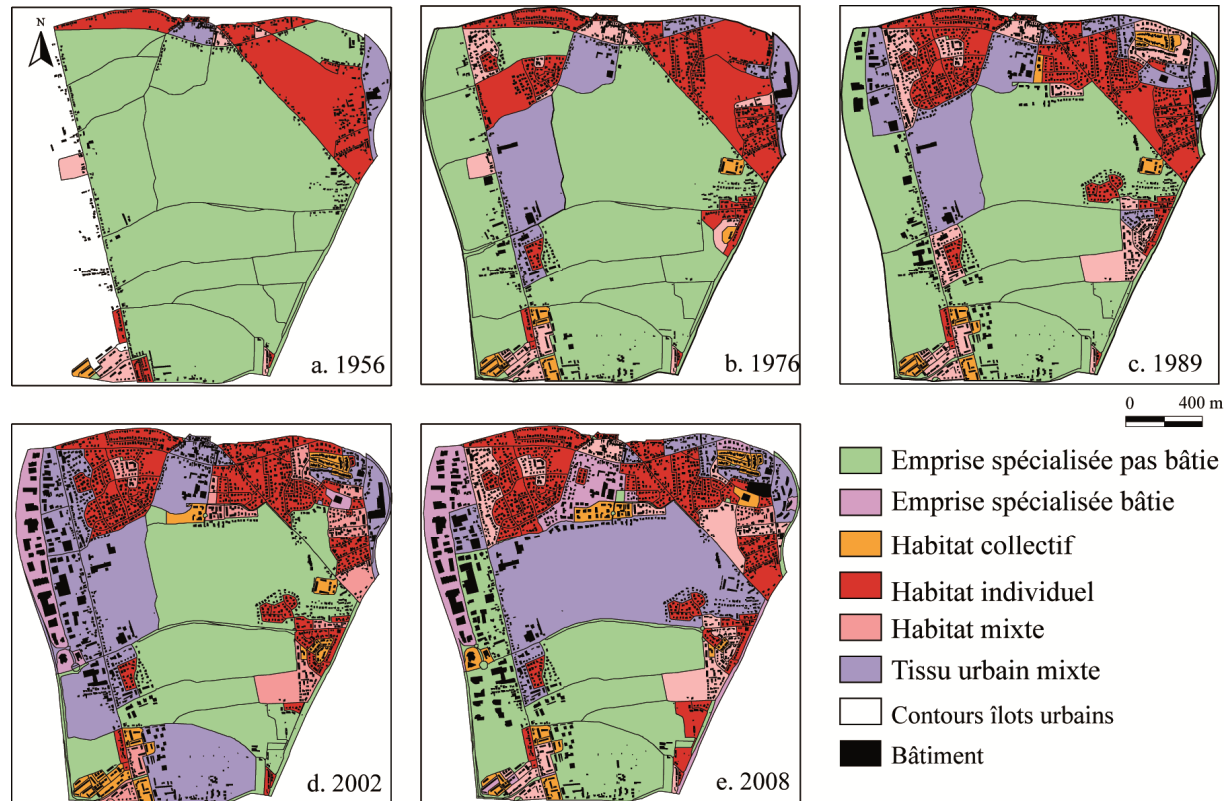


Figure 5. Classification des tissus urbains entre 1956 et 2008.

Ces premiers tests sur l'agglomération de Strasbourg ont également mis en évidence l'importance de l'échantillonnage. D'autres tests, appliquant le modèle le plus pertinent (Test 2c) appris sur Strasbourg, ont été appliqués à la ville d'Orléans (seconde villes de test dans le cadre du projet GéopenSim). Les premiers résultats ont montré des résultats de classification non concluant. On peut en déduire que ce type de procédure ne pourra être appliqué directement à d'autres agglomérations, et devra passer obligatoirement par une étape d'échantillonnage. Ceci reste cependant à confirmer par l'application du modèle de Strasbourg à une ville plus comparable (en termes de taille, de morphologie et de distribution des îlots urbains) comme par exemple Toulouse ou Montpellier.

6. Bibliographie

- ADEUS, 2003, « URBI et ORBI » L'étalement urbain : état des lieux, Dimension Villes & Territoires, n°39, 11 p.
- Baltasavias E.P., 2004, Object extraction and revision by image analysis using geodata and knowledge: current status and steps towards operational systems, *ISPRS Journal of Photogrammetry & Remote Sensing*, 58, 129-151.
- Bauer T., Steinnocher K. (2001): Per parcel land use classification in urban areas applying a rule-based technique. In: *GeoBIT/GIS* 6(2001), pp. 24-27.
- Bessy-Piétry P., 2000, Recensement de la population 1999 - Les formes de la croissance urbaine, *Insee Première*, n°701.
- Blaschke, T., Burnett, C., Pekkarinen, A., 2004. *New contextual approaches using image segmentation for object-based classification*. In: De Meer, F., de Jong, S. (Eds.), *Remote Sensing Image Analysis: Including the spatial domain*. Kluwer Academic Publishers, Dordrecht, pp. 211-236.
- Blokeel H., De Realt L., 1998, Top-down induction of first-order logical decision trees, *Artificial Intelligence*, vol. 101, n°1-2, 12p.
- Bondu A. and Lemaire V., 2007, Etat de l'art sur les méthodes statistiques d'apprentissage actif. RNTI, *Numéro spécial sur l'apprentissage et la fouille de données*, 2007.
- Bouziani M., Goïta K., Heb D., in press 2010, Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge, *ISPRS Journal of Photogrammetry and Remote Sensing*, in press, doi:10.1016/j.isprsjprs.2009.10.002
- Chelghoum N., Zeitouni K., Boulmakoul A., 2002, *Fouille de données spatiales par arbre de décision multi-thèmes*, EGC, 6 p.
- Curie F., Mas A., Perret J., Puissant A., Ruas A., s 2010, *Simuler la densification du tissu urbain au moyen d'un processus de peuplement*, SAGEO 2010, 15 p.
- Donnay, J.P., Barnsley, M.J., Longley, P.A., 2001. *Remote Sensing and Urban Analysis*. Taylor and Francis, London and New York., 315 p.
- Geomediterranée, 2004, BD OCS – Constitution d'une base de données 'occupation du sol' à partir d'images satellitaires. Rapport méthodologique, 34 p.

- Holland, D.A., Boyd, D.S., Marshall, P., 2006. Updating topographic mapping in Great Britain using imagery from high-resolution satellite sensors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60 (3), 212_223.
- IGN, 2002, Spécifications de contenu de la BDTOPO Pays, version 1.2., 65 p
- Jacquin, A., Misakova, L., Gay, M., 2008, A hybrid object-based classification approach for mapping urban sprawl in periurban environment. *Landscape and Urban Planning* 84 (2), 152-165.
- Kurtz C., Passat N., Gancarski P., Puissant A., in press 2010, Multiresolution region-based clustering for urban analysis, *International Journal of Remote Sensing*, 17 p.
- Lachiche N. and P. Flach. 2002: a true first-order bayesian classifier. In Springer-Verlag, editor, 12th International Conference on Inductive Logic Programming, pages 133–148, 2002.
- Lachiche N., 2010, Propositionalisation, In Encyclopedia of Machine Learning, Springer, 6p.
- Lesbueguerues J., Lachiche N., Braud A., Puissant A., Skupinski G., Perret J., 2009, A platform for spatial data labeling in an urban context, *International Opensource Geospatial Research Symposium*, 11 p.
- Nicot B., 1996, Une mesure de l'étalement urbain en France 1982-90, *Revue d'Economie Régionale et Urbaine*, n°1, pp. 71-98.
- Perret J., Ruas A., Boffet-Mas A., 2009, *Understanding urban dynamics: the use of vector topographic database and the creation of spatiotemporal databases*, International Cartographic Conference - ICC 2009, Santiago, Chile, November 2009, 11 p.
- Puissant A., Mas A., Skupinski G., Raimond A.-M., Perret J., Weber C., 2010, Méthodes de constitution de bases de données historiques pour le suivi spatiotemporel des tissus urbains, *Géomatique Expert*, Juillet-Aout 2010, 10 p.
- Quinlan J. R., 1993, C4.5 : Programs for machine Learning, Morgan Kaufmann.
- Saib A., 2009, Répondre à l'étalement urbain : construire un grand territoire métropolitain transfrontalier, *Dimension villes & territoires*, Adeus, n°63, 8 p.
- Salleb A. and Vrain C., 2000, *An Application of Association Rules Discovery to Geographic Information Systems*. In Springer-Verlag, editor, Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, pages 613 – 618.
- Witten I., Frank E., 2005, *Data mining: Practical machine learning tools and techniques*, second edition, Morgan Kaufmann, 50 p.
- Zeitouni K., Yeh L., Le data mining spatial et les bases de données spatiales, *Revue internationale de géomatique, Numéro spécial sur le Data mining spatial*, Vol 9, N°4, 2000.