



HAL
open science

Classification et évolution des tissus urbains à partir de données vectorielles

Anne Puissant, Nicolas Lachiche, Grzegorz Skupinski, Agnès Braud, Julien Perret, Annabelle Mas

► **To cite this version:**

Anne Puissant, Nicolas Lachiche, Grzegorz Skupinski, Agnès Braud, Julien Perret, et al.. Classification et évolution des tissus urbains à partir de données vectorielles. *Revue Internationale de Géomatique*, 2011, 21 (4), pp.513-532. 10.3166/RIG.15.513-532 . hal-03766249

HAL Id: hal-03766249

<https://hal.science/hal-03766249>

Submitted on 31 Aug 2022

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 et évolution des tissus urbains à partir de données vectorielles

Anne Puissant* — Nicolas Lachiche** — Grzegorz Skupinski*,
Agnès Braud** — Julien Perret*** — Annabelle Mas****

* Laboratoire Image, Ville, Environnement, Université de Strasbourg

3, rue de l'argonne, 67 000 Strasbourg, France {prenom.nom}@live-cnrs.unistra.fr

** Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection,
Université de Strasbourg - Pôle API Bd Sébastien Brant 67412 Illkirch Cedex,
France {prenom.nom}@unistra.fr

*** Laboratoire COGIT, IGN

73 avenue de Paris, 94160 Saint Mandé Cedex, France {prenom.nom}@ign.fr

**** Laboratoire CEDETE, Université d'Orléans

10, rue de tours, BP 46257, 45065 Orléans cedex2, {prenom.nom}@univ-orleans.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 deux méthodes de fouille de données relationnelles (la cardinalisation et les quantiles) 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, mettent en évidence l'intérêt de ce type de méthodes pour classer les tissus urbains (plus de 80% de taux de bon classement sur les données historiques). A partir de ces résultats, une analyse statistique des évolutions est présentée.

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 two relational data mining methods (Cardinalisation and Quantiles) 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 (80% of global accuracy on historical database). From the classification results, a statistical analysis of evolutions is described.

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

KEYWORDS: urban fabric, morphology, vector databases, classification, decision tree, evolution.

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 proposer des scénarios de développement urbain ou de développer des concepts différents pour trouver des réponses adaptées au défi de l'étalement urbain.

Le projet 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 (Perret *et al.*, 2009). 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 ensuite analysées afin d'identifier des règles d'évolution à différents niveaux méso-géographiques (e.g. la ville ou l'îlot) et des statistiques sur des objets spécifiques comme les bâtiments (Curie *et al.* 2010). Dans ce projet, l'identification et la caractérisation des îlots sont deux étapes préalables à l'analyse de leur évolution. Celles-ci constituent un problème de fouille de données relationnelles. L'identification des îlots peut être automatisée par une procédure de classification alors que la caractérisation revient à extraire des connaissances sur les tissus urbains.

Dans ce contexte, l'objectif de cet article est de proposer une méthode de classification automatique des îlots urbains permettant d'identifier le type de tissu urbain et de caractériser sa morphologie. Deux méthodes d'apprentissage supervisé symbolique (la cardinalisation et les quantiles), permettant d'obtenir un modèle de classification représenté sous la forme de règles de décision, sont testées sur l'agglomération de Strasbourg. 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. Une analyse des évolutions des îlots est ensuite présentée à la section 6. Les conclusions et perspectives sont finalement exposées à la section 7.

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-OCCS (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).

Partant du constat que la classe d'un îlot s'exprime souvent en fonction de nombres ou de proportions de ses bâtiments présentant une certaine caractéristique (par exemple 10 « grands » bâtiments ou 10% de « grands » bâtiments), deux nouvelles approches qui introduisent cette expressivité (*la cardinalisation* et *les quantiles*) ont été proposées (Braud et Lachiche, 2011). Celles-ci s'inscrivent dans le cadre d'une branche de la fouille de données relationnelles appelée propositionalisation (Lachiche, 2010) qui permet d'enrichir le langage utilisé pour représenter les règles. Elle vise à simplifier un problème relationnel en le transformant en un problème attribut-valeur, c'est à dire à construire une table unique à partir de la base de données relationnelle. La complexité est alors moindre et un algorithme d'apprentissage attribut-valeur tel qu'un arbre de décision peut s'appliquer. Le problème est de construire cette nouvelle table, avec de nouveaux attributs appelés *features* capturant les informations contenues dans plusieurs tables. Considérons l'extrait de nos données ci-dessous :

Table Ilots

IDilot	densité	aire	élongation	convexité	classe
9601	0,194	6812	0,772	0,921	« h_indiv »
9602	0,455	11119	0,470	0,916	« h_coll »
...

Table Bâtiments

IDbat	aire	élongation	convexité	IDilot
4519	122	0,765	1,00	9601
4521	122	0,752	1,00	9601
4528	119	0,948	1,00	9601
4537	112	0,918	0,999	9601
4866	2239	0,772	0,595	9602
...

L'objectif est d'enrichir la table îlots avec des *features* en synthétisant les données concernant les bâtiments de ces îlots (par exemple des statistiques sur l'aire de ses bâtiments). Ces deux approches construisent deux types de *features* :

- la cardinalisation produit des *features* correspondant au seuil minimal sur un attribut de la table 'Bâtiments' pour que l'îlot contienne au moins k bâtiments vérifiant ce seuil (par exemple l'aire minimale telle que l'îlot contienne au plus 3 bâtiments d'aire inférieure ou égale à ce seuil) énumération de premier niveau.

- les quantiles produisent des *features* correspondant au seuil minimal sur un attribut de la table 'Bâtiments' pour que l'îlot contienne au moins une fraction (proportionnelle à un nombre k qui fixe le nombre de quantiles) du nombre total de ses bâtiments vérifiant ce seuil, par exemple, pour des quartiles (k=4), les aires minimales telles qu'au moins 25%, 50% et 75%, respectivement, des bâtiments de l'îlot aient des aires inférieures ou égales à ces seuils.

Ces deux approches sont testées dans cet article. Elles s'appliquent aux 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, comme dans l'exemple ci-dessus.

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 créées sur quatre zones tests réparties dans les trois couronnes de l'agglomération (Figure 1). La zone 3 (axe de développement vers l'Allemagne) et la zone 4 (quartier européen) sont deux zones au contact direct du centre-ville (péricentre) qui ont fait et font encore l'objet d'importantes opérations de rénovation urbaine. La zone 1, localisée dans la deuxième couronne, est représentative d'un étalement urbain soumis à des contraintes naturelles (remontée de la nappe phréatique). La zone 2 représente une commune périurbaine de la troisième couronne soumise à un étalement urbain continu. Les données historiques ont été produites en fonction des supports disponibles (photographies aériennes et cartes topographiques), 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 prises 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 et n°9 – 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 données topographique (objets de type 'réseaux de communication' regroupant le réseau routier, ferré et hydrographique).

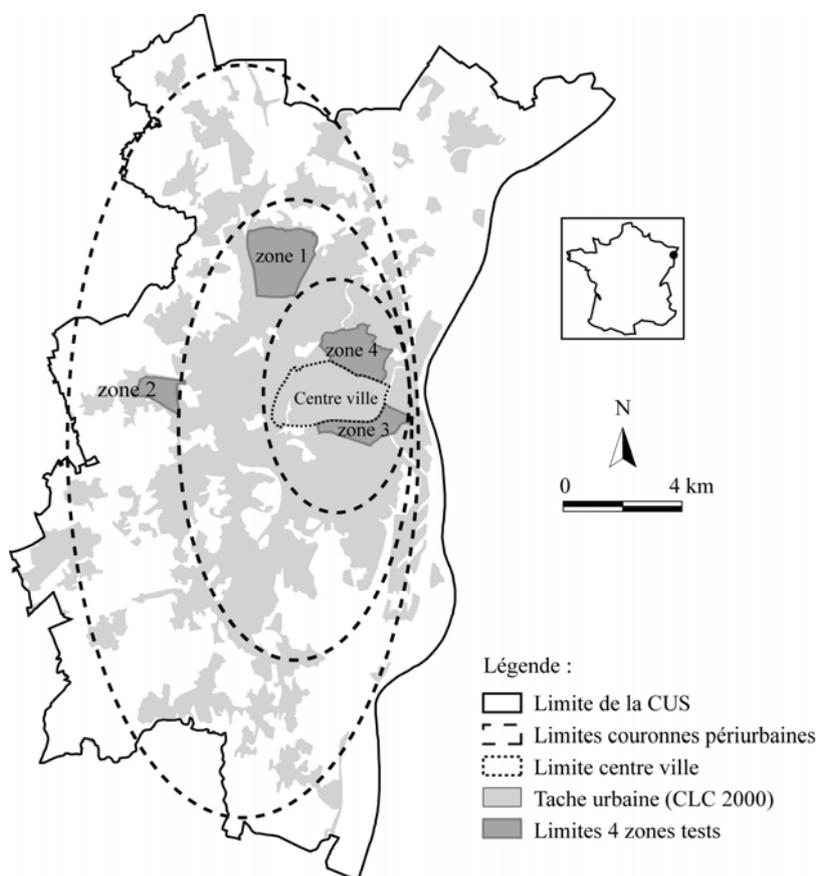


Figure 1. Localisation des quatre zones tests dans la Communauté urbaine de Strasbourg (CUS).

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>Mélange de classes d'habitat (classe 1, 2, 3) et de la classe 6 (emprises spécialisées bâties)</i>
6. Emprises spécialisées bâties : <i>emprises militaires, zones industrielles, commerciales, emprises scolaires et universitaires, emprises hospitalières, emprises publiques ou patrimoniales.</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 définies pour cartographier la ville au 1/10 000^e

Des mesures morphologiques sont ensuite calculées pour caractériser d'une part les bâtiments (*e.g.* l'aire, l'élongation et la convexité) et d'autre part l'organisation spatiale des bâtiments dans un îlot (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'entraînement et à la validation) est créée à partir d'un module d'étiquetage open source fondé sur Geoxygene (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), les deux méthodes testées (la cardinalisation et les quantiles) permettent 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.

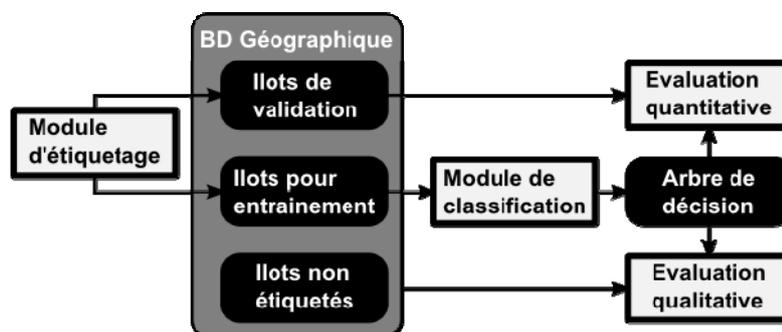


Figure 2. Etapes de la méthode de classification

Les résultats de la classification sont ensuite validés d'une part, qualitativement en analysant l'arbre de décision produit et d'autre part quantitativement par validation croisée. Lorsque l'on dispose d'un seul jeu d'îlots étiquetés (entraînement et validation 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 tailles semblables, en prenant soin que chaque groupe contienne un nombre d'îlots de chaque classe proportionnel aux nombres présents dans le jeu complet (Witten et Frank, 2005). Le nombre de 'paquets', en général 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 quatre zones (cf. 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 (entraînement 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). Il s'agit également d'identifier la méthode la plus stable et la plus performante en termes de pourcentages de bon classement. Le second objectif est d'appliquer le modèle appris le plus pertinent (taux de classement le plus élevé), à la base de données historique (section 5.2). Les résultats sont également évalués qualitativement par l'interprétation visuelle et par l'analyse de l'arbre de décision (règles produites).

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

Pour chaque méthode testée (la cardinalisation et les quantiles), trois tests de classification sont appliqués à l'ensemble des îlots urbains de l'agglomération et aux quatre zones tests. 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'exemples d'îlots urbains (Tableau 2). Ces îlots (entraînement 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 5% du nombre total d'îlots. Le nombre d'exemples pour les six classes thématiques comporte, pour ce test, au minimum 30 exemples par classe afin d'assurer la significativité statistique des mesures (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 les quatre zones tests comme échantillons (entraînement et validation confondus) qui représentent également 5 % du nombre total d'îlots. L'objectif est de vérifier si la répartition des exemples influence le résultat de la classification.

Afin de vérifier la stabilité de chaque modèle, un **troisième test** de classification a classé les 4 zones d'intérêt en utilisant les échantillons d'entraînement sélectionnés sur l'ensemble de l'agglomération.

<i>2002</i>	<i>Test 1</i>	<i>Test 2</i>
<i>Classes d'îlots urbains</i>	<i>Agglo</i>	<i>4 zones</i>
Habitat pavillonnaire	40	91
Habitat collectif	39	52
Habitat discontinu mixte	71	59
Tissu discontinu mixte	60	60
Emprises spécialisées bâties	30	41
Emprises spécialisées peu ou pas bâties	34	50
<i>Nb total d'îlots d'entraînement</i>	<i>274</i>	<i>319</i>

Tableau 2. Récapitulatif du nombre d'îlots étiquetés (entraînement et validation confondues) selon les tests effectués

L'analyse du taux de bon classement (Tableau 3) montre que la méthode d'échantillonnage influence le résultat de classification, quelle que soit la méthode utilisée (test 1 et test 2). Un échantillonnage mieux réparti sur l'ensemble de la zone à classer permet, comme on pouvait s'y attendre, d'obtenir un meilleur taux de bon classement. On peut noter que la différence de taux de bon classement obtenue avec la méthode des quantiles, est plus faible (4%) que celle obtenue avec la cardinalisation (9%). Les résultats du troisième test mettent en évidence une diminution du taux de bon classement lorsque les échantillons d'entraînement sélectionnés sur l'ensemble de l'agglomération sont utilisés pour classer les 4 zones tests. Toutefois, si l'on considère le fonctionnement 'moyen', autrement dit la stabilité de la méthode, le plus performant en terme de bon classement est la méthode des quantiles entraînée sur les 4 zones tests. Afin de vérifier ce résultat, nous avons appliqué la méthode des quantiles (test 2) sur l'ensemble des îlots de l'agglomération en 1989. Le taux de bon classement obtenu confirme la performance du modèle qui atteint également 76 % de bon classement contre 73 % pour la cardinalisation.

La Figure 3 présente les résultats de classification sur les quatre zones tests. La comparaison avec une classification experte montre que la majorité des îlots urbains de type individuel et collectif sont bien classés (par exemple au nord de la zone 1 – Figure 3). Les erreurs de classification apparaissent essentiellement sur les classes d'habitat et de tissus mixtes. La méthode confond également l'habitat collectif et l'habitat mixte car le principal critère morphologique les différenciant est la taille des bâtiments. L'analyse de la classe d'emprise spécialisée peu ou pas bâtie remet en cause la pertinence du découpage en îlots. En effet, comme le montre la figure 1a et plus précisément l'îlot central de grande dimension (zone de culture, dont la partie

centrale est non constructible en raison des remontées de nappe phréatique), certains bâtiments en périphérie appartiennent à cet îlot central alors que morphologiquement, si l'on fait abstraction du découpage en îlot, ceux-ci sont semblables aux îlots contigus.

Cardinalisation (2002)	Entraîné sur agglomération	Entraîné sur quatre zones
Testé sur agglomération	73% (<i>test 1</i>)	64% (<i>test 2</i>)
Testé sur quatre zones	64% (<i>test 3</i>)	/
	69%	64%
Quantiles (2002)	Entraîné sur agglomération	Entraîné sur quatre zones
Testé sur agglomération	80% (<i>test 1</i>)	76% (<i>test 2</i>)
Testé sur quatre zones	66% (<i>test 3</i>)	/
	73 %	76 %

Tableau 3. Résultats de classification en termes de taux de bon classement (exprimé en pourcentages) selon les tests effectués (données 2002) – (en gras – taux de bon classement moyen)

L'analyse des résultats de l'arbre de décision (obtenus à partir du modèle issu la méthode des quantiles – test 2) 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 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 (Curie *et al.*, soumis 2011). D'autres tests sont en cours afin de vérifier si d'autres méthodes tenant compte notamment du voisinage ou des relations spatiales permettraient d'extraire une connaissance réellement exploitable.

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

Le modèle construit le plus satisfaisant (test 2 - quantile) 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 4. Afin de vérifier que notre affirmation sur la stabilité du modèle se justifie, nous avons

également appliqué le test 1 – quantiles sur la base de données historique. Les résultats de 2002 ne sont pas mentionnés, puisque ces îlots ont contribué à construire le modèle appliqué aux autres dates.

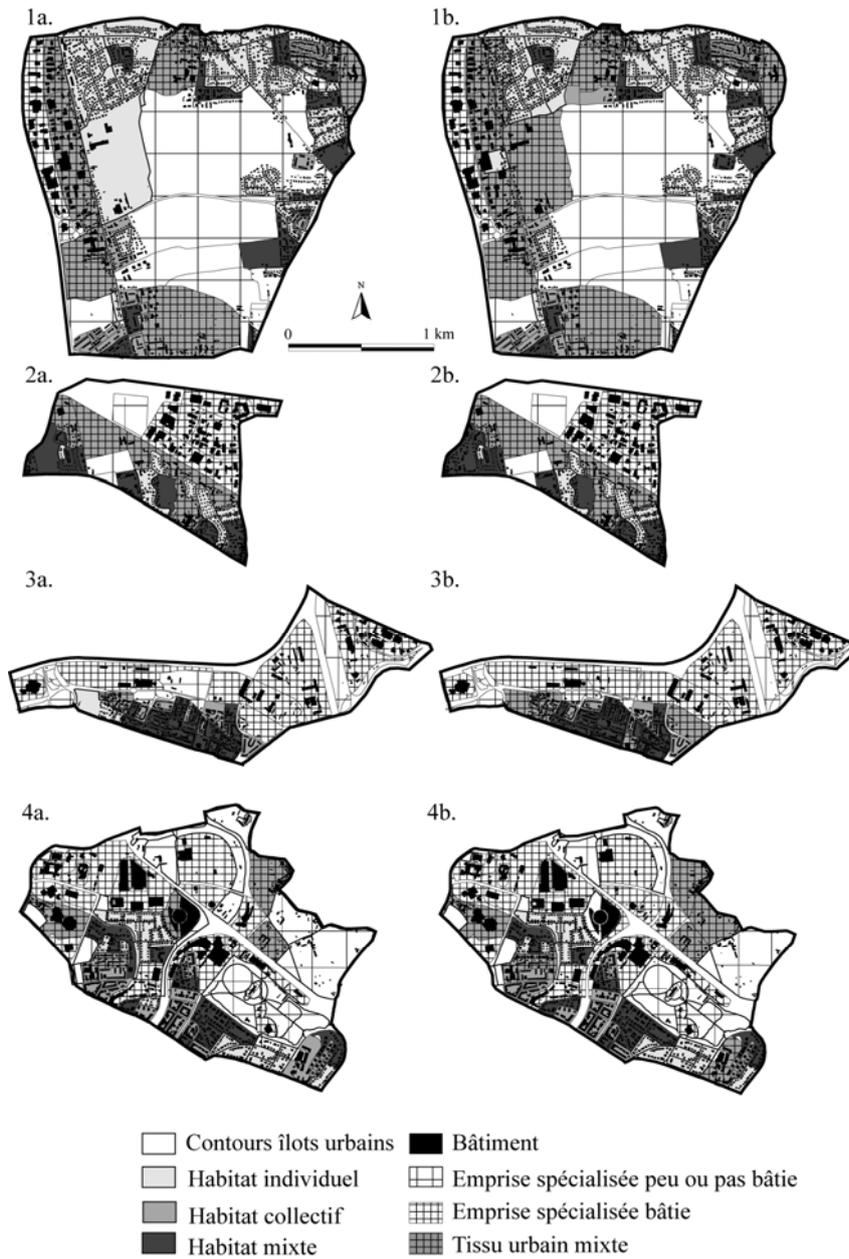


Figure 3. (a) *Extrait de la classification (ensemble de l'agglomération - méthode des quantiles entraînée sur les 4 zones)* (b) *Validation – étiquetage expert.*

4 zones	Test 1 - Cardinalisation	Test 2 - Quantiles
2008	66	76
2002	/	/
1989	61	88
1976	58	82
1966	59	84
1956	64	81

Tableau 4. *Taux de bon classement (%) des quatre zones, selon les dates disponibles.*

Les résultats montrent que le taux de bon classement aux dates antérieures à 2002 (excepté en 2008) est supérieur à 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 afin d'analyser leur évolution. Les résultats du test 1 confirment la moins bonne stabilité et donc la moins bonne performance de la méthode de cardinalisation avec des taux de bon classement inférieurs en moyenne de 20% par rapport à la méthode des quantiles.

La figure 4, qui présente les résultats de classification de la zone 1 aux différentes dates, 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. Analyse des évolutions

La simulation des dynamiques urbaines nécessite de disposer de règles d'évolution qui peuvent être définies soit par des experts géographes soit par l'analyse des statistiques acquises sur une zone donnée. Ces règles sont définies dans le cadre du projet GeOpenSim au niveau de l'îlot. Elles doivent fixer l'état objectif auquel doit tendre l'îlot au cours de la simulation à travers deux principales caractéristiques : la densité objectif et le type objectif (habitat individuel, habitat collectif, etc).

Dans ce contexte, l'objectif est d'analyser les classifications aux différentes dates (section 5), comme données d'entrées, afin de proposer cet état objectif lors d'une simulation. Une analyse statistique des matrices de changements (quantification du nombre d'îlot évoluant d'un état 1 à un état 2), permet de

répondre à cet objectif et d'identifier des types d'évolution (densification, spécialisation d'un tissu urbain en habitat, mixité, etc).

Dans un premier temps, les changements sont calculés entre 1989 et 2002 sur l'ensemble de l'agglomération (Tableau 5). Dans un second temps, les matrices de changement sont calculées sur les 4 périodes (1956-1976, 1976-1989, 1989-2002 et 2002-2008) et sur les 4 zones d'intérêt (Tableau 6a, 6b, 6c, 6d). L'analyse des distributions marginales permet de quantifier et d'identifier les types évolutions.

Sur l'ensemble de l'agglomération (23154 îlots), 5% des îlots ont subi un changement (Tableau 5). Un tiers des changements correspond à une spécialisation des emprises peu ou pas bâties. En effet, ces îlots de faible densité (< 0.05 – section 5) sont les premiers à subir une transformation géométrique, *a priori* due à une modification (construction) d'un nouveau réseau communication (découpage de l'îlot). Pour les deux-tiers d'îlots modifiés entre 1989 et 2002, un phénomène de mixité est observé avec une transformation de l'habitat mixte majoritairement en tissu urbain mixte (mixité des fonctions) et une transformation de l'habitat pavillonnaire en habitat mixte (mixité sociale). Ces résultats sur les emprises peu ou pas bâties sont à nuancer puisqu'ils représentent seulement 3% des îlots de ce type alors que les changements de l'habitat et du tissu mixte représentent près de 30% chacun.

Sur les 4 zones d'intérêts (2391 îlots), 10% des îlots ont changé de classe thématique (Tableau 6). Sur l'ensemble de la période (1956-2008), deux-tiers des changements correspondent ici à une modification des emprises peu ou pas bâties. Les autres principales modifications rejoignent les résultats précédents, puisque les premiers îlots à être modifiés sont les classes d'habitat et de tissu mixte. Les résultats montrent également que les changements observés sur les classes d'habitat individuel, mixte et collectif représentent un tiers des changements. L'analyse de ces matrices par périodes met en évidence une décroissance progressive du pourcentage de changements (tableau 7). Ces écarts sont dus en partie à l'inégalité des périodes analysées. Le calcul du pourcentage des changements par an permet d'identifier que la période de 1976-1989 est la période qui a connu le plus de changements. En effet, cette période correspond à la vaste période de construction de nouveaux logements à la fois individuels et collectifs dans la deuxième voire troisième couronne périurbaine. Les chiffres (Tableau 6) montrent en effet que respectivement 45% et 19% d'îlots peu ou pas bâti sont devenus de l'habitat individuel et collectif.

Ces statistiques issues de l'analyse des distributions marginales peuvent être facilement utilisées afin de choisir le type objectif d'un îlot lors d'une simulation. Elles permettent de confirmer et surtout de quantifier les types d'évolutions des tissus urbains sur les 50 dernières années. Ces résultats pourraient être affinés, en tenant compte de la distance au centre afin de mieux localiser les types de changements observés. Des premiers tests sont en cours.

Type d'évolution	Période	% changements	% changement par an
1956-1976	20 ans	20%	1
1976-1989	13 ans	17%	1,3
1989-2002	13 ans	7%	0,5
2002-2008	6 ans	2%	0,3

Tableau 7. Taux de changement des îlots urbains selon les périodes.

7. Conclusion et perspectives

Dans cet article, deux méthodes d'apprentissage supervisé symbolique ont été testées 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 font partie du module d'analyse spatiale proposé dans la plate-forme open-source GeopenSim. La méthode des quantiles a permis de bien classer près de 80% des tissus urbains de l'ensemble de l'agglomération et des zones représentatives (en termes d'évolution) à différentes dates. La typologie proposée est donc pertinente pour ce type d'analyse. Toutefois, les règles extraites ne permettent pas 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 ont permis 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 méthodes permettant de tenir compte des relations spatiales entre les bâtiments par exemple (Chelgoum *et al.*, 2002). Des tests de classification ont été également appliqués à la ville d'Orléans (seconde ville test du projet GeOpenSim), à partir des îlots d'entraînement de Strasbourg, afin de vérifier la généralité de la méthode. Les résultats de classification sont non-concluants. On peut en déduire qu'une étape d'échantillonnage est nécessaire afin de tenir compte des particularités morphologiques de chaque ville. Ceci reste cependant à confirmer par l'application du modèle de Strasbourg à une ville comme Toulouse ou Montpellier, plus proche de Strasbourg en termes de taille de morphologie et de distribution des îlots urbains.

En ce qui concerne l'analyse des évolutions, elle a permis de quantifier les changements et d'analyser les trajectoires d'un îlot au cours du temps (par période). Ces statistiques sont utiles dans l'étape de simulation pour fixer des objectifs d'évolution du type d'un îlot. Cette analyse a été réalisée en considérant 4 périodes. Il serait dès lors intéressant d'analyser la trajectoire d'un îlot au cours du temps afin d'identifier des transitions 'type' (état1 – état 2 – état3). Ces recherches ont fait l'objet d'autres développements dans le cadre de GeOpensim et ont permis de proposer un outil de suivi et de visualisation de ces transitions (Gancarski *et al.*, soumis 2011).



Figure 4. Classification des tissus urbains de la zone 1 entre 1956 et 2008.

Nb îlots		Tissus urbains en 2002						Total	Nb îlot (état 1 => état 2)	Changement (%)	Part de changement (% en ligne)	
		Emp. peu b.	Emp. b.	H. col.	H ind.	H mixte	T. mixte					
Tissus urbains en 1989	Emprise peu bâtie	19132	95	95	168	116	42	19648	516	45%	3%	
	Emprise bâtie	9	626	27	0	12	19	693	67	6%	10%	
	H. collectif	10	33	484	0	28	16	571	87	8%	15%	
	H. individuel	19	2	10	898	102	19	1050	152	13%	14%	
	H mixte	43	17	51	59	610	60	840	230	19%	27%	
	Tissu mixte	2	23	17	6	47	257	352	95	8%	27%	
Total		19215	831	686	1131	915	413	23154	1147	100%		
		% total d'îlots changés entre 1989 et 2002								5%		

Tableau 5. Matrice de changements (exprimé en nombre d'îlots) entre 1989 et 2002, sur l'ensemble de l'agglomération.

Nb îlots		Tissus urbains en 2008						Total	Nb îlots (état 1 => état 2)	Changements (%)	Part de changement (% en ligne)	
		Emp. Peu b.	Emp. b.	H. col.	H ind.	H mixte	T. mixte					
Tissus urbains en 1956	Emprise peu bâtie	1345	25	30	45	29	16	1490	145	60%	10%	
	Emprise bâtie	2	127	6	0	0	4	139	12	5%	9%	
	H. collectif	9	3	124	1	4	1	142	18	7%	13%	
	H. individuel	13	0	1	281	15	2	312	31	13%	10%	
	H mixte	9	4	3	10	205	5	236	31	13%	13%	
	Tissu mixte	0	2	1	1	2	66	72	6	2%	8%	
		1378	161	165	338	255	94	2391	243	100%		
		% total d'îlots changés entre 1956 et 2008								10%		

Tableau 6. Matrice de changements (exprimé en nombre d'îlots) entre 1956 et 2008, sur les quatre zones tests.

7. 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.
- Braud A. and Lachiche N., 2011,. Propositionaliser des attributs numériques sans les discrétiser, ni les agréger. In *Extraction et Gestion des Connaissances (EGC 2011)*, pp 437-442.
- 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.
- Blockeel H., De Reult 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.
- Gancarski P., Petitjean F., Puissant A., soumis 2011, Spatial evolution of urban fabric under Time Warping using multi-temporal topographic databases, *CAGEO*.
- 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 BDTOPPO 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., 2010, Multiresolution region-based clustering for urban analysis, *International Journal of Remote Sensing*, vol. 31, n°22, pp. 5941-5973.
- Lachiche N., 2010, Propositionalisation, In *Encyclopedia of Machine Learning*, Springer, 6p.
- Lesbueguerries 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 Kauffmann.
- Saïb 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.