

ÉCOLE DOCTORALE DES SCIENCES DE LA TERRE  
ET DE L'ENVIRONNEMENT (ED 413)

Thèse présentée par

**Simon Rougier**

Pour obtenir le grade de  
Docteur de l'Université de Strasbourg  
Discipline : Géographie

Préparée au

Laboratoire Image, Ville, Environnement UMR 7362

**Apport des images satellites  
à Très Haute Résolution Spatiale couplées  
à des données géographiques multi-sources  
pour l'analyse des espaces urbains**

**Soutenue le 28 septembre 2016**

*Membres du jury :*

**Anne Puissant**

**Sandrine Glatron**

**Jean- Christophe Foltête**

**Sébastien Gadal**

**Christiane Weber**

**David Sheeren**

Maître de conférence, HDR, Université de Strasbourg

Chargée de recherche, HDR, CNRS, Université de Strasbourg

Professeur, Université de Bourgogne Franche-Comté

Professeur, Université d'Aix-Marseille

Directrice de recherche, CNRS, TETIS, Montpellier

Maître de conférence, INP-ENSAT, Toulouse

Directrice

Directrice

Rapporteur

Rapporteur

Examinatrice

Examineur



## Remerciements

La rédaction de ces remerciements signe pour moi la fin de la rédaction de cette thèse. Quatre ans difficiles, mais qui auront été enrichissants et m'auront permis de faire de superbes rencontres. Je souhaite donc remercier l'ensemble des personnes qui auront contribué directement ou indirectement, de près ou de loin à ce travail.

Je tiens tout d'abord à remercier ma directrice de thèse Anne Puissant qui m'a permis de réaliser cette thèse à la suite de mon Master. Merci beaucoup de m'avoir fait confiance, de m'avoir trouvé un financement, de m'avoir suivi depuis mon premier stage il y a plus de cinq ans maintenant et de m'avoir si souvent remis en selle. J'ai beaucoup appris avec toi et te remercie pour ta bonne humeur permanente. Je remercie aussi ma codirectrice Sandrine Glatron d'avoir bien voulu m'encadrer. Merci pour tes corrections, ta gentillesse et ton humanité.

Mes remerciements vont aussi à la Zone Atelier Environnementale Urbaine et à son ancienne directrice Christiane Weber pour m'avoir offert un contrat qui m'a permis de m'inscrire pour une quatrième année. Je remercie également le CNES pour la mise à disposition des images Pléiades dans le cadre de la Recette Thématique Utilisateurs (RTU) et l'IGN pour les pré-traitements effectués sur les images. Merci aussi au service Géomatique et Connaissance du territoire de l'Eurométropole de Strasbourg et à CIGAL pour la mise à disposition des données sur le territoire de Strasbourg.

Je tiens aussi à remercier les membres du jury qui ont accepté d'évaluer ce travail de thèse : Jean-Christophe Foltête, Sébastien Gadal, Christiane Weber et David Sheeren.

Cette thèse m'a aussi permis de participer à un programme de recherche, le projet VALI-URB et je remercie l'ensemble des participants à ce projet : Laurence Hubert-Moy, Jean Nabucet, Julien Deniau et Léo Vétillard de l'Université de Rennes, Éléonore Wolff, Émilie Hanson et Omar Benarchid de l'Université de Bruxelles qui ont permis au début de cette thèse de nombreux échanges enrichissants. Un grand merci à Omar pour ses codes R qui m'ont permis d'évaluer mes segmentations de manière supervisée dans le chapitre 3 et à Ruben Van De Kerchove qui avec les siens m'ont permis de découvrir le package caret.

Cette thèse est l'aboutissement de quatre années de travail au Laboratoire, Image, Ville et Environnement et je remercie l'ensemble de ses membres pour leur accueil, leur

aide, leur gentillesse et leur soutien. Je pense particulièrement à Estelle, Arnaud, Aziz, Grégoire, Alexis, Sophie, Adrien, Aurélie, Christophe, Jacky, Yasmina, Émile.

Je pense aussi à tous les doctorants passés, présents ou futurs peut-être avec qui j'ai noué de vraies relations d'amitié au cours de ces années et toujours prêts pour boire un café, une bière, fumer une clope. Merci à Frank, Jennifer, Sophie, Sébastien, Valentin, Léa, Lucie. Et aussi à tous mes collègues de bureau Pierre, André, Wissal, David, Filsa, Pierrick, Laurent. Heureusement que vous étiez là, je n'y serais pas arrivé sans vous !

Je remercie une nouvelle fois André sans qui cette thèse ne serait pas la même. Merci pour tout le temps que tu as passé à me montrer tes méthodes d'apprentissage actif, mais aussi à les faire évoluer avec moi, à répondre à mes questions et de m'avoir envoyé dans de bonnes directions. En Master tu m'as aussi fait découvrir R, il m'aura valu quelques migraines, mais il ne m'aura pas lâché !

Je pense aussi très fort à tous mes amis toujours dans l'Ouest qui m'auront permis de décompresser à chaque fois qu'on pouvait se voir et qui ont toujours eu une oreille attentive : Ronan, Noémie, Guigui, Didie, Guillaume, Aline, Tigrou, Manu, Céline, Hélène, Romain, Aline... On se revoit bientôt !!!

Merci aussi à Liliane et Pierre, plein de bons moments passés avec vous et encore beaucoup à venir. Et merci Lili pour les îlots de Rennes ; la photo-interprétation ce n'est pas mon truc !

Un grand merci à ma famille, mon père, Judith, Benjamin pour avoir toujours été là, de m'avoir écouté, tant soutenu et avoir à peu près tout géré ce que je ne pouvais pas faire d'ici.

Cette rédaction s'approche de la fin et je tiens à remercier du fond du cœur mon grand-père qui nous a quittés en juin dernier peu avant la finalisation de cette thèse. Je sais que tu aurais aimé assister à ça et je suis triste que tu ne sois pas là pour le voir. On parlait peu, on n'était pas toujours d'accord, mais tu m'as toujours poussé à reprendre mes études et tu m'as soutenu financièrement et moralement pendant ces dix longues années d'études. Pour beaucoup, tu as fait de moi ce que je suis aujourd'hui. AVE PAPI !

Enfin Sara. Merci d'avoir accepté de me suivre à Strasbourg et de recommencer tes études à zéro ; de toujours avoir été là pour me soutenir ; d'avoir supporté ma mauvaise humeur, mes frustrations, ma colère, pas toujours exprimées, mais souvent présentes. Merci pour tous ces moments passés avec toi depuis six ans, ne change pas. Une nouvelle vie commence pour nous...

## Liste des abréviations

**ACP** : Analyse en Composante Principale  
**ADEUS** : Agence de Développement et d'Urbanisme de l'agglomération Strasbourgeoise  
**ANR** : Agence Nationale de la Recherche  
**APUR** : Atelier Parisien d'Urbanisme  
**ASP** : Agence de Services et de Paiement  
**AUCAME** : Agence d'études d'Urbanisme de Caen-Métropole  
**AUDIAR** : Agence d'urbanisme de Rennes  
**BDOCS** : Base de données d'occupation des sols de la région Alsace  
**CAH** : Classification Ascendante Hiérarchique  
**CART** : *Classification and Regression Tree*  
**CGDD** : Commissariat Général au Développement Durable  
**CIGAL** : Coopération pour l'Information Géographique en Alsace  
**CLC** : Corine Land Cover  
**COV** : Composés Organiques Volatils  
**CNES** : Centre National d'Études Spatiales  
**ED2** : *Euclidean Distance 2*  
**EEA** : *European Environment Agency*  
**ESP** : *Estimation of Scale Parameter*  
**GLCM** : *Grey-Level Co-Occurrence Matrix*  
**GMES** : *Global Monitoring for Environment and Security*  
**GNG** : *Growing Neural Gaz*  
**HRL** : *High Resolution Layers*  
**HRS** : Haute Résolution Spatiale  
**IGN** : Institut national de l'information géographique et forestière  
**INSEE** : Institut National de la Statistique et des Études Économiques  
**IR-MAD** : *Iterative Re-weighted Multivariate Alteration Detection*  
**IRIS** : Îlots Regroupés pour l'Information Statistique  
**k-NN** : *k-Nearest Neighbor*  
**KDD** : *Knowledge Discovery in Databases*  
**LCZ** : *Local Climate Zone*  
**Lidar** : *light detection and ranging*  
**LIVE** : Laboratoire, Image, Ville, Environnement  
**MANOVA** : *Multivariate Analysis of Variance*

**MCSAL** : *Multi-Class Spatial Active Learning*  
**MCSAL<sub>D</sub>** : *Multi-Class Spatial Active Learning considering Diversity*  
**MCSAL<sub>S</sub>** : *Multi-Class Spatial Active Learning integrating Stratification*  
**MEDDE** : Ministère de l'Environnement, de l'Énergie et de la Mer  
**MNE** : Modèle Numérique d'Élévation  
**MNS** : Modèle Numérique de Surface  
**MNT** : Modèle Numérique de Terrain  
**MRS** : Moyenne Résolution Spatiale  
**MSAVI2** : *Modified Soil-Adjusted Vegetation Index 2*  
**NG** : *Neural Gaz*  
**NSR** : *Number of Segment Ratio*  
**OBIA** : *Object-Based Image Analysis*  
**OCS GE** : Occupation du Sol à Grande Échelle  
**OOB** : *Out-Of-Bag*  
**OSM** : OpenStreetMap  
**PLU** : Plan Local d'Urbanisme  
**PSE** : *Potential Segmentation Error*  
**RF** : *Random Forest*  
**RGE** : Référentiel à Grande Échelle  
**ROC<sub>LV</sub>** : *Rate Of Change of Local Variance*  
**RPG** : Registre Parcellaire Graphique  
**RTU** : Recette Thématique Utilisateurs  
**SCoT** : Schéma de Cohérence Territoriale  
**SIG** : Système d'Information Géographique  
**SOeS** : Service de l'Observation et des Statistiques  
**SOM** : *Self Organizing Map*  
**SRCE** : Schéma Régional de Cohérence Écologique  
**SRGS** : *Stratified Random Grid Sampling*  
**SVM** : *Support Vector Machine*  
**THRS** : Très Haute Résolution Spatiale  
**VS** : Variable Synthétique  
**ZNIEFF** : Zones Naturelles d'Intérêt Écologique Faunistique et Floristique







# Sommaire

Remerciements .....	i
Liste des abréviations .....	iii
Sommaire .....	vii
Introduction générale .....	1
<b>PARTIE I</b>	
<b>Structures et formes urbaines : concepts, typologies, données disponibles et état de l'art .....</b>	<b>5</b>
Introduction de la partie I .....	7
<b>CHAPITRE 1 : Structure spatiale urbaine et morphologie urbaine : définitions et concepts .</b>	<b>9</b>
<b>CHAPITRE 2 : Données géographiques et méthodes existantes pour la cartographie de la structure urbaine .....</b>	<b>37</b>
Conclusion de la partie I et positionnement de recherche .....	71
<b>PARTIE II</b>	
<b>Extraction de la végétation urbaine à partir d'images satellites à Très Haute Résolution Spatiale .....</b>	<b>77</b>
Introduction de la partie II .....	79
<b>CHAPITRE 3 : Sélection de la segmentation optimale pour l'extraction des éléments verts urbains .....</b>	<b>83</b>
<b>CHAPITRE 4 : Comparaison de stratégies d'échantillonnage pour la classification d'images THRS .....</b>	<b>111</b>
<b>CHAPITRE 5 : Généralisation de l'approche orientée-objet par apprentissage actif pour la classification de la végétation .....</b>	<b>137</b>
Conclusion de la partie II .....	163
<b>PARTIE III</b>	
<b>Cartographie et analyse quantitative de la structure urbaine à partir de données multi-sources .....</b>	<b>167</b>
Introduction de la partie III .....	169
<b>CHAPITRE 6 : Création d'une base de données d'éléments urbains multi-sources .....</b>	<b>171</b>
<b>CHAPITRE 7 : Application de méthodes d'extraction de connaissances .....</b>	<b>203</b>
Conclusion de la partie III .....	253
Conclusion générale et perspectives de recherche .....	257
Références bibliographiques .....	263
Liste des figures .....	298
Liste des tableaux .....	300
Annexes .....	301
Table des matières .....	343



## Introduction générale

Les villes sont confrontées à de nombreuses problématiques environnementales qui les obligent à repenser les politiques urbaines. En effet, ce sont elles et leurs activités qui sont les plus grandes consommatrices de ressources naturelles, d'énergies et sont les plus émettrices de gaz à effet de serre. Mais les villes représentent aussi le plus grand levier pour limiter l'empreinte de l'homme sur l'environnement. La configuration des tissus, c'est-à-dire l'arrangement spatial des éléments urbains, a des effets directs sur la consommation en énergie des bâtiments, la formation de l'îlot de chaleur urbain, mais aussi sur le bien-être et les comportements des habitants, par exemple sur leurs modes de déplacement (Traisnel, 2001 ; Oke, 1988 ; Tzoulas *et al.*, 2007 ; Cervero et Kockelman, 1997).

De plus, les villes de l'ensemble de la planète ne cessent de croître d'un point de vue spatial et donc de repousser leurs limites même si leur population n'augmente pas voire diminue (EEA, 2006). Ce phénomène appelé étalement urbain consomme énormément de territoires agricoles et naturels et nécessite de revoir les politiques urbaines et les pratiques urbanistiques afin d'aboutir à une gestion plus économe de l'énergie et du foncier. Une solution qui a été envisagée est de densifier les villes, mais ce processus se heurte à la volonté de plus en plus forte des citoyens d'avoir des villes plus vertes (Clergeau et Hubert-Moy, 2011).

L'étalement urbain en plus de consommer des ressources foncières fragmente les paysages agricoles et naturels et donc les habitats des populations végétales et animales (EEA, 2006). Cette fragmentation est un des processus majeurs qui peut expliquer la diminution de la biodiversité en raison d'une réduction de la taille de ces habitats et de leur isolement (Fahrig, 2003). Ainsi est apparue la volonté de conserver et de maintenir une infrastructure verte à l'échelle européenne permettant de laisser circuler les espèces et endiguer cette perte de biodiversité (Jongman *et al.*, 2004). En France, cette idée d'infrastructure verte s'est traduite dans les lois Grenelle 1 (2009) et 2 (2010) par la prise en compte des Trames Verte et Bleue. Ces trames ont pour objectifs de conserver les continuités écologiques et doivent être intégrées dans les divers documents d'aménagement et d'urbanisme à tous les échelons territoriaux. Les communes des espaces urbains et périurbains sont donc aussi concernées puisque les futurs plans locaux d'urbanisme devront en tenir compte. La conservation de la biodiversité, bien que celle-ci soit ordinaire (Daniel *et al.*, 2013), et le maintien des connectivités écologiques est important dans

ces espaces, car ils permettent le déplacement d'espèces, de comprendre et d'apporter une réponse aux changements environnementaux, l'éducation à l'environnement et d'améliorer le bien-être des citoyens (Dearborn et Kark, 2010).

Avec le phénomène d'étalement urbain et avec la conservation des trames vertes et bleues, les gestionnaires des espaces urbains doivent disposer de données fiables, précises et à jour pour pouvoir identifier les enjeux et effectuer un suivi de leur territoire. Ce contexte nécessite donc de pouvoir caractériser quantitativement et qualitativement les espaces urbains afin d'effectuer un suivi régulier de ses évolutions. La cartographie et les Systèmes d'Information Géographique (SIG) sont des outils précieux d'aide à la décision (Joliveau, 2004) qui peuvent être très adaptés dans ce contexte. Toutefois, si des bases de données géographiques d'occupation et d'utilisation du sol existent et sont même relativement nombreuses, elles sont confrontées à plusieurs problèmes qui peuvent limiter leur usage dans les espaces urbains. En effet, elles nécessitent une échelle adaptée qui doit permettre d'obtenir une représentation fine de la ville, de sa structure, et doivent être mise à jour régulièrement.

Ainsi l'objectif thématique de cette thèse est de mieux comprendre comment s'articulent la trame grise (les tissus urbains et ses éléments) et la trame verte dans les espaces urbains et périurbains afin de pouvoir mieux analyser et représenter la structure urbaine et que cette caractérisation tienne compte de ces deux éléments. Pour répondre à cet objectif, il est nécessaire d'avoir des données géographiques à grande échelle (du 1/5000<sup>e</sup> au 1/10000<sup>e</sup>) qui permettent une cartographie précise de ces deux trames. Plusieurs questions peuvent ainsi être formulées auxquelles cette thèse essaiera de répondre.

Quelles données géographiques mobiliser pour la cartographie des espaces urbains ? Quelles échelles de représentation ? Sont-elles adaptées à l'analyse quantitative de la structure urbaine ? Sur quelles méthodes s'appuyer afin de cartographier les trames grises et les trames vertes ainsi que leurs éléments ? Quelles sont leurs limites et comment peuvent-elles être améliorées afin de les rendre plus opérationnelles ?

Nous formulons ici l'hypothèse que les bases de données géographiques existantes ne sont pas toutes adaptées à la caractérisation quantitative des espaces urbains, mais que les images satellites ou aériennes permettent de combler ce manque. Toutefois, la seconde hypothèse est que les méthodes existantes ne sont pas forcément adaptées ou du moins qu'elles pourraient être améliorées. Ainsi l'objectif méthodologique de cette thèse est de proposer une méthodologie adaptée à la cartographie dans un premier temps des éléments des trames, notamment les éléments verts, et de la structure urbaine à une échelle du 1/5000<sup>e</sup> au 1/10000<sup>e</sup>. L'objectif est que la méthodologie développée soit la plus automatique possible afin de la rendre opérationnelle et rapide à mettre en place afin d'être utilisée pour réaliser un suivi régulier des espaces urbains.

Afin de répondre à ce double objectif à la fois thématique et méthodologique, cette thèse s'articule en trois parties.

La partie I présente le cadre conceptuel à savoir la morphologie urbaine, mais aussi l'état actuel des connaissances en ce qui concerne les méthodes de cartographie de la structure urbaine. Le chapitre 1 permet de définir les notions et les concepts mobilisés et de présenter les différents éléments constitutifs de la structure urbaine. Les sites d'études qui sont représentatifs de nombreux espaces urbains d'Europe de l'Ouest sont ensuite présentés (chapitre 2) ainsi que les bases de données disponibles pour ces sites. Ces données y sont évaluées. L'identification des verrous scientifiques et méthodologiques conduit enfin à préciser les objectifs spécifiques et les hypothèses de travail.

La seconde partie (partie II) a pour objectif de proposer une méthodologie d'extraction de la végétation urbaine à partir d'images satellites à Très Haute Résolution Spatiale (THRS). En effet, les bases de données qui permettent de décrire la végétation ne sont pas adaptées du fait de leur échelle de représentation et de leur non-exhaustivité. Cette méthodologie se place dans le cadre d'une démarche orientée-objet et une classification supervisée par apprentissage actif. Le chapitre 3 vise définir une segmentation adaptée pour l'identification de la végétation urbaine. Avec une classification supervisée, le choix des données d'entraînement est essentiel et le chapitre 4 permet de proposer des méthodes d'apprentissage actif afin de sélectionner ces données. Les méthodes développées dans ces deux chapitres sont construites sur des sites tests et le chapitre 5 généralise ces méthodes à l'ensemble de nos sites d'études afin d'obtenir une cartographie adaptée de la végétation urbaine pouvant être utilisée pour caractériser quantitativement la structure urbaine de ces sites.

La dernière partie (partie III) doit répondre aux objectifs généraux de cette thèse. Une base de données est constituée dans le chapitre 6 permettant d'analyser quantitativement et de cartographier la structure urbaine. Cette base de données s'appuie sur des données géographiques existantes et sur les classifications des images satellites issues de la partie II. Un maillage de l'espace doit être défini et une série d'indicateurs calculés permettant de rendre compte de l'organisation des éléments entre eux complétée par d'autres indicateurs afin de caractériser quantitativement la structure urbaine. Le chapitre 7 propose une méthodologie originale de traitement pour la cartographie de la structure urbaine qui se veut la plus automatique possible et permette une représentation du 1/5000<sup>e</sup> au 1/10 000<sup>e</sup>. Cette méthode est évaluée puis validée afin de vérifier sa reproductibilité et son opérationnalité. Enfin, une conclusion à ce travail présente les différentes perspectives de recherche.



# Partie I

## Structures et formes urbaines : concepts, typologies, données disponibles et état de l'art

Introduction de la partie I.....	7
<b>CHAPITRE 1</b> : Structure spatiale urbaine et morphologie urbaine : définitions et concepts.....	9
<b>CHAPITRE 2</b> : Données géographiques et méthodes existantes pour la cartographie de la structure urbaine.....	37
Conclusion de la partie I et positionnement de recherche.....	71





## Introduction de la partie I

Cette partie a pour objectif de présenter le cadre thématique, de définir les notions et concepts essentiels à ce travail et de préciser ses objectifs spécifiques.

Le chapitre 1 nous permet de définir notre objet d'étude, la forme urbaine, et de détailler l'ensemble des éléments qui constituent sa structure : la trame grise et la trame verte. Nous précisons dans ce chapitre ces deux trames et les évolutions dans leurs conceptions et leurs considérations qui ont créé en ville une diversité de formes importantes. Il présente aussi les enjeux et les intérêts d'étudier ces deux trames et ainsi la nécessité d'obtenir une cartographie à grande échelle de cette structure.

Le second chapitre de cette partie présente les sites d'étude qui sont retenus et les données qui peuvent être mobilisées afin de cartographier les trames grises et vertes urbaines. Ces données qui sont géographiques y sont définies plus en détail. Cette analyse met en lumière le manque d'information sur la ville à l'échelle où nous souhaitons réaliser l'analyse de la structure urbaine (du 1/5000<sup>e</sup> au 1/10 000<sup>e</sup>). Ainsi ce chapitre montre les différentes méthodes existantes dans la littérature pour la cartographie de ces trames. Il permet aussi d'en identifier les verrous afin de rendre ces méthodes opérationnelles et de préciser les objectifs spécifiques et les hypothèses de cette thèse.



# Chapitre 1

## Structure spatiale urbaine et morphologie urbaine : définitions et concepts

### Sommaire

---

Introduction . . . . .	10
1.1 Formes urbaines et morphologie urbaine . . . . .	10
1.2 La trame grise . . . . .	11
1.2.1 La notion de tissu urbain . . . . .	12
1.2.2 Intérêts de l'étude des tissus urbains . . . . .	16
1.2.3 Typologies grises . . . . .	20
1.3 La trame verte . . . . .	25
1.3.1 La prise en compte de la végétation en urbanisme . . . . .	25
1.3.2 La notion de trame verte . . . . .	27
1.3.3 Multifonctionnalité de la végétation urbaine . . . . .	30
1.3.4 Typologies vertes . . . . .	32
Conclusion . . . . .	34

---

## Introduction

L'objectif de ce chapitre est de définir les différentes notions mobilisées. Ainsi la section 1.1 nous permet de montrer la polysémie de la notion de forme urbaine et de son analyse, la morphologie urbaine. Nous y précisons la définition que nous avons retenue et notre objet d'étude. Nous détaillons ensuite dans la section 1.2 le premier élément de la structure urbaine à savoir la trame grise. Cette section nous permet de mieux définir la notion de tissus urbains ainsi que l'intérêt d'étudier la ville à cette échelle et les typologies utilisées pour les qualifier. La section 1.3 revient elle sur le deuxième élément que nous intégrons à la structure urbaine : la trame verte. Cette section présente la place et l'évolution qu'a connue cette trame dans les espaces urbains et une partie des fonctions qui lui sont associées.

### 1.1 Formes urbaines et morphologie urbaine

La morphologie urbaine se réfère à deux acceptations. La première est un synonyme de la forme urbaine tandis que la seconde fait référence à l'étude de ces formes (Merlin et Choay, 1988). Nous retenons la deuxième acceptation dans la suite de cette thèse. Ainsi pour Allain (2004), la morphologie urbaine se définit comme « *l'étude de la forme physique de la ville, de la constitution progressive de son tissu urbain et des rapports réciproques des éléments de ce tissu qui définissent des combinaisons particulières, des figures urbaines (rues, places et autres espaces publics...).* Cette réalité complexe analysable à différents niveaux d'échelle et de plusieurs points de vue, traverse les cloisonnements disciplinaires ». Cette définition insiste sur le caractère pluridisciplinaire de l'étude des formes urbaines où se retrouvent aussi bien l'architecture, l'urbanisme, l'histoire, la géographie, l'économie, la climatologie, la sociologie, etc. Cette multitude d'approches entraîne de fait une multitude de définitions pour la notion de forme urbaine.

Elle est une notion polysémique et ambiguë et peut être synonyme, selon les disciplines ou les acteurs qui l'utilisent, de forme sociale urbaine, représentation urbaine, modèle urbain, type urbain, plan urbain, composition urbaine ou tissu urbain (Raynaud, 1999). Lévy (2005) voit cinq grandes approches qui utilisent cette notion et qui influent sur sa définition :

- La forme urbaine comme forme du paysage urbain qui étudie l'espace urbain par la perception visuelle, son esthétisme, sa dimension culturelle.
- La forme urbaine comme forme sociale qui s'intéresse particulièrement à la répartition des groupes sociaux dans les espaces urbains. Les analyses fonctionnelles de la ville dont l'objectif est d'étudier l'utilisation du sol, c'est-à-dire la répartition des activités (économique, résidentielle, récréative, etc.), peuvent être intégrées à cette approche
- La forme urbaine comme forme bioclimatique où c'est la dimension environnementale de la ville qui est au centre de l'étude. Plusieurs disciplines sont ici concernées

telles que la climatologie urbaine (îlot de chaleur urbain, pollution atmosphérique) ou l'écologie urbaine.

- La forme urbaine comme forme des tissus urbains qui s'intéresse à l'organisation et les interrelations entre les différents éléments tels que le réseau viaire, les bâtiments et le parcellaire.
- La forme urbaine comme forme des tracés où la géométrie du plan de la ville est au centre de l'étude.

Pour cet auteur, l'ensemble de ces approches amène à considérer la forme urbaine comme une notion polymorphique et polysémique dont les différents registres sont interdépendants. La forme urbaine est donc un objet complexe qui prend un caractère systémique dont l'ensemble des registres peut être pris en compte simultanément afin d'obtenir une « forme unitaire globale ».

La définition de Allain (2004) met aussi l'accent sur les échelles d'analyse de la forme urbaine. Ces échelles sont très diverses et les formes urbaines peuvent ainsi être analysées à micro échelle, celle des bâtiments et de leur interrelation (APUR, 2007), à méso échelle où l'organisation des éléments au niveau du bloc urbain ou du quartier sont considérés (Puissant *et al.*, 2011 ; Tran *et al.*, 2014) ou à macro échelle, celle de la ville ou de l'agglomération (Tsai, 2005 ; Le Néchet, 2015). Toutefois, la complexité même de la ville fait qu'il est difficile de se concentrer uniquement sur un niveau d'analyse et adopter une approche multi échelle est nécessaire pour comprendre son fonctionnement (Smith et Crooks, 2010 ; Le Néchet et Aguiléra, 2012).

Dans les analyses multi échelle de la ville, une approche traditionnelle consiste à commencer par le tout pour ensuite voir les parties et enfin revenir au tout (Salat, 2011). Toutefois pour mieux saisir la ville et sa complexité, l'idéal pour Salat est de commencer par les parties puis de voir leurs relations. Pour cet auteur, l'étude de la ville par les tissus urbains est spécifiquement adaptée puisqu'il s'agit d'une échelle d'analyse qui se situe entre celle « *de l'architecture des bâtiments et les grands tracés de l'urbanisme, qui est le cadre de la vie quotidienne* ». De plus pour reprendre les différents registres de formes énoncées par Lévy (2005), nous considérons ici la forme urbaine comme celle des tissus urbains, que nous appelons la trame grise, en y intégrant leur forme bioclimatique dans le sens où les trames vertes et bleues y sont intégrées. La combinaison de ces différentes trames constitue ce que nous appelons la structure urbaine (figure 1.1).

## 1.2 La trame grise

Cette section vise à définir plus spécifiquement les éléments de la trame grise urbaine. La notion de tissu urbain (§ 1.2.1), les éléments qui le composent (§ 1.2.1.1) sont tout d'abord présentés ainsi que les grands courants urbanistiques qui au cours du temps ont produit des formes urbaines différenciées (§ 1.2.1.2). Nous montrons dans la section 1.2.2 l'influence que peut avoir la forme urbaine sur l'environnement urbain

et sur les comportements des habitants afin de montrer l'intérêt que représente son analyse. Enfin nous revenons sur les typologies des tissus utilisées dans la littérature, car ils offrent des lectures différentes de la trame grise et de la ville (§ 1.2.3).

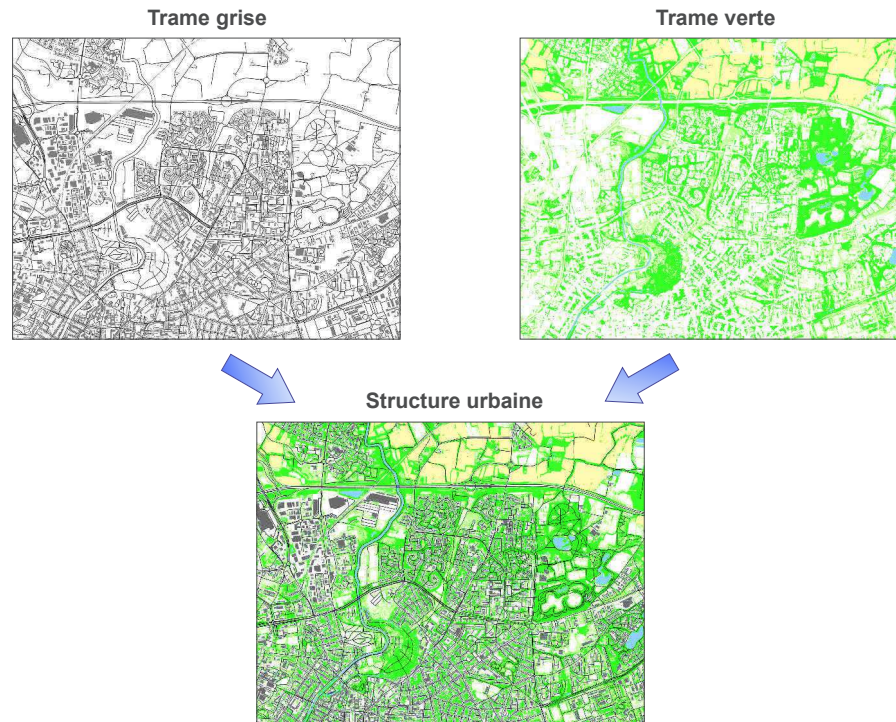


Figure 1.1 – Trames et structure urbaines (sources : BD TOPO® – 2015 ; RPG 2012)

### 1.2.1 La notion de tissu urbain

La notion de tissu urbain est une « métaphore qui fait référence au tissage – le textile –, ou à la biologie – les tissus végétaux, osseux. Le terme de tissu urbain entraîne une double acceptation. Il s'agit d'une vision locale qui « oublie » momentanément l'organisation de l'ensemble, l'armature, le squelette, pour s'intéresser au remplissage, à la substance. Il s'agit d'une organisation qui présente à la fois une forte solidarité entre les éléments et une capacité à s'adapter, à se modifier, à se transformer. Appliqué à la ville, le terme de tissu évoque la continuité et le renouvellement, la permanence et la variation » (Panerai et al., 1999).

Pour Merlin et Choay (1988) le tissu est « l'expression physique de la forme urbaine. Il est constitué par l'ensemble des éléments physiques qui contribuent à celle-ci – le site, le réseau viaire, la division parcellaire, le rapport entre les espaces bâtis et non bâtis, la dimension, la forme et le style des bâtiments – et par des rapports qui relient ces éléments. [...] La notion de tissu urbain et donc à la fois statique (état des formes urbaines à un moment donné) et dynamique (porteuse d'évolution de ces formes urbaines) ».

L'étude des tissus est, comme nous l'avons mentionné précédemment, l'analyse de l'organisation de ses éléments constitutifs ainsi que des interrelations qui existe entre eux. Un tissu urbain est ainsi une portion de ville « homogène » selon les caractéristiques des différents éléments qui le composent et qui peuvent être, eux, hétérogènes. On peut ainsi parler par exemple de tissus haussmanniens, ouverts ou fermés, de grands ensembles ou pavillonnaires. Ce sont donc les formes, mais aussi les différentes organisations possibles entre ces éléments qui définissent différents types de tissus.

### 1.2.1.1 Éléments constitutifs des tissus

Les tissus urbains sont le résultat de l'articulation de trois éléments : les réseaux de voies, les découpages fonciers et les bâtiments (Panerai *et al.*, 1999 – figure 1.2). Le réseau viaire assure la desserte de la ville par une structure hiérarchisée de voies allant de la ruelle pour assurer la desserte locale en passant par les boulevards et les avenues pour les liaisons intra-urbaines, aux autoroutes pour celles entre les villes. Ce réseau constitue, d'un point de vue global, l'armature de la forme urbaine. D'un point de vue plus local c'est lui qui permet l'organisation du tissu urbain. La voie est aussi importante du fait qu'elle constitue l'espace public, celui pratiqué par les habitants, et s'oppose donc à l'espace privé qui est matérialisé par les découpages fonciers. Les voies permettent de donner beaucoup d'indications sur les tissus et sont les éléments les plus permanents dans la ville du fait que les constructions se maintiennent sur leur axe de développement et conservent ainsi leur emplacement (Salat, 2011).

Le négatif du réseau viaire constitue les parties privées de la ville qui sont découpées par un maillage parcellaire qui est aussi d'une très grande importance pour la compréhension des tissus et de leur constitution. En effet, ces découpages portent une marque historique importante qui peut exprimer des structures socio-économiques anciennes. Les relations de ce maillage avec la rue sont ce qui permet la constitution du tissu urbain (Panerai *et al.*, 1999). Il est important de distinguer parcelle cadastrale et parcelle d'usage (Allain, 2004). La parcelle cadastrale est un découpage du sol utilisé pour son appropriation. C'est aussi la maille élémentaire de l'urbanisme réglementaire permettant de calculer le coefficient d'emprise au sol des bâtiments. Une parcelle cadastrale appartient à un propriétaire, mais ne correspond pas nécessairement à la réalité architecturale. Un bâtiment peut être sur plusieurs parcelles qui ainsi forment une parcelle d'usage. Un ensemble contigu de parcelles cadastrales qui se trouvent donc délimitées par les voies de circulation constitue un îlot. C'est une unité très souvent utilisée pour la constitution des tissus. À un niveau supérieur, le quartier est un regroupement d'îlots similaire d'un point de vue sémantique qui peut lui aussi être utilisé pour décrire les formes urbaines.

Enfin le dernier élément constitutif du tissu urbain est le bâtiment. Son architecture, sa position par rapport à la rue, aligné ou en retrait, permettent de constituer des tissus urbains homogènes. Pour Panerai *et al.* (1999), c'est la relation des bâtiments à la rue qui

permet d'obtenir une solidarité entre eux bien qu'ils ne soient pas forcément de même hauteur ou de la même époque et qui permet de constituer des tissus homogènes.

Les différentes configurations possibles entre ces trois éléments permettent une diversité très importante de formes urbaines. Les formes de tissus urbains ont fortement évolué au cours du temps du fait de l'évolution des croyances, des conceptions philosophiques, esthétiques et techniques qui combiné avec l'inertie de ces formes offrent aujourd'hui dans les villes une grande variété de tissus.



**Figure 1.2** – Les trois éléments constitutifs des tissus urbains : les voies (a), le découpage foncier (b) et les bâtiments (c) (source : BD Parcellaire® – 2015) (d'après Panerai *et al.*, 1999)

### 1.2.1.2 Formes des tissus, évolutions et tailles

Allain (2004) distingue six grands héritages urbanistiques. Le premier forme la ville médiévale ou traditionnelle. Celle-ci n'est pas liée à de grands projets urbanistiques imposés par un pouvoir, mais est le fait des habitants où la diversité des constructions réalisées sur des micro parcelles est très importante. Ce type de tissu est en général peu homogène, la densité de bâtiments est très importante et les façades sont accolées sur



les rues qui sont en général étroites et non rectilignes. Le terme de croissance organique est souvent utilisé pour décrire la constitution de ces tissus.

Le second héritage urbanistique identifié par Allain est l'idéal urbain classique. Cet idéal apparaît pendant la Renaissance et la ville devient ici considérée comme une œuvre d'art. Une place très importante est ainsi donnée aux monuments et aux perspectives.

La ville libérale apparaît avec la révolution industrielle dans un contexte de forte croissance urbaine, de révolution technologique et de forte mobilisation du capital. Haussmann et Cerda sont les urbanistes les plus représentatifs de ce courant. C'est le début de l'urbanisme réglementaire et technicien qui avait pour objectif de répondre aux problèmes d'hygiène et de ségrégation urbaine tout en conservant une dimension esthétique héritée de la ville classique importante.

Le quatrième courant urbanistique est celui des cités-jardins proposées par Howard. C'est ici une réponse à l'urbanisation importante des villes et aux pollutions liées à leur industrialisation. L'objectif annoncé est de proposer une ville avec une densité assez faible permettant de tirer avantage des bénéfices de la ville et de ceux de la campagne sans leurs désavantages. Dans leur conception, une place très importante est donnée aux espaces verts ainsi qu'à la maîtrise du foncier et ce courant influencera beaucoup par la suite la conception des villes nouvelles. Pour Boutefeu *et al.* (2008) les écoquartiers qui sont aujourd'hui de plus en plus plébiscités empruntent beaucoup à ce courant urbanistique.

Vient ensuite le courant que Allain (2004) désigne par la ville fordiste et fonctionnaliste. Les formes urbaines sont construites de manière industrielle, répétitive et en grandes séries à partir de la seconde moitié du 20<sup>e</sup> siècle. Ce courant, dont Le Corbusier est le plus célèbre représentant, vise à sectoriser la ville selon quatre fonctions : l'habitation, les loisirs, le travail et la circulation. Ses principes vont être rassemblés dans la Charte d'Athènes de 1941. En France, ce courant a été très largement mis en œuvre au lendemain de la Seconde Guerre Mondiale afin de répondre à la crise du logement et il donnera naissance à la politique des grands ensembles entre les années 50 et 70 qui se caractérise par la construction de barre et de tours. Il a été très critiqué par les urbanistes considérant cette vision comme destructrice des tissus urbains, ayant uniformisé et banalisé les formes, produisant de la non-ville (Salat, 2011).

C'est ainsi qu'est née la réaction postmoderne qui est le dernier courant mentionné par Allain (2004). Ce courant n'en est pas un au sens strict du terme puisqu'il est très divers, mais insiste sur le côté destructeur du fonctionnalisme sur la texture et l'effacement du sens des villes. Cet urbanisme vise à mieux intégrer les tissus urbains dans le contexte existant en conciliant les vertus des tissus urbains traditionnels et les contraintes de la modernité. Toutefois si ce courant prend une importance de plus en plus grande, la croissance des villes se fait encore pour l'essentiel de manière sérielle du fait que se sont principalement les promoteurs immobiliers, les géomètres et les spécialistes du marketing (urbanistes commerciaux) plutôt que les urbanistes qui aujourd'hui construisent les villes (Mangin, 2004).

Une autre forme d'urbanisation non mentionnée par Allain (2004) est constituée par le développement très important de l'habitat pavillonnaire et de la maison individuelle associé à celui des grandes zones commerciales en périphérie des villes (Mangin, 2004). Le développement de l'habitat pavillonnaire est pourtant un des facteurs les plus importants de consommation d'espace agricole et naturel en Europe (EEA, 2006). Il répond à un rejet des grands ensembles favorisé en France par le développement important des infrastructures de transport routier (autoroute, rocade) à partir des années 70 et une augmentation du taux de motorisation des ménages (Mangin, 2004). Cette période marque aussi le désengagement de l'État du secteur locatif et le début des politiques d'accès à la propriété de maisons individuelles. Favorisées par la création de voies rapides, par l'installation de cette nouvelle population éloignée des centres et par un foncier peu cher, les grandes zones commerciales ou d'activités se sont encore davantage développées dans ces périphéries.

L'ensemble de ces différents modes de construction des villes au cours du temps s'est fait sur des îlots de tailles très différentes (Allain, 2004). De manière générale, le maillage de la ville est plus fin (des îlots plus petits) dans les espaces centraux et devient plus lâche dans les espaces périphériques. En moyenne la taille de la maille est ainsi de 250 à 500 m par 250 m de côté dans les tissus les plus éloignés du centre à 250 m sur 100 ou 120 m en zone péricentrale pour arriver à des moyennes de 120 m sur 60 ou 60 sur 30 dans les centres urbains (figure 1.3). Toutefois de fortes différences peuvent exister notamment dans les centres-ville ou les quartiers péricentraux où de grands îlots peuvent être présents tels que pour des monuments, des parcs ou des services publics (lycée, hôpital, etc.).

L'étude des formes urbaines a connu un ralentissement à partir des années 70 (Lévy, 2005). Toutefois les nouvelles préoccupations environnementales, la problématique du développement durable nécessitent de repenser les formes urbaines pour créer les villes de demain. Ainsi l'étude du tissu urbain et ses implications socio-environnementales connaissent aujourd'hui un regain d'intérêt.

### **1.2.2 Intérêts de l'étude des tissus urbains**

La ville est actuellement confrontée à des défis environnementaux majeurs. En effet, les problématiques de la maîtrise de l'étalement urbain, la sauvegarde des territoires agricoles, la consommation d'énergie, la problématique des gaz à effet de serre et la maîtrise des pollutions urbaines sont des défis majeurs que le développement urbain doit intégrer.

Les activités urbaines sont les principales consommatrices d'énergie en France. En 2014 le secteur résidentiel représentait 30 % du total d'énergie consommée, le secteur tertiaire 15 % et les transports 33 % (figure 1.4 – Commissariat général au développement durable, 2015a).



**Figure 1.3** – Exemple de tissus urbains caractéristiques (source : base de données référentiel 2000, Eurométropole de Strasbourg)

Il en va de même pour les émissions de gaz à effet de serre. La première source d'émission en France en 2014 est le transport qui en représente 28 % suivie des bâtiments résidentiels, tertiaires, commerciaux et institutionnels dont la part est de 20 % (figure 1.5 – Commissariat général au développement durable, 2015b). Ce contexte associé aux problématiques de réchauffement climatique nécessite donc de penser les villes de manière à ce qu'elle consomme globalement moins d'énergie tant au niveau du bâtiment qu'au niveau des transports.

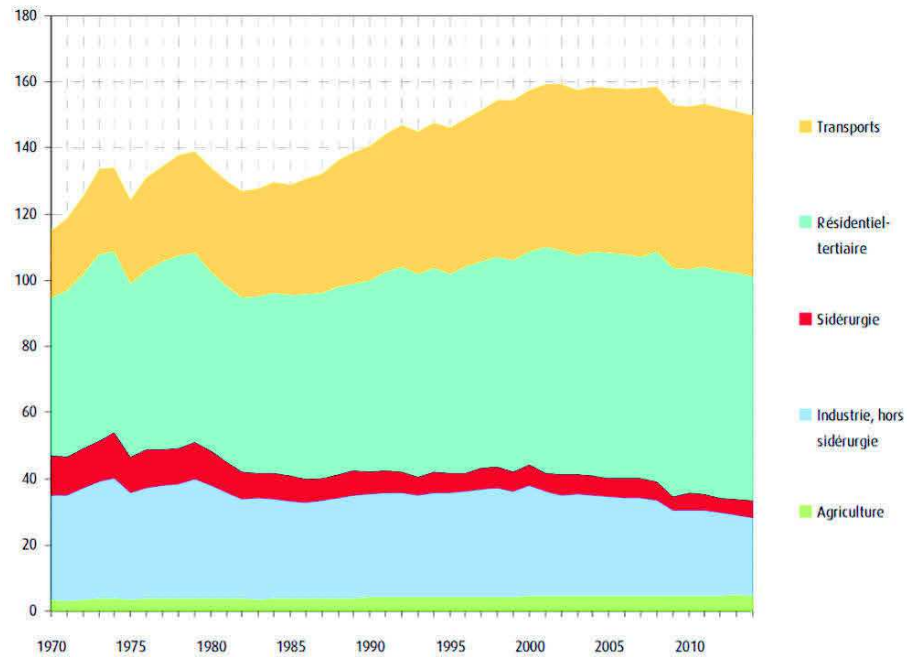
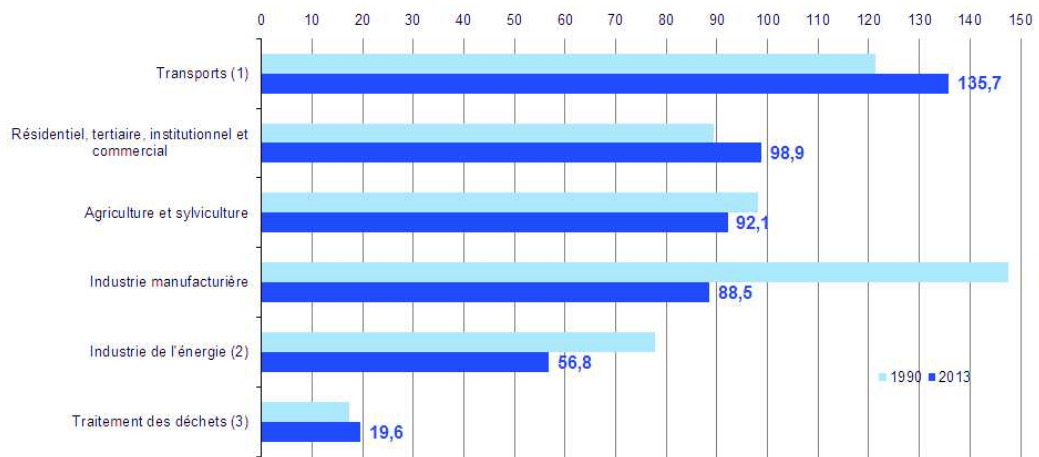


Figure 1.4 – Évolution de la consommation finale énergétique en France par secteur en millions de tonnes équivalents pétrole (de Commissariat général au développement durable (2015a))



Notes : hors UTCF (utilisation des terres, leurs changements et la forêt) ; (1) aérien et maritime : trafic domestique uniquement ; (2) y compris incinération des déchets avec récupération d'énergie ; (3) hors incinération des déchets avec récupération d'énergie, et hors captage de biogaz.  
 Champ : France métropolitaine, départements d'Outre-mer, Saint Martin (périmètre Protocole de Kyoto).  
 Source : Citepa (inventaire CCNUCC, format "Plan Climat"), juin 2015.

Figure 1.5 – Émissions de gaz à effet de serre en France en millions de tonnes équivalents CO2 (de Commissariat général au développement durable (2015b))

De nombreuses recherches ont ainsi eu pour objectif de comprendre le lien qui peut exister entre la mobilité et les formes urbaines. Ce sont principalement des facteurs socio-économiques qui jouent sur les modes de transport (Schwanen *et al.*, 2004 ; Pouyanne, 2004), mais l'impact des formes n'en est pas moins négligeable. En général elles montrent que des formes urbaines denses favorisent les mobilités douces (marche à pied, vélo) et les transports en commun et diminuent ainsi le taux de motorisation des ménages et réduisent la distance de déplacement (Cervero et Kockelman, 1997 ; Lin et Yang, 2009 ; Tran *et al.*, 2014). C'est particulièrement vrai dans les centres-ville anciens qui n'ont pas été conçus pour les modes de transport routier (Van de Coevering et Schwanen, 2006). De plus la diversité des fonctions urbaines apparaît être un facteur déterminant sur les modes de mobilités (Cervero et Kockelman, 1997 ; Lin et Yang, 2009). Toutefois ces études sont souvent réalisées à des échelles très locales, celles de l'îlot ou du quartier, mais des études plus globales manquent pour mieux comprendre les effets de la forme urbaine sur les comportements de mobilité (Le Néchet et Aguiléra, 2012) et intègrent surtout les trajets quotidiens en oubliant ceux plus longs et moins fréquents pour les loisirs par exemple. Certains chercheurs ont aussi montré que la construction de ville dense pouvait entraîner une congestion plus importante des axes de transports et donc une pollution plus importante, une diminution d'accès aux espaces verts naturels et des coûts de logement plus élevé pour les habitants (Breheny, 1997 ; Burton, 2000). De plus, certaines formes urbaines telles que les maisons individuelles pourraient diminuer le nombre de déplacements du fait de l'accès à un jardin (Holden et Norland, 2005).

D'autres études tentent de trouver le lien entre forme urbaine et consommation des bâtiments. Tout comme pour la mobilité, beaucoup d'études arrivent à la conclusion que la densité de bâti, la compacité des bâtiments et leur contiguïté permet de réduire les consommations (Adolphe, 2001 ; Traisnel, 2001). Mais la qualité des matériaux de construction apparaît être un facteur tout aussi important permettant d'expliquer ces consommations (Traisnel, 2001 ; APUR, 2007). Ces recherches se concentrent principalement sur les besoins en chauffage des bâtiments et uniquement sur la consommation d'énergie. Les besoins en climatisation apparaissent aussi diminués dans le cas de ville dense du fait d'une moins grande surface extérieure amenant donc moins d'échange thermique avec l'extérieur (Hui, 2001). À l'inverse, moins d'échange avec l'extérieur ne permet pas de profiter des apports solaires directs pour le chauffage (Robinson et Stone, 2004) et l'ombre portée des bâtiments dans la ville dense peut aussi entraîner une augmentation des besoins en électricité notamment pour la lumière (Hui, 2001).

Les formes urbaines peuvent aussi avoir un impact important sur la formation de l'îlot de chaleur urbain. Il s'agit d'un microclimat en milieu urbain qui entraîne une augmentation de la température de l'air par rapport aux espaces naturels ou agricoles à proximité. Plusieurs facteurs sont en causes telles que les activités urbaines, les matériaux, etc. Ces effets montrent des différences de températures pouvant atteindre 12 °C principalement la nuit, car la température diminue moins en ville du fait de l'absorption de chaleur plus importante par les matériaux urbains et qui se trouve dégagée la nuit (Oke, 1988). La ville dense est la plus touchée par ce phénomène

microclimatique, car elle est généralement plus minérale avec moins de végétation. De plus ces formes perturbent et peuvent limiter la circulation de l'air au niveau des rues ce qui est aussi un facteur aggravant de l'îlot de chaleur urbain (Oke, 1988), mais peut aussi entraîner une stagnation des pollutions atmosphériques dans ces tissus (Weber *et al.*, 2014). Combiné avec une augmentation globale des températures, cet effet pourrait être encore plus important à l'avenir et avoir un impact sur la santé (Sakhy *et al.*, 2011) et sur les consommations électriques nécessaire aux climatisations (Masson *et al.*, 2014).

La ville apparaît ici comme un objet particulièrement complexe et il est donc difficile de proposer des formes urbaines optimales. En effet, chaque forme de développement a ses avantages permettant ainsi d'optimiser certains paramètres qui s'accompagne nécessairement d'effet négatif sur les autres (Bonhomme, 2013 ; Martilli, 2014).

Bien qu'implicite, l'exposé sur les tissus urbains réalisés jusqu'ici utilisait la notion de type ou de typologie par exemple lorsque la ville dense, les différentes époques de construction, etc. sont mentionnées. La prochaine section vise donc à définir les notions de types et de typologie et montrer une partie de la diversité qu'il est possible de considérer pour décrire la ville et ses tissus.

### 1.2.3 Typologies grises

Pour Salat (2011) les « *notions de type et de typologie forment un des outils majeurs de l'analyse urbaine. De même que la systématique en botanique ou en minéralogie précède toute la réflexion sur les formations végétales ou sur les structures géologiques, la typologie des plans, des parcelles et surtout du bâti permet de se dégager de la collection d'objets isolés pour en opérer un classement requérant un certain niveau d'abstraction. Ce classement permet de rendre compte des objets complexes que sont les tissus urbains. La typologie classique commence par la description des éléments (maillages, parcellaires, maisons et immeubles) puis par leur classification en fonction de certains critères jugés pertinents* ». L'objectif de ce classement en type est donc de mettre en ordre afin de réduire la complexité du problème étudié, mais cela nécessite au préalable de faire le choix de certains indicateurs pour construire cette typologie (Coenen-Huther, 2007).

Pour l'analyse des tissus urbains, de nombreuses typologies ont ainsi été construites. Selon les objectifs de l'étude, ces typologies peuvent être plus ou moins complexes c'est-à-dire avec plus ou moins de types différents et utiliser plus ou moins d'indicateurs pour les décrire. L'APUR (2007) par exemple construit une typologie en neuf classes des îlots de la ville de Paris basé sur un seul indicateur. La date de construction des bâtiments à partir de relevé de terrain est utilisée et les neuf types correspondent à une période de construction. L'objectif de cette étude est d'estimer les consommations en énergie de bâtiments et cette typologie permet donc d'obtenir une indication sur les matériaux utilisés pour leur construction. La densité est aussi un indicateur important et très souvent utilisé, notamment par les agences d'urbanisme, pour réaliser une typologie des tissus urbains. De nombreuses mesures de densité bâtie peuvent être calculées selon

que sont pris en compte ou non les espaces publics, la surface au sol, de plancher ou de logement d'un bâtiment (Salat, 2011). L'AUCAME (2010) utilise ainsi trois indicateurs de densité : la densité de logement, le coefficient d'occupation du sol (surface de plancher / surface de l'îlot) et le coefficient d'emprise au sol (surface au sol / surface de l'îlot) qui, combinés avec la part d'espaces publics, permettent de décrire les tissus urbains types de l'agglomération caennaise. Elle analyse avec ces indicateurs une typologie composée de six grands types correspondant à des époques de construction chacun se trouvant lui-même composé de deux types différents liés à leur morphologie.

Badariotti (2005) propose une typologie des tissus urbains de Strasbourg basée sur l'époque de leur construction : le tissu dense du centre-ville, les tissus de la fin du 19<sup>e</sup> et du début de 20<sup>e</sup> siècle composés du type haussmannien, du pavillonnaire et du tissu industriel d'avant-guerre, et enfin les tissus d'après-guerre composés des grands ensembles, des zones industrielles et commerciales et des lotissements. Son objectif est de vérifier si les mesures de fractalité proposés par Frankhauser (1990) peuvent rendre compte des différences et des ressemblances entre les types de tissus sélectionnés.

Dans le cadre de recherche en climatologie urbaine, certains auteurs ont proposé de construire des typologies simplifiées afin de pouvoir mieux étudier par modélisation les apports que représentent certains tissus sur la consommation énergétique des bâtiments. Arantes *et al.* (2016) par exemple comparent six configurations de tissus (figure 1.6) et leur assemblage au sein de quartier avec pour objectif d'optimiser la densité de population, l'accès à l'ensoleillement et la performance énergétique.

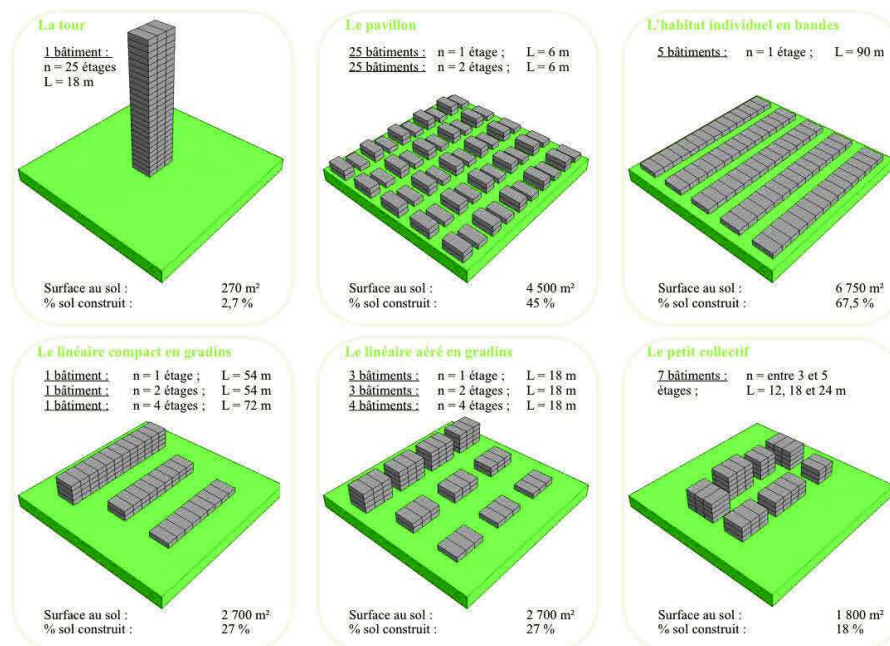


Figure 1.6 – Exemple de morphologie simplifiée pour l'analyse de la consommation des bâtiments (de Arantes *et al.*, 2016)

Toujours pour des applications en climatologie urbaine, Stewart et Oke (2012) proposent une typologie composée de dix types pour les espaces urbains et de sept pour les espaces agricoles et naturels (figure 1.7), les « *Local Climate Zones* » (LCZ). Leur objectif est que ces LCZ permettent d'étudier les effets et l'intensité des îlots de chaleur urbains en utilisant une typologie unifiée afin de rendre possible la comparaison entre différents espaces urbains. La construction de ces LCZ repose ainsi sur une quantité limitée d'indicateurs afin de pouvoir être facilement reproductibles, ici dix basés principalement sur la morphologie des tissus, leur composition, mais aussi leur albédo et l'utilisation humaine d'énergie. Cette typologie connaît actuellement un succès important, tant pour l'analyse des climats urbains que pour la cartographie des tissus, comme en témoigne le nombre d'études qui l'utilise. Ce succès peut s'expliquer par son caractère générique, facilement compréhensible et culturellement neutre (Bechtel *et al.*, 2015).

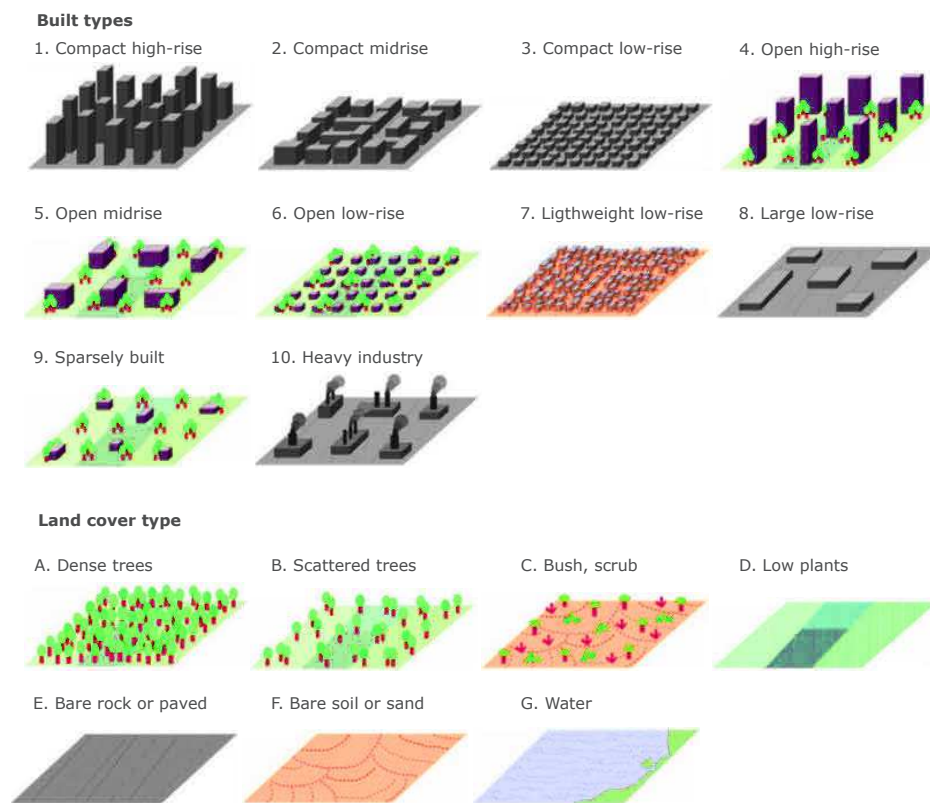


Figure 1.7 – Les *Local Climate Zones* proposées par Stewart et Oke (2012)

D'autres auteurs ont aussi proposé des typologies qui peuvent être appliquées sur l'ensemble des tissus urbains. Berghauser Pont et Haupt (2005) proposent ainsi une typologie en huit classes qui peuvent être obtenues simplement à partir de quatre indicateurs et de leur projection sur la *Spacemate* (figure 1.8). Ces quatre indicateurs sont la densité de plancher (*FSI*), la densité de surface au sol du bâti (*GSI*), le ratio d'espace ouvert (*OSR*) et le nombre moyen d'étages (*L*).



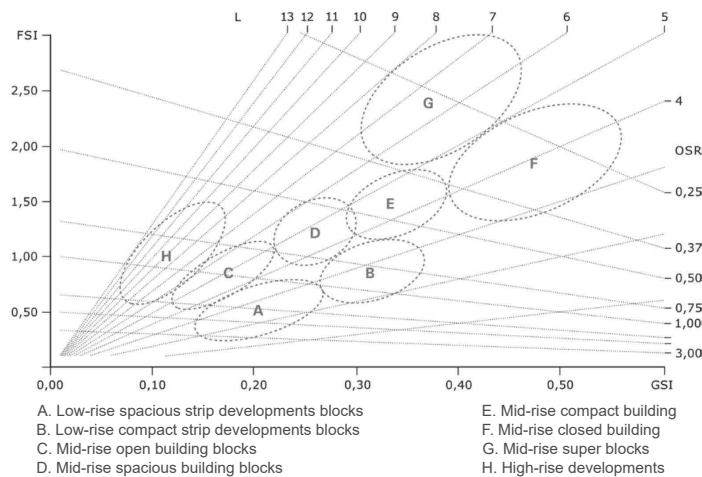


Figure 1.8 – La Spacemate de Berghauer Pont et Haupt (2005)

Avec des objectifs de cartographie des tissus urbains, de nombreuses typologies différentes ont été proposées. Une typologie très souvent considérée pour cartographier les structures urbaines utilise la densité comme critère principal. Steiniger *et al.* (2008) proposent ainsi une typologie en cinq classes à partir de la forme des bâtiments : centre urbain, urbain, suburbain, rural et industriel. Rousseaux *et al.* (2012) en formulent une similaire pour la cartographie des tissus de Nantes composée de six classes : centre-ville, péricentre, résidentiel dense et peu dense, habitat isolé et zone industrielle, à laquelle Bernabé *et al.* (2013) sur ce même site d'étude ajoutent le type grand ensemble. Toujours pour la cartographie des tissus, Puissant *et al.* (2011) proposent une typologie assez simple permettant de prendre en compte la diversité des formes de Strasbourg (tableau 1.1). Cette typologie est régulièrement utilisée comme dans les recherches de Walde *et al.* (2014) ou de Baro *et al.* (2016). Toutefois, elle ne considère pas la diversité des formes des habitats collectifs et individuels.

Tableau 1.1 – Typologie proposée par Puissant *et al.* (2011) pour la cartographie des îlots au 1/10000<sup>e</sup>

Type d'îlot 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
Mélange des classes 2 (pavillonnaire) et 3 (grand ensemble)
5. Tissu urbain discontinu mixte
Mélange de classes d'habitat (classe 1, 2, 3) et de la classe 6 (emprises spécialisées bâties)
6. Emprises spécialisées bâties
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)
8. Réseau de communication
9. Réseau hydrographique

D'autres auteurs proposent des typologies plus diversifiées pour ces types prenant en compte, pour l'habitat individuel, la densité (Herold *et al.*, 2003b) ou la contiguïté des bâtiments (Hermosilla *et al.*, 2012) et, pour l'habitat collectif, la dimension des immeubles (Long et Kergomard, 2005 ; Sahraoui *et al.*, 2014). Meinel *et al.* (2009) proposent une typologie en dix classes permettant de mieux rendre compte de la diversité de forme des tissus résidentiels (figure 1.9). Cette typologie a souvent été reprise et adaptée à la configuration urbaine du site étudié faisant ainsi varier le nombre de types de tissus urbains (Hussain *et al.*, 2012 ; Bonhomme, 2013 ; Hecht *et al.*, 2015).

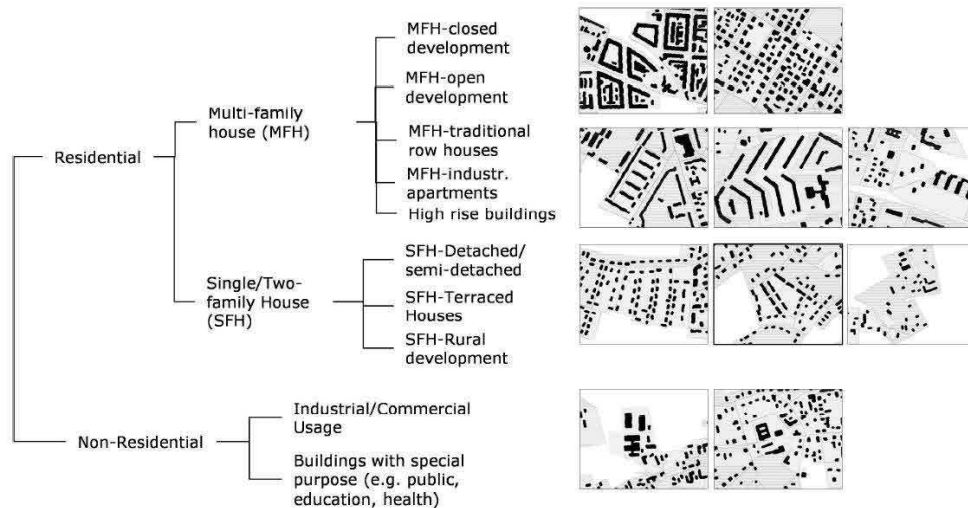


Figure 1.9 – Typologie des tissus urbains proposée par Meinel *et al.* (2009) (de Hecht *et al.*, 2013a)

Les typologies présentées jusqu'à présent se sont surtout attachées à décrire les tissus urbains d'un point de vue morphologique. Il existe toutefois de nombreuses typologies pour la cartographie des espaces urbains qui tentent, à partir de ces formes, de déduire les fonctions de la ville (Zhan *et al.*, 2000 ; Bauer et Steinnocher, 2001 ; Hu et Wang, 2013). Wu *et al.* (2007) en propose une en neuf types afin de cartographier la ville d'Austin au Texas (habitat individuel et collectif, commercial, industriel, bureau, bâtiments publics, espaces ouverts, transport et espace non développé). Ce type de classification permettant de cartographier la fonction des tissus est très demandé par les gestionnaires de ces espaces pour la planification urbaine. Toutefois le lien entre la structure des tissus urbains et leur fonction socio-économique n'est pas direct et peut parfois être injustifié du fait de la mixité des fonctions que peut remplir une portion de ville homogène d'un point de vue morphologique (Van de Voorde *et al.*, 2011).

Loin de proposer une vision exhaustive des typologies utilisées pour décrire la trame grise, cette section a permis d'en montrer toute la diversité. Les typologies sont construites selon plusieurs critères. Elles nécessitent d'être adaptées aux sites d'études et à l'application qui en sera faite, mais aussi aux données et aux méthodes utilisées pour la

cartographie des tissus urbains (chapitre 2). Peu de ces typologies considèrent explicitement les éléments « naturels », végétation et surface en eau, dans la description des espaces urbains. Ces éléments sont aussi très peu utilisés dans la construction des tissus comme le met en lumière le chapitre 2. En effet dans l'étude des tissus, une dichotomie entre espaces construits et non construits est souvent faite, mais pose ainsi la question de la composition de ces espaces ouverts. Pourtant ces éléments permettent aussi de structurer l'espace urbain et peuvent être caractéristiques de certains tissus, notamment la végétation qui peut être très différente selon l'utilisation, mais aussi les époques de constitution des tissus urbains (Herold *et al.*, 2003b ; Wang *et al.*, 2013). La section suivante s'attache donc à préciser la place de la végétation en ville et son intérêt d'un point de vue environnemental. Elle montre comment la trame verte est un élément constitutif de la structure urbaine.

### 1.3 La trame verte

Cette section présente la notion de trame verte en milieu urbain. Cette notion est apparue récemment en France suite au Grenelle de l'Environnement, mais la végétation urbaine a depuis longtemps été considérée dans la ville. Nous présentons tout d'abord la place de la végétation dans l'urbanisme au cours du temps (§ 1.3.1) puis dans la section 1.3.2 l'émergence de la notion de trame verte en France qui n'est pas spécifique aux espaces urbains. Les trames vertes en milieux urbains peuvent remplir une multitude de fonctions qui sont présentées lors de la section 1.3.3. Enfin, nous présentons comme nous l'avons fait pour la trame grise, quelques typologies utilisées pour décrire ces trames vertes que ce soit dans les espaces naturels ou dans les espaces urbains (§ 1.3.4).

#### 1.3.1 La prise en compte de la végétation en urbanisme

La place accordée à la végétation en ville a connu de grandes transformations au cours du temps dans les politiques et dans les courants urbanistiques. Mehdi *et al.* (2012) distinguent trois grandes périodes de l'utilisation de la végétation en ville depuis le 19<sup>e</sup> siècle qui sont l'urbanisme fonctionnaliste, l'urbanisme environnemental et l'urbanisme écologique. Toutefois, la végétation était aussi prise en compte dans les périodes précédentes. Pour reprendre les héritages urbanistiques identifiés par Allain (2004) (§ 1.2.1.2), dans les villes médiévales la végétation était très peu présente à cause de la contiguïté des maisons, des rues étroites et des places de petites dimensions malgré l'existence de quelques jardins dans le noyau urbain (Claval, 1986). Dans les faubourgs toutefois, la végétation est présente pour ses services de production agricole et qui, lors de l'agrandissement des fortifications, se retrouve en partie intégrée dans l'enceinte de la ville.

Dans la ville classique, les jardins d'agrément se développent notamment à proximité des palais puis des hôtels particuliers à partir du 17<sup>e</sup> siècle, mais ne sont pas encore

destinés à l'ensemble des citoyens. C'est à partir de la destruction des fortifications des villes que des boulevards plantés d'arbres apparaissent et constituent des lieux de promenade. Au 18<sup>e</sup> siècle, de nombreux jardins sont créés au sein des villes et sont de plus en plus ouverts à l'ensemble de la population urbaine avec des préoccupations hygiénistes qui commencent à apparaître (Claval, 1986).

Le 19<sup>e</sup> siècle est celui de l'émergence de l'urbanisme fonctionnaliste de la prise en compte de la végétation décrit par Mehdi *et al.* (2012) principalement avec le courant hygiéniste qui déplore les conditions de vie des classes populaires et l'antagonisme ville-campagne. Pour ces auteurs deux visions de l'urbanisme coexistent : celle qui place la végétation au centre de l'aménagement et où ville est organisée par un réseau important des parcs et de jardins et un second courant qui intègre la végétation comme une infrastructure, qui vient compléter les fonctions rendues par les divers autres équipements urbains.

Les urbanistes de ce premier courant sont nombreux. Par exemple, Frederick Olmsted veut créer un ensemble de parcs urbains qui seraient tous reliés par de larges promenades plantées (*parkways*) afin de former un système de parcs urbains qui serait l'élément organisateur de la ville (Toublanc et Bonin, 2012). Ces principes sont repris par Forestier au début du 20<sup>e</sup> siècle avec l'idée que ce système de parcs devait aussi intégrer les espaces naturels situés à proximité des villes. Howard propose aussi à cette période les cités jardins dont l'idée est de mêler ville et campagne (§ 1.2.1.2). Ces cités devaient contenir en leur centre un grand parc qui intégrerait l'ensemble des services urbains. Il propose aussi le concept de ceinture verte (*green belt*) dont l'objectif est de préserver une couronne d'espaces non bâtis autour des villes pour éviter leur agglomération (Toublanc et Bonin, 2012). Ces ceintures vertes, qui peuvent être forestières ou agricoles, ont eu un écho particulièrement grand et beaucoup de villes notamment en France s'en sont pourvus (Donadieu et Fleury, 2003).

Le deuxième courant de pensée identifié par Mehdi *et al.* (2012) considère plus la végétation comme un équipement urbain qu'un élément organisateur de la ville. C'est le cas d'Eugène Hénard qui propose, pour la ville de Paris à l'époque de la ville libérale (§ 1.2.1.2), que l'ensemble des habitants soit à moins de 500 m d'un espace vert et à 1 km d'un parc urbain. Dans la ville qu'Allain (2004) décrit comme fordiste et fonctionnaliste, la végétation a aussi une place importante. Le Corbusier par exemple laisse une grande place aux piétons et imagine les secteurs d'habitation comme des parcs où seraient élevés des immeubles de grande hauteur entourés d'espaces végétalisés. Pour ces deux courants, l'espace vert est vu comme un équipement urbain qui doit répondre aux besoins d'exercice physique des citoyens et comme un complément de logement dans les grands ensembles. Ces théories ont entraîné une uniformisation de ces espaces végétalisés (Merlin et Choay, 1988).

À partir des années 1990, les préoccupations environnementales et paysagères prennent de l'ampleur et apparaissent dans les documents d'urbanisme (Toublanc et Bonin, 2012). Le terme de rupture verte y apparaît dont l'objectif est de maîtriser l'étalement urbain. C'est l'urbanisme que Mehdi *et al.* (2012) appellent environnemental qui s'appuie sur les

concepts de développement et d'urbanisme durable. Les espaces végétalisés en milieu urbain sont toujours associés à une fonction urbanistique et à une fonction sociale permettant d'offrir des paysages quotidiens de qualité et des espaces de nature récréatifs au sein de la ville. Mais ces espaces doivent aussi jouer leur rôle dans la protection de l'environnement. Ils remplissent ainsi des fonctions multiples (§ 1.3.3). La tendance est aussi dans les années 2000 à la densification des espaces urbains afin de protéger les écosystèmes non urbains, mais qui remet en cause la place laissée au vert dans les villes. De plus que ce soit dans les lois ou dans le discours urbanistique « *les espaces verts ne sont pas considérés en tant qu'écosystème, mais toujours en qualité d'équipement urbain essentiellement à vocation sociale* » (Mehdi et al., 2012).

C'est aussi à la même période que de nombreux travaux scientifiques prennent en compte la végétation urbaine sous l'angle combiné de l'écologie du paysage et de l'aménagement (Arrif et al., 2011). Apparaît alors un nouvel urbanisme écologique (Blanc, 2009) qui se définit comme « *le développement de modes durables de conception du milieu urbain, qui entraîne un nouveau rapport entre l'urbain et la nature [...] et s'enrichit du processus de mise en œuvre de trames vertes urbaines* » (Arrif et al., 2011).

### 1.3.2 La notion de trame verte

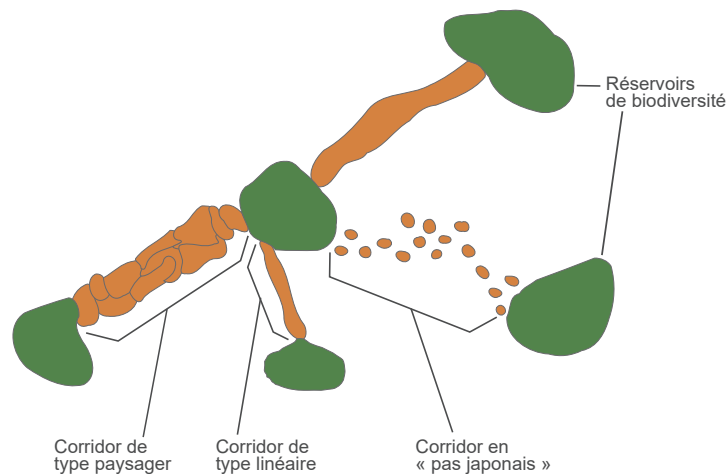
La notion de trame verte associée à la trame bleue apparaît en France à l'issue du Grenelle de l'environnement et des lois Grenelle 1 (2009) et 2 (2010). Elles sont une réponse à la perte de biodiversité constatée à l'échelle planétaire qui se traduit par une disparition sans précédent d'espèces animales et végétales et de milieux naturels (Allag-Dhuisme et al., 2010). Cette diminution de la biodiversité est principalement la conséquence de sept facteurs (Millennium Ecosystem Assessment, 2005) :

- L'accroissement de la demande en nourriture et en eau liée à l'augmentation de la population.
- L'augmentation des pollutions et des déchets liée aux activités humaines.
- Les échanges commerciaux internationaux.
- Le changement climatique.
- La surexploitation des ressources naturelles.
- Le développement d'espèces invasives.
- Les changements d'occupation et d'utilisation du sol.

Ce dernier facteur a une influence sur la biodiversité soit directement en détruisant des espaces naturels soit indirectement par la fragmentation et l'isolement des habitats (Clergeau, 2007). Les villes sont particulièrement en cause puisque leur développement s'est traduit par un phénomène d'étalement urbain parfois très important, mais aussi la construction de grandes infrastructures de transports dont une conséquence a été de fragmenter les paysages naturels et agricoles (EEA, 2006). La fragmentation de ces espaces et la disparition des corridors écologiques peuvent être un frein à la libre circulation des espèces et à la continuité des processus écologiques (Gilbert-Norton

*et al.*, 2010). Ces corridors permettent de connecter les habitats naturels d'un même type entre eux et sont nécessaires pour les populations animales et végétales afin d'assurer un brassage génétique et une population suffisamment nombreuse, mais aussi de migrer si les conditions du milieu changent ce qui est de plus en plus probable dans le contexte de réchauffement climatique. Depuis les années 1990, ce constat a mis en évidence la volonté de conserver et de développer les infrastructures vertes au niveau européen (Jongman *et al.*, 2004). En France, cela s'est traduit dans les lois Grenelle 1 et 2 par la décision de mettre en place un réseau de Trames Vertes et Bleues.

Les trames vertes et bleues recouvrent deux acceptations : une première écologique et une deuxième qui est un ensemble de mesures pour prendre en considération les continuités dans les politiques d'aménagement du territoire (Combrette, 2016). Les trames sont composées de deux éléments principaux (figure 1.10). Les premiers sont les réservoirs de biodiversité qui correspondent aux espaces où les espèces peuvent assurer leur cycle de vie et à partir desquels elles se dispersent (Allag-Dhuisme *et al.*, 2010). Les seconds sont les corridors écologiques qui relient les réservoirs permettant ainsi aux espèces de migrer. Ils peuvent être linéaires, en pas japonais, terrestres, aériens ou aquatiques. L'objectif écologique des trames vertes et bleues en France est de conserver ces habitats et d'assurer leur connectivité.



**Figure 1.10** – Exemple d'éléments de la Trame verte et bleue : réservoirs de biodiversité et types de corridors (de Allag-Dhuisme *et al.*, 2010)

La connectivité se divise en deux composantes : une connectivité structurelle et une connectivité fonctionnelle. La connectivité structurelle correspond à la contiguïté des éléments physiques du paysage (Sordello *et al.*, 2014) et représente ainsi la partie cartographiable de l'arrangement spatial de la trame (Combrette, 2016). Elle se base donc souvent sur l'utilisation d'indicateurs quantitatifs (CREAT, 2010) ou sur le déplacement hypothétique d'espèces (Rudnick *et al.*, 2012). La connectivité fonctionnelle fait référence au degré auquel le mouvement des individus existe réellement dans le paysage même si les taches d'habitats sont éloignées l'une de l'autre (Burel et Baudry, 1999). Elle

est donc dépendante du potentiel de mouvement des populations lié à la connectivité structurelle, mais aussi de leur dynamique (Sordello *et al.*, 2014). Pour l'évaluation fonctionnelle des connectivités, il n'est pas possible de la réaliser avec l'ensemble des espèces et il est nécessaire de faire le choix d'une espèce ou d'une guildes d'espèces. Toutefois, les capacités de dispersion sont assez mal connues et il est donc plus facile d'utiliser une approche fonctionnelle et paysagère et d'utiliser l'entrée espèce comme un moyen de communication et d'information à partir d'une espèce emblématique d'un territoire pour expliquer la nécessaire sauvegarde de la biodiversité (Liénard et Clergeau, 2011). La notion de connectivité est donc essentielle pour la définition des corridors et nécessite donc l'identification préalable des réservoirs de biodiversité ainsi que la cartographie des éléments qui peuvent potentiellement constituer ces continuités.

En France depuis 2009, les trames vertes et bleues doivent être prises en compte dans les documents d'urbanisme et d'aménagement à toutes les échelles territoriales. Cette prise en compte s'appuie sur une logique descendante d'emboîtement d'échelle qui se décompose en quatre niveaux. Le premier niveau est celui de l'État qui fixe les grandes orientations stratégiques pour la préservation et le développement des continuités écologiques. C'est ensuite aux régions d'intégrer ces grandes orientations et de définir les enjeux régionaux pour les intégrer dans leur schéma régional de cohérence écologique (SRCE) qui devait être réalisé pour 2012. Le SRCE doit ainsi proposer une cartographie synthétique des trames vertes et bleues au 1/100 000<sup>e</sup> (§ 2.1.3). Les documents d'urbanisme aux niveaux plus fins, les SCoT (schéma de cohérence territoriale – 1/25 000<sup>e</sup>) au niveau intercommunal et au communal les PLU (plan local d'urbanisme – 1/5000<sup>e</sup>), doivent intégrer eux aussi ces trames en étant conformes ou compatibles avec les orientations fixées dans le SRCE. La prise en compte des trames vertes et bleues suit donc une logique verticale, mais nécessite aussi d'être horizontale puisque les continuités écologiques ne peuvent être simplement limitées à une entité administrative et nécessite que la continuité soit aussi territoriale. Elle engage un nombre d'acteurs très important qui peut expliquer les retards pris dans la réalisation des SRCE qui ont cependant aujourd'hui tous été adoptés en métropole excepté en Picardie (annexe A) et leur intégration dans les SCoT et les PLU.

Pour prendre en compte les trames vertes et bleues, les aménageurs à tout niveau s'appuient souvent sur des zonages réglementaires existants tels que les réserves naturelles nationales, les parcs naturels régionaux, les ZNIEFF (zones naturelles d'intérêt écologique faunistique et floristique), le réseau Natura 2000, etc. (Cormier *et al.*, 2010). Ces zonages sont pour la plupart créés par l'État, les régions ou les départements. Aux échelons inférieurs, la définition des continuités écologiques est différente notamment dans les espaces urbains puisque les collectivités locales n'ont pas la maîtrise de ces zonages, donc d'outils adaptés à la protection de la trame, et que les processus écologiques sont complètement différents dans ces milieux (Clergeau et Hubert-Moy, 2011). La biodiversité dans les espaces urbains est souvent qualifiée d'ordinaire (Daniel *et al.*, 2013) et les continuités peuvent être composées d'éléments très petits à ces échelons qui nécessitent des méthodes d'identification qui ne sont les mêmes qu'aux échelles

régionales. Maurel *et al.* (2013) insistent par exemple sur la place des arbres isolés et de leur pied qui peuvent permettre le déplacement d'espèces animales et végétales avec des corridors en pas japonais (figure 1.10). Les jardins privés ont aussi une place à prendre dans la définition de la trame verte (Daniel *et al.*, 2013 ; Riboulot-Chetrit, 2015) bien qu'ils soient généralement intégrés aux espaces artificialisés et sont donc, même lorsque la cartographie de la trame est réalisée à une échelle très grande, souvent considérés comme des barrières infranchissables (Liénard et Clergeau, 2011). Ces continuités apparaissent aussi plus efficaces lorsque de nombreuses espèces végétales sont présentes et lorsqu'elles sont composées de plusieurs strates de végétation (Bergoënd *et al.*, 2013), mais les modes de gestion de ces espaces végétalisés urbains privés ou publics ont une très forte influence sur cette efficacité (Maurel *et al.*, 2013 ; Riboulot-Chetrit, 2015).

La logique descendante de la définition de ces trames dans les documents d'aménagement et d'urbanisme rend aussi difficile l'appropriation aux niveaux communal et intercommunal des trames vertes et bleues (Cormier *et al.*, 2010). En effet, la traduction à un échelon plus fin des directives régionales est difficile et le risque est fort de ne plus retrouver qu'une partie incomplète des données descendantes. De plus la définition des trames vertes peut être très différente selon les niveaux. Aux niveaux de l'État et des régions, les trames vertes et bleues conservent leur sens d'origine à savoir favoriser la dispersion des espèces et ont donc une visée écologique forte. Aux niveaux plus fins, les trames ont plus une visée de sauvegarde du paysage et du patrimoine naturel et agricole et, plus spécifiquement dans les espaces urbains, une vision hygiéniste et d'amélioration du cadre de vie qui correspond plus à l'histoire de la prise en compte de la végétation dans ces espaces (Arrif *et al.*, 2011 ; Toublanc et Bonin, 2012 ; Clergeau et Blanc, 2013). Ces différentes définitions des trames vertes et bleues ne sont toutefois pas antinomiques puisque, dans l'idée du développement durable, le caractère multifonctionnel peut être mis en avant (Ahern, 1995 ; Mehdi *et al.*, 2012). La section suivante présente donc la multifonctionnalité ainsi que la notion associée de services écosystémiques que peuvent représenter les trames et plus spécifiquement la trame verte dans les milieux urbains.

### 1.3.3 Multifonctionnalité de la végétation urbaine

L'idée de multifonctionnalité des éléments naturels a émergé tout d'abord avec les forêts dans les années 60 avec l'idée qu'elle devait assurer la production de bois tout en prenant en compte les problématiques environnementales puis s'est ensuite étendue aux espaces agricoles dans les années 90 et s'est élargie à tous les types de paysage (Selmi *et al.*, 2013). Ces fonctions peuvent être sociales, écologiques, esthétiques, environnementales, paysagères, etc. (Blanc *et al.*, 2012). En même temps est apparue la notion de services écosystémiques qui désigne l'ensemble des bénéfices directs et indirects rendus à la société par les fonctions des écosystèmes (Costanza *et al.*, 1997). Le Millennium Ecosystem Assessment (2005) a classé ces services en quatre types différents qui ne sont pas spécifiques aux espaces urbains ou à la végétation :



- Les services de support qui sont ceux nécessaires à tous les autres services écosystémiques tels que la production d'oxygène, de matières premières, la formation du sol, etc.
- Les services d'approvisionnement qui correspondent aux produits issus des écosystèmes tels que la nourriture, les énergies ou l'eau douce.
- Les services de régulation des processus environnementaux tels que l'amélioration de la qualité de l'air, la régulation du climat, la limitation de l'érosion, la purification de l'eau.
- Les services culturels qui sont les avantages immatériels issus des écosystèmes à travers les loisirs et les expériences esthétiques ou éthiques.

Toutefois, la végétation urbaine ne propose pas que des services écosystémiques, mais aussi des desservices (Selmi *et al.*, 2013). En effet, les arbres peuvent être allergènes, produire des composés organiques volatils (COV) qui contribuent à la production d'ozone (Mehdi *et al.*, 2013) et, en perturbant les flux atmosphériques, peuvent limiter les échanges avec l'atmosphère et donc augmenter localement la pollution de l'air (Wania *et al.*, 2012). Les continuités peuvent aussi permettre le développement d'espèces invasives.

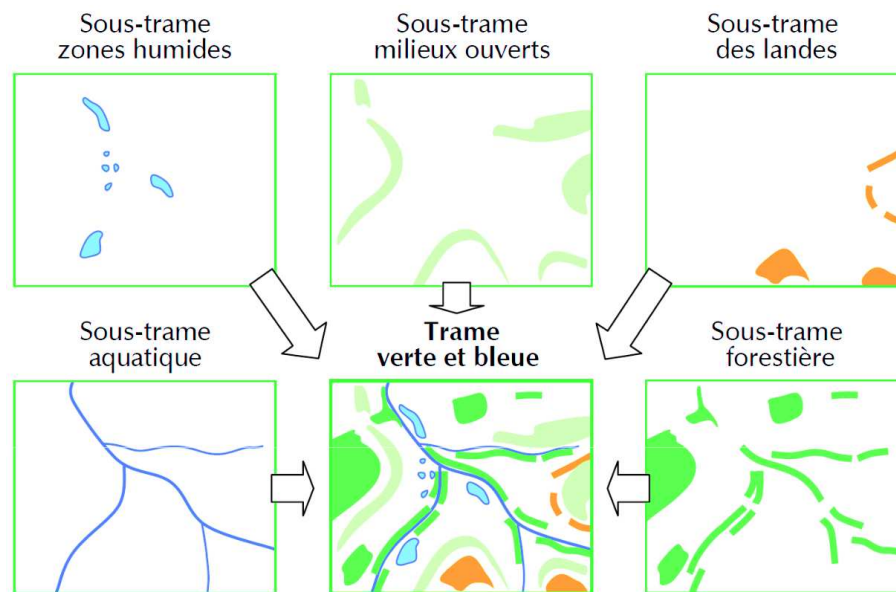
Dans la définition des trames vertes notamment par les collectivités urbaines, les services culturels sont souvent ceux qui sont mis en avant (Clergeau et Blanc, 2013). Elles permettent ainsi de maintenir ou de développer des espaces récréatifs et participent donc à l'amélioration du cadre de vie des citoyens. Elles ont aussi une influence positive sur la santé (Tzoulas *et al.*, 2007 ; Taylor *et al.*, 2015). En améliorant la qualité environnementale de ces espaces, les trames vertes participent aussi à la revalorisation du foncier (Morancho, 2003).

Avec les problématiques environnementales auxquelles sont confrontées les villes, les services écosystémiques de régulation des processus environnementaux prennent une place de plus en plus importante dans la recherche, mais aussi dans les politiques de la ville. La végétation a en effet une influence importante sur le climat urbain et les échanges entre la ville et l'atmosphère (Mehdi *et al.*, 2013). Tout d'abord, elle permet une diminution de l'îlot de chaleur urbain grâce à plusieurs processus et permet ainsi la création d'îlots de fraîcheur lorsque les surfaces végétalisées sont suffisamment grandes (ADEUS, 2014). La végétation arborée tout d'abord permet par effet d'ombrage, mais aussi grâce à l'évapotranspiration des arbres de réduire les températures qui peuvent être comprises entre 1 et 2 °C. Toutefois en milieux urbains l'évapotranspiration des plantes est réduite, car les sols contiennent moins d'eau liée au ruissellement qui est plus important dans ces espaces (Gill *et al.*, 2007), mais la mise en place de végétation notamment arborée permet de retenir ces eaux de pluie (Armson *et al.*, 2013). De plus, la végétation urbaine permet de réduire les pollutions, sonores par effet d'écran (Fang et Ling, 2003), mais aussi atmosphériques grâce à la captation de carbone et de polluants (Nowak, 1994. ; Selmi *et al.*, 2016)

L'ensemble de ces services produits par la végétation urbaine pousse ainsi les décideurs à mieux considérer les trames vertes dans les politiques publiques en y intégrant leur caractère multifonctionnel. La demande est forte pour obtenir des cartographies de ces trames afin de les intégrer et nécessite en milieu urbain une identification exhaustive et à jour du fait de la taille des éléments constitutifs qui peuvent être de petites dimensions telle que des arbres isolés. Toutefois, le choix des trames à identifier est nécessaire, car elles peuvent être multiples comme le montre la section suivante.

### 1.3.4 Typologies vertes

Tout comme pour l'analyse des tissus urbains, de nombreuses typologies ont été proposées pour la cartographie des trames vertes et bleues. Ces trames sont en effet composées de sous-trames qui correspondent chacune à des milieux ou des habitats et peuvent donc être très différentes selon les territoires étudiés. Dans la proposition du comité opérationnel trame verte et bleue, Allag-Dhuisme *et al.* (2010) proposent une déclinaison en cinq sous-trames principales : milieux forestiers, landes, milieux ouverts, zones humides et aquatiques (figure 1.11).



**Figure 1.11** – Exemple de Trames verte et bleue composées de sous-trames écologiques spécifiques (de Allag-Dhuisme *et al.*, 2010)

Dans l'élaboration des SRCE, les typologies des sous-trames peuvent être différentes selon les régions et leur nombre varier de quatre pour les régions Haute-Normandie et Île-de-France à dix pour le Centre-Val de Loire et le Nord-Pas-de-Calais (annexe A). Dans ces documents, les sous-trames sont assez générales et des approches qui se concentrent plus sur une approche d'identification des milieux peuvent les compléter. Hostein *et al.*

(2014) proposent ainsi une cartographie des milieux agropastoraux du Massif central en 29 classes et utilisent des images aériennes et satellites ainsi que des données Lidar (*light detection and ranging*) pour la réaliser (§ 2.1.1). Sellin *et al.* (2015) utilisent aussi des images satellites afin de cartographier la végétation par une approche habitat leur permettant d'obtenir une cartographie en 4 niveaux dont le plus fin est composé de 22 types d'habitats (tableau 1.2).

**Tableau 1.2** – Typologie proposée par Sellin *et al.* (2015) pour la cartographie des formations végétales par télédétection

Niveau 1 (Occupation du sol)	Niveau 2 (Grands types de végétation)	Niveau 3 (Types de formations végétales)
Milieus non végétalisés		
Végétations artificielles	Cultures	
	Plantations Parcs et jardins	
Végétations naturelles et semi-naturelles	Forêts	Forêts mésophiles (caducifoliées)
		Forêts hygrophiles (caducifoliées)
		Forêts de conifères
		Forêts de sclérophylles
		Forêts mixtes
	Fourrés arbustifs	Fourrés mésophiles Fourrés hygrophiles
	Fourrés nains	Landes et garrigues
		Tourbières à chaméphytes Fourrés crassulescents
	Végétations aquatiques marines	Herbiers marins
	Végétations herbacées	Prairies mésophiles
Prairies hygrophiles		
Pelouses sur sable non fixé		
Pelouses sur sable fixé		
Roselières et cariçaiies Franges et lisières Prés salés		
Milieus aquatiques avec ou sans végétations amphibies associées		

Toutefois, ce type d'approche n'est pas le plus fréquent et les recherches sur les trames vertes et bleues s'intéressent souvent à la cartographie d'une sous-trame spécifique. Vacquie et Houet (2012) vont ainsi identifier les zones humides dans des milieux montagnards des Pyrénées à partir d'images satellites et cartographier trois types de zones humides : les prairies inondables, les prairies et landes humides et les tourbières. Daniel *et al.* (2013) se concentrent sur la sous-trame arborée afin de comparer trois villes de l'ouest de la France, Angers, Rennes et Nantes et analysent ses évolutions et son intégration dans les documents d'urbanisme. La trame forestière est aussi celle considérée par Hubert-Moy *et al.* (2012). Dans cette recherche l'objectif est de réaliser la cartographie de cette trame à trois niveaux : régional, intercommunal et communal, et de vérifier si les données géographiques existantes peuvent permettre une cartographie exhaus-

tive et suffisamment détaillée pour évaluer la connectivité structurelle à ces différents niveaux.

Les typologies des SRCE ainsi que des recherches qui viennent d'être présentées ne sont pas spécifiquement consacrées aux milieux urbains. La végétation est beaucoup moins présente avec une densité de plus en plus importante au fur et à mesure que la densité de bâti décroît (Clergeau, 2007). Pour Bolund et Hunhammar (1999) les éléments qui constituent les espaces végétalisés en milieu urbain sont de sept types différents :

- Les arbres sur rue qui peuvent être isolés ou en alignement et souvent entourés de sol artificialisé.
- Les parcs et les espaces verts gérés souvent composés d'herbe, de grands arbres et d'autres plantes. Ces auteurs incluent dans ce type les équipements sportifs tels que les terrains de jeux ou les golfs.
- Les forêts urbaines qui sont les éléments les plus denses et les moins gérés.
- Les zones cultivées et les jardins ouvriers utilisés pour la production.
- les zones humides composées de marais et marécages.
- Les lacs et les mers.
- Les rivières et les cours d'eau.

Pour la cartographie des trames vertes urbaines de quatre sites d'études : Toulouse, Pierrefite-Nestalas, Muret et Strasbourg, Combrette (2016) les définit à partir de trois éléments : les arbres, les arbustes et l'herbe et les évalue grâce à la construction d'un graphe paysager (Foltête *et al.*, 2012). Ces graphes lui permettent de mesurer les connectivités potentielles des trames vertes et de simuler les effets d'un projet d'aménagement sur ces connectivités. De même dans ce travail nous nous intéressons spécifiquement à la trame verte et considérons donc l'ensemble des espaces végétalisés urbains comme étant constitutifs des trames.

## Conclusion

Ce premier chapitre a permis de présenter notre objet d'étude, l'analyse des formes urbaines que nous définissons comme l'analyse à l'échelle des tissus de l'organisation de la structure urbaine. Cette structure urbaine est composée de deux éléments principaux : la trame grise et la trame verte. Pour ces deux trames, nous nous intéressons à la matérialité physique des éléments qui les composent et qui constituent le paysage urbain. La trame grise est composée principalement des bâtiments et du réseau de communication qui peut être routier ou ferré tandis que la végétation arborée, arbustive et herbacée ainsi que les cultures qui peuvent être très présentes dans les espaces périurbains sont les éléments que nous considérons comme constitutifs de la trame verte.

L'objectif de cette thèse étant la cartographie de la structure urbaine à l'échelle des tissus, le prochain chapitre permet de définir les données géographiques, leurs caractéristiques

et de vérifier si les bases de données existantes peuvent être utilisées afin d'analyser quantitativement la structure urbaine par ses trames grises et vertes.



# Chapitre 2

## Données géographiques et méthodes existantes pour la cartographie de la structure urbaine

### Sommaire

---

Introduction . . . . .	38
2.1 Les données géographiques . . . . .	38
2.1.1 Types de données géographiques . . . . .	38
2.1.2 Évolution de l'offre en imagerie optique . . . . .	39
2.1.3 Notion d'échelle et lien avec la résolution des images . . . . .	41
2.2 Sélection des sites d'études et données disponibles . . . . .	43
2.2.1 Sites d'études . . . . .	43
2.2.2 Bases de données géographiques disponibles . . . . .	45
2.3 Cartographie de la structure urbaine : aspects méthodologiques . . . . .	54
2.3.1 Évolution des approches pour l'analyse d'image : du pixel à l'objet . . . . .	55
2.3.2 Cartographie des éléments verts . . . . .	57
2.3.3 Cartographie des tissus urbains . . . . .	61
Conclusion . . . . .	70

---

## Introduction

Le chapitre précédent a permis de préciser les différentes notions essentielles de ce travail. Il a aussi permis de préciser l'objet d'intérêt de cette thèse, la structure urbaine que nous considérons comme la combinaison de la trame grise et de la verte. Pour analyser la structure urbaine, il est important de disposer d'une cartographie précise de la ville au niveau des tissus et des éléments urbains qui soit précise et à jour.

L'objectif de ce chapitre est donc de définir dans un premier temps les données géographiques et leurs caractéristiques (§ 2.1). La section 2.2 présente les sites d'études ainsi que les données géographiques disponibles pour ces sites et qui permettent de caractériser la ville. Cette section nous permet aussi de vérifier si ces bases de données sont adaptées à l'analyse de la structure urbaine à l'échelle des tissus urbains. Enfin la section 2.3 présente un état de l'art des méthodes existantes pour la cartographie de la trame verte et de la trame grise.

## 2.1 Les données géographiques

Cette première section permet de définir ce qu'est une donnée géographique. Dans un premier temps, nous présentons les différents types de données disponibles et précisons l'évolution de leur mode de production et de leur utilisation (§ 2.1.1). La section 2.1.2 présente plus en détail les caractéristiques des images satellites optiques et leurs évolutions. Enfin, nous définissons une notion essentielle pour les données géographiques : l'échelle et revenons sur sa relation avec la résolution (§ 2.1.3).

### 2.1.1 Types de données géographiques

Une donnée est géographique si elle permet de décrire un objet, un événement, un phénomène spatialement référencé sur la surface de la Terre (Puissant, 2003). Elles peuvent être de nature diverse : décrire l'occupation du sol, des caractéristiques socio-économiques, des découpages administratifs, etc. Deux représentations spatiales sont possibles pour les données géographiques, une vectorielle et une matricielle (raster, image).

La représentation vectorielle représente les objets géographiques de manière continue et non discrétisée. Trois formes peuvent être utilisées : le polygone qui permet de représenter une surface, la ligne pour une limite administrative ou une route par exemple, le point qui peut se référer à une adresse. Pour chacun des objets représentés par ces formes, il est possible d'associer des informations attributaires permettant de les décrire.



La représentation matricielle de l'espace discrétise l'espace géographique en mailles (les pixels) pour lesquelles une ou plusieurs valeurs numériques sont associées. Ces rasters peuvent être thématiques et concerner ainsi l'occupation du sol, les espaces verts, etc., mais peuvent aussi être des données brutes provenant de photographies aériennes ou de satellites. Ces images de télédétection sont issues d'un processus d'acquisition d'information par l'intermédiaire d'un capteur qui va acquérir un rayonnement électromagnétique qui peut être émis ou réfléchi par les objets. Ces images peuvent être obtenues de deux manières différentes. La première est dite passive, c'est-à-dire que le capteur va uniquement enregistrer des longueurs d'onde qui sont réfléchies ou émises naturellement et ils sont donc dépendants d'une source d'énergie extérieure (solaire, dégagée par les matériaux). Ces capteurs optiques sont ceux qui nous intéressent dans la suite de cette thèse. La deuxième est active ce qui signifie que le capteur est associé à une source d'énergie qui va envoyer un signal qui sera ensuite enregistré par ce capteur. Il s'agit par exemple des radars ou des Lidars qui se différencient principalement par la longueur d'onde qui est émise.

La production de ces données géographiques, et plus spécifiquement les données brutes, était en général dominée par des services gouvernementaux tels que l'IGN en France (Institut national de l'information géographique et forestière). Toutefois une révolution a eu lieu sur les modes d'acquisition, de production et de distribution de ces données (Puissant, 2014). Le développement de l'internet à haut débit a permis une démocratisation de l'accès aux données géographiques. D'abord consommateurs, les individus sont devenus producteurs de données. Volontairement tout d'abord avec le développement de plateformes collaboratives de cartographie comme OpenStreetMap (OSM) par exemple et parfois malgré les individus à cause de la géolocalisation de plus en plus permanente qu'entraîne la consommation des nouvelles technologies. Les données géographiques sont de plus en plus accessibles au plus grand nombre avec la mise en place de la gratuité d'un nombre important de données. L'IGN met à disposition une partie de ses données pour toute mission de services publics ainsi que pour l'enseignement et la recherche. Récemment, la France s'est aussi engagée dans l'ouverture des données publiques, « l'open data », afin qu'elles puissent être réutilisées librement. Ce mouvement initié par l'État se décline aussi localement et de plus en plus de collectivités locales s'engagent à rendre accessibles leurs données.

La production des images satellites a aussi connu de nombreuses évolutions. En effet de nombreuses entreprises sont rentrées dans ce marché entraînant par exemple une multiplication des satellites privés d'Observation de la Terre. La section suivante présente les évolutions qu'a connues ce secteur en rappelant tout d'abord les caractéristiques inhérentes aux images.

### **2.1.2 Évolution de l'offre en imagerie optique**

La mise en orbite du premier satellite d'observation de la Terre, le satellite Landsat, a été effectuée en 1972. Au premier janvier 2016, 343 satellites d'Observation de la Terre

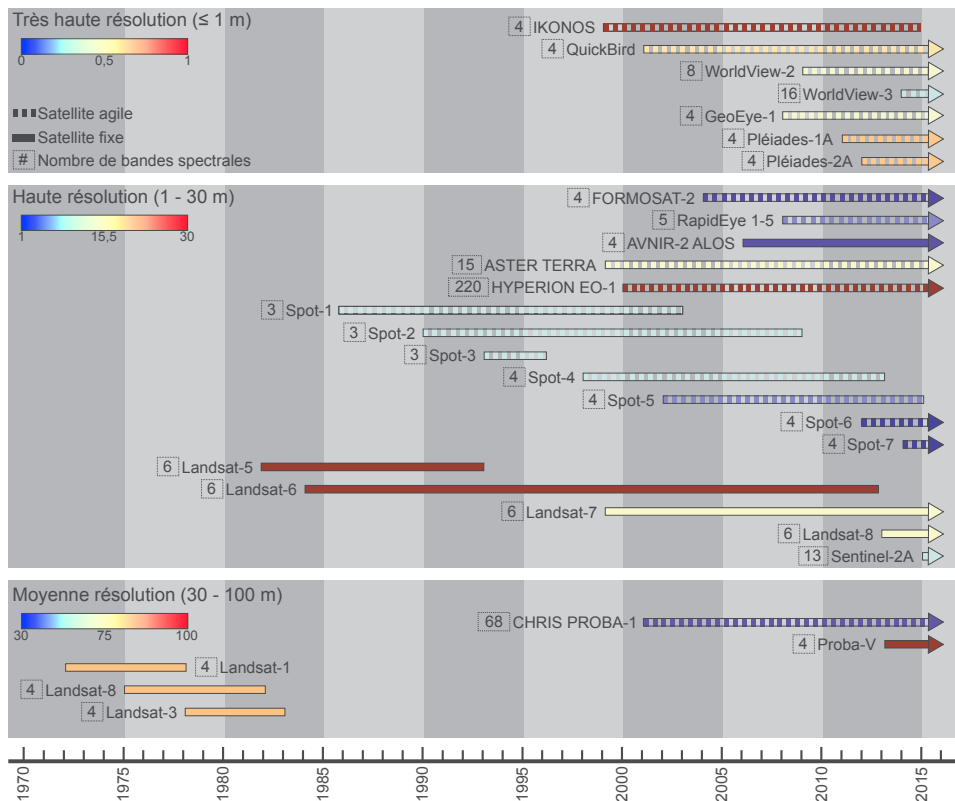
sont en orbite incluant ceux passifs et actifs (*UCS Satellite Database – Union of Concerned Scientists*) et les images qui en sont issues ont connu de nombreuses avancées depuis le premier satellite Landsat. Les images satellites optiques sont caractérisées par trois résolutions : la résolution spectrale, la résolution temporelle et la résolution spatiale.

La résolution spectrale d'un capteur optique caractérise sa capacité à distinguer différentes longueurs d'onde du spectre électromagnétique. Une image est ainsi caractérisée par un certain nombre de bandes spectrales chacune correspondant à un intervalle de longueur d'onde que le capteur peut enregistrer. Un capteur caractérisé par une bonne résolution spectrale est donc un capteur qui peut acquérir une image dans une grande partie du spectre électromagnétique et dont les intervalles pour chaque bande sont les plus étroits possible afin d'obtenir une signature spectrale précise de l'objet étudié. Trois types d'images peuvent être acquises : les images panchromatiques qui n'ont qu'une bande spectrale souvent dans l'ensemble du spectre visible, les images multispectrales qui sont caractérisées par un nombre de bandes inférieur à dix et au-delà des images hyperspectrales qui peuvent être composées de plusieurs centaines de bandes. Le premier satellite Landsat était caractérisé par quatre bandes spectrales et certains capteurs comme Hyperion en ont actuellement plus de 200 (figure 2.1).

La résolution temporelle désigne la fréquence de revisite d'un satellite, c'est-à-dire le temps qui lui est nécessaire pour faire l'acquisition d'une image d'un même site. En général, les satellites avec une plus faible résolution spatiale ont une capacité de revisite meilleure du fait qu'ils balayent des champs géographiques plus larges. Afin d'améliorer la résolution temporelle des satellites, le CNES a par exemple créé une constellation de satellites identiques, les satellites Pléiades, qui sont sur la même orbite, mais décalés à  $180^\circ$  l'un de l'autre. De plus, il est possible d'améliorer cette résolution en rendant les satellites plus agiles dans leur prise de vue afin de viser des zones qui sont décalées du nadir (point représentatif de la verticale descendante). Le satellite GeoEye-1 par exemple a un angle maximal d'inclinaison  $60^\circ$  par rapport au nadir et les satellites Pléiades  $30^\circ$ . Pour ces derniers, cette caractéristique et le fait qu'ils soient en constellation leur permet d'avoir une revisite quotidienne. Cette résolution est un des grands avantages de l'imagerie satellitaire par rapport à celle des images aériennes qui nécessite la programmation d'une campagne de prise de vue dont la fréquence de mise à jour est par exemple de trois à cinq ans pour les données images de l'IGN.

La dernière caractéristique des images est leur résolution spatiale. La résolution spatiale définit la taille du plus petit élément sur une image c'est-à-dire le pixel. Cette caractéristique des images a connu de très fortes évolutions depuis 40 ans puisque le premier Landsat avait une résolution spatiale de 60 m et que certains satellites aujourd'hui proposent des résolutions sub métriques inférieures à 0,5 m. Selon cette résolution on parlera de satellites à moyenne résolution (MRS) lorsqu'elle est comprise entre 30 et 100 m, de satellites à haute résolution (HRS) quand elle est comprise entre 1 et 30 m et enfin de satellites à très haute résolution spatiale (THRS) pour celles qui sont inférieures à 1 m. Ces seuils peuvent être différents selon les auteurs. La résolution spatiale des

images devrait encore s'améliorer dans les années à venir puisque le CNES étudie aujourd'hui la faisabilité d'un satellite, THR-NG (très haute résolution de nouvelle génération), qui permettrait d'obtenir une résolution de 30 cm (CNES, 2016). Cette caractéristique est essentielle puisqu'elle conditionne les objets visibles sur les images et donc l'échelle de représentation et d'utilisation des bases de données extraites de ces images. La section suivante présente ainsi la notion d'échelle telle qu'elle est utilisée en géographie et plus précisément en cartographie et sa relation avec la résolution spatiale des images.



**Figure 2.1** – Chronologie des capteurs satellites optiques et thermiques multi et hyperspectraux (adaptée de Houborg *et al.* (2015))

### 2.1.3 Notion d'échelle et lien avec la résolution des images

L'échelle en géographie est une notion abondamment utilisée et polysémique (Brunet *et al.*, 1993). Elle peut correspondre à l'intensité d'un phénomène telle que l'échelle de Richter, mais peut aussi indiquer une forme de hiérarchie et être plutôt un synonyme de niveau ou d'échelon. Cela peut être le cas lorsque l'on parle du territoire de certaines collectivités tel que l'échelle communale, l'échelle départementale, etc. En cartographie, l'échelle désigne le rapport entre la mesure d'une distance sur une carte et la mesure de

distance sur le terrain. Une échelle au 1/200<sup>e</sup> signifie que 1 cm correspond dans la réalité à 200 cm. Le rapport d'échelle étant une fraction, une grande échelle correspondra à une portion d'espace réduite où les éléments sont cartographiés avec une grande précision et une petite échelle représentera donc une portion d'espace plus vaste.

Pour une représentation cartographique d'un phénomène ou de l'occupation du sol, le choix d'une échelle ne fera pas apparaître les mêmes informations, le même niveau de détail ou la même nomenclature. Il est ainsi nécessaire que l'échelle sélectionnée soit en adéquation avec l'objet étudié, mais aussi avec l'instrument utilisé pour sa mesure (Marceau, 1999). En effet, une échelle trop grande pour une étendue elle-même de grande dimension, peut rendre illisible une carte où chaque objet figuré serait donc trop petit. L'inverse ne serait pas non plus adapté, car elle ne permettrait pas de voir les spécificités locales et résulterait visuellement par de vastes plages colorées. Il est donc nécessaire de trouver un compromis entre la résolution qui désigne comme pour l'image la taille du plus petit élément et l'étendue géographique d'autant plus si des traitements informatiques sont réalisés (Forriez, 2010).

Pour l'analyse de la ville et selon les applications ou les objectifs, les échelles peuvent fortement varier. Avec un objectif de gestion telle que l'entretien du réseau, des transports, etc., une grande échelle est nécessaire (1/200<sup>e</sup> au 1/1000<sup>e</sup>) alors que pour l'analyse des espaces urbains avec pour objectif la planification ou des opérations d'aménagement une échelle du 25 au 1/50 000<sup>e</sup> est suffisante (Puissant, 2003). Pour l'étude des formes urbaines, ces échelles peuvent être très différentes selon la forme étudiée (§ 1.1 – figure 2.2) :

- du 1/25 000<sup>e</sup> au 1/50 000<sup>e</sup> pour l'analyse de la tache urbaine
- du 1/10 000<sup>e</sup> au 1/25 000<sup>e</sup> pour l'étude des formes au niveau du quartier
- du 1/2000<sup>e</sup> au 1/5000<sup>e</sup> pour l'analyse des îlots urbains
- du 1/500<sup>e</sup> au 1/2000<sup>e</sup> pour le niveau de la parcelle et du bâtiment

Les bases de données géographiques sont caractérisées par une échelle de représentation et d'utilisation conseillée qui permet donc une représentation cartographique, mais cette échelle est aussi directement liée à la résolution de cette base. En effet, une base de données à petite échelle a en général une résolution faible, ce qui se traduit lors de leur construction par une surface minimum de cartographie qui va être assez élevée telle que nous le présentons dans la section 2.2.2. Souvent obtenues par photo-interprétation d'images, ces bases de données ont une échelle qui est donc fortement influencée par la résolution spatiale des images utilisées puisqu'elle-même détermine les objets identifiables sur ces images. Le choix de l'image est donc déterminant pour construire les données à une échelle adaptée à l'objet qui est étudié.

Cette section a permis de définir ce qu'est une donnée géographique, comment elle est caractérisée, par un mode de représentation, une échelle, une résolution. Le secteur de l'information géographique connaît actuellement de profonds bouleversements tant dans sa production que dans son utilisation. Pour l'étude du milieu urbain, ces données

sont particulièrement intéressantes, mais nécessitent qu'elles soient à une échelle adaptée et qu'elles soient mises à jour régulièrement pour être utilisées par les collectivités ou les professionnels qui gèrent ces espaces. La section suivante présente les sites d'études retenus dans cette thèse et les données géographiques disponibles afin de vérifier si ces bases sont adaptées pour l'analyse de la structure urbaine.

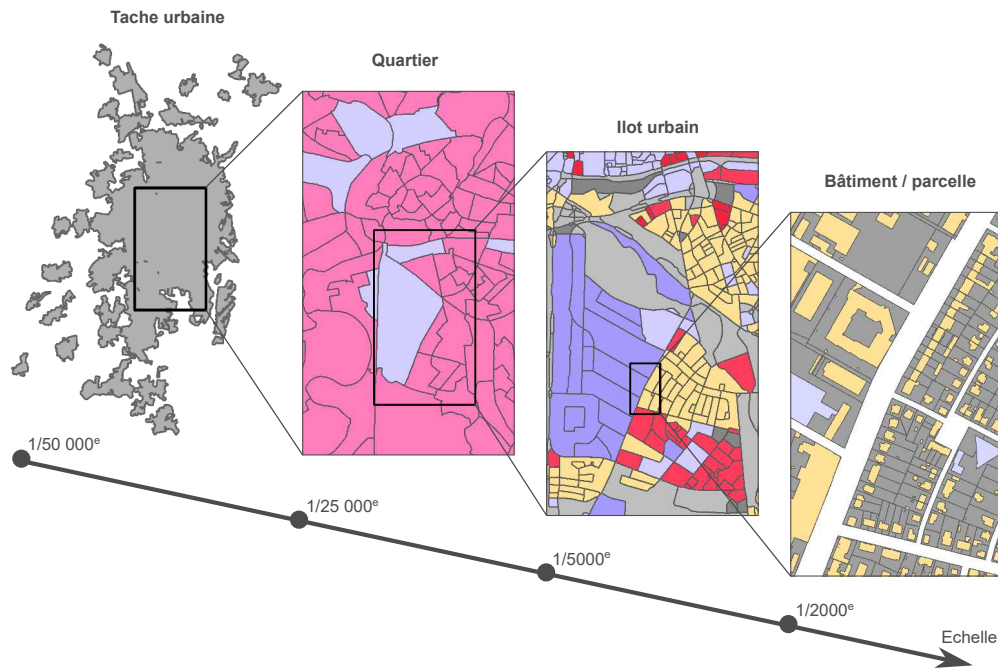


Figure 2.2 – Les échelles d'analyse des formes urbaines

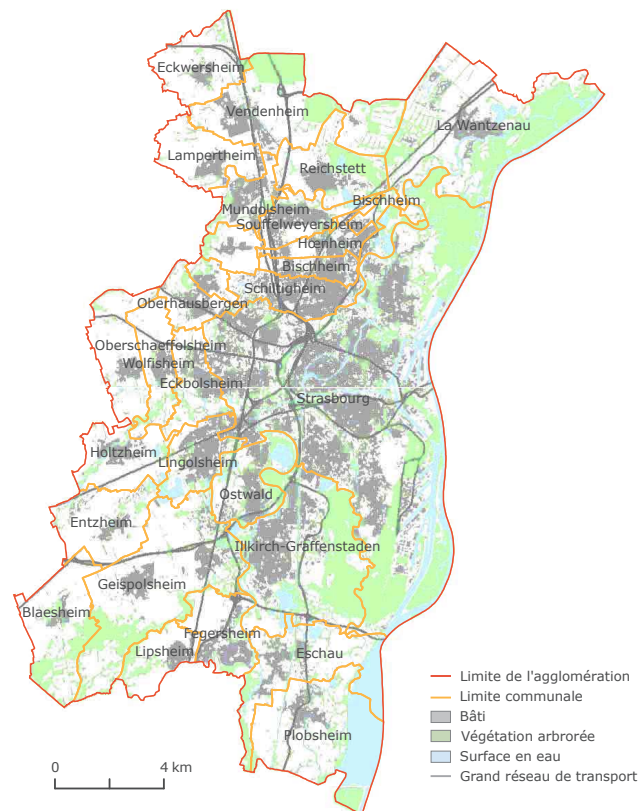
## 2.2 Sélection des sites d'études et données disponibles

L'objectif de cette section est de présenter les sites d'études sélectionnés (§ 2.2.1) pour l'analyse de la structure urbaine qui se décompose en une trame grise et une trame verte. De plus, les bases de données disponibles pour cartographier ces trames sont présentées et évaluées qualitativement par l'analyse de leur nomenclature, de leur échelle de représentation et de leur mode de production (§ 2.2.2).

### 2.2.1 Sites d'études

Deux sites d'étude sont sélectionnés : l'Eurométropole de Strasbourg et Rennes Métropole. Le site de Strasbourg est représentatif de nombreuses villes de l'Europe de l'Ouest avec un centre dense hérité du Moyen Âge et des extensions urbaines organisées

de manière radioconcentrique (figure 2.3). De nombreuses données sont disponibles pour ce site ainsi que des connaissances du terrain au sein du Laboratoire Image, Ville, Environnement (LIVE). Le site de Rennes a une morphologie urbaine différente avec une croissance urbaine autour de centres villageois situés en périphérie (figure 2.4). Des images satellites THRS sont disponibles pour ces deux sites d'études grâce au programme d'accompagnement de la Recette Thématique Utilisateurs (RTU) des images satellites Pléiades du CNES. Dans ce cadre, un programme de recherche a été mis en place entre les universités de Bruxelles, Strasbourg et Rennes : VALI-URB<sup>1</sup>.



**Figure 2.3** – Présentation du site de l’Eurométropole de Strasbourg (source : BD TOPO® 2014, IGN)

L’Eurométropole de Strasbourg est une agglomération composée de 28 communes et située dans l’est de la France (figure 2.3). Elle s’étend sur 316 km<sup>2</sup> avec une population

1. Urban sprawl and ecological corridors : Consolidation of VHRS remote sensing methods. Programme Stereo II BELSPO / CNES (Belgian Federal Science Policy Office / Centre National d’Études Spatiales). En collaboration avec ANAGEO-IGEAT de l’Université Libre de Bruxelles (Analyse Géospatiale – Institut de Gestion de l’Environnement et d’Aménagement du Territoire), le LETG-COSTEL de l’université de Rennes 2 (Littoral, Environnement, Télédétection, Géomatique – Climat et occupation du sol par télédétection) et le LIVE de l’Université de Strasbourg (Laboratoire, Image, Ville et Environnement)

de 473 375 habitants en 2012. La structure de l'agglomération est globalement mono-nucléique (Badariotti, 2005) avec une île centrale qui regroupe l'essentiel de la ville médiévale classique. Elle est limitée à l'est par la présence de la frontière franco-allemande matérialisée par le Rhin. Les tissus urbains strasbourgeois se caractérisent aussi par la présence des tissus haussmanniens de la Neustadt dont la construction a débuté à partir de 1870. Ces tissus sont délimités par une zone non construite correspondant à une ancienne zone militaire transformée en ceinture verte et aujourd'hui occupée par les grands réseaux de transport. Autour de cette ceinture existaient de nombreux centres villageois rattrapés par l'urbanisation le long des axes de communication. De nombreuses zones commerciales, industrielles, mais aussi portuaires sont présentes notamment aux abords du Rhin. Cette agglomération est caractérisée par un réseau hydrographique assez dense qui, en plus du Rhin, est traversé par l'Ill et la Bruche ainsi que de nombreux canaux, d'étangs et de gravières. Ce réseau s'associe à une présence arborée importante par des forêts inondables et de multiples ripisylves le long de ces cours d'eau. Plusieurs parcs urbains sont présents dans Strasbourg ainsi que de nombreux jardins ouvriers participant à la présence de végétation en ville. Le territoire agricole situé en périphérie est principalement composé de culture en openfield.

L'agglomération de Rennes Métropole est une agglomération de l'ouest de la France. Elle est constituée de 43 communes et s'étend sur un territoire de 705 km<sup>2</sup> (figure 2.4). En 2013, cette agglomération compte 426 502 habitants. Elle est souvent décrite comme une ville-archipel signifiant une structure générale polynucléaire avec de nombreux centres villageois entourés de territoire agricole. Elle s'organise autour d'un centre historique datant pour sa majeure partie de l'époque médiévale et de la ville classique du 18<sup>e</sup> siècle. Au cours du 19<sup>e</sup> et du début du 20<sup>e</sup> siècle de nombreux lotissements sont créés au sein de la ville de Rennes. Après la Seconde Guerre Mondiale, la ville connaît un fort développement de population grâce à la construction de nombreux quartiers d'habitat collectif de type grand ensemble. À partir de 1968, une ceinture verte autour de la ville de Rennes est mise en place afin de limiter le développement urbain ce qui lui donnera par la suite la dénomination de ville-archipel. Ce site est aussi marqué par un réseau hydrographique dense dont les deux principaux cours d'eau sont L'Ille et la Vilaine dont la confluence est dans le centre de Rennes et de larges zones humides située au sud-ouest de cette commune. Les espaces verts prennent une place importante dans la ville de Rennes et ceux publics occupent environ 17 % de la surface communale (AUDIAR, 2015). La végétation arborée apparaît beaucoup plus fragmentée que dans l'agglomération strasbourgeoise. Moins de grandes forêts sont présentes. Par contre, le territoire agricole est beaucoup plus bocager qu'à Strasbourg. Se trouvent ainsi un réseau de haies dense ainsi qu'un habitat dispersé beaucoup plus présent.

### **2.2.2 Bases de données géographiques disponibles**

Cette section présente les différentes bases de données disponibles qui peuvent être mobilisées pour la cartographie de la trame grise et de la trame verte. Nous présentons

tout d'abord les bases de données d'occupation du sol qui cartographient de manière continue l'espace des deux agglomérations (Corine Land Cover, Urban Atlas, base de données occupation des sols d'Alsace). Dans un second temps, les bases de données qui représentent les éléments des trames grises et vertes de manière non continue sont analysées. Cette section permet d'évaluer si ces bases répondent aux besoins pour l'analyse de la structure urbaine et la cartographie des trames grises et vertes.

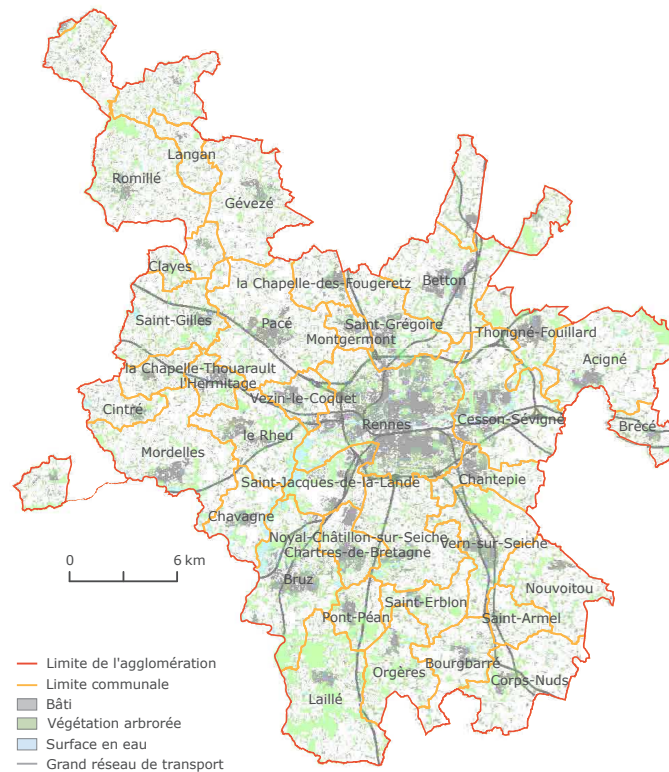


Figure 2.4 – Présentation du site de Rennes Métropole (source : BD TOPO® 2015, IGN)

### 2.2.2.1 Bases de données géographiques spatialement continues

Les bases présentées et évaluées dans cette section sont continues sur l'ensemble des sites d'études, c'est-à-dire que l'ensemble de l'espace est cartographié et que les polygones sont tous jointifs. Deux de ces bases sont disponibles pour Rennes et Strasbourg tandis que la dernière présentée est une base locale créée par la région Alsace.

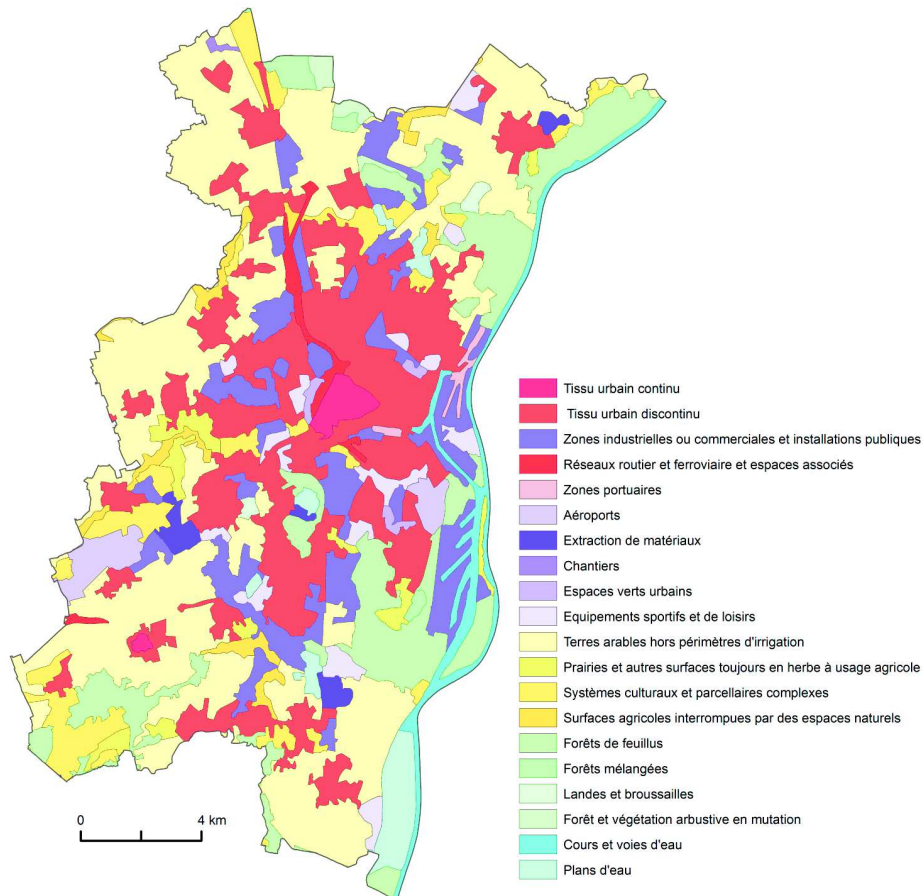
#### *Base de données Corine Land Cover – CLC*

La première, la couche CLC, est une base de données géographique d'occupation du sol produite pour 38 États européens. Quatre millésimes sont actuellement disponibles :



1990, 2000, 2006, 2012. En France, sa production est assurée par le Service de l'Observation et des Statistiques du Commissariat Général au Développement Durable et est mise gratuitement à disposition sur leur site. La nomenclature de CLC se décompose en quatre niveaux, le dernier est composé de 50 classes qui sont identiques pour l'ensemble des territoires concernés par cette base (EEA, 2007). Son échelle d'utilisation conseillée est de l'ordre du 1/100 000<sup>e</sup>, la surface minimum de représentation cartographique est de 25 ha et sa précision est de 25 m.

Cette base de données est produite par photo-interprétation à partir d'images provenant pour l'année 2012 des satellites IRS Resourcesat-1 et Resourcesat-2 livrées avec une résolution de 25 m, de la constellation RapidEye avec des images à une résolution de 25 m et complétées par des images du capteur IRS AWiFS à 60 m (Büttner et Kosztra, 2012). La figure 2.5 présente un extrait de CLC pour l'agglomération de Strasbourg. Seules les classes présentes sur ce territoire sont représentées dans la légende, mais l'ensemble de la nomenclature de cette base peut être trouvé en annexe B.1.

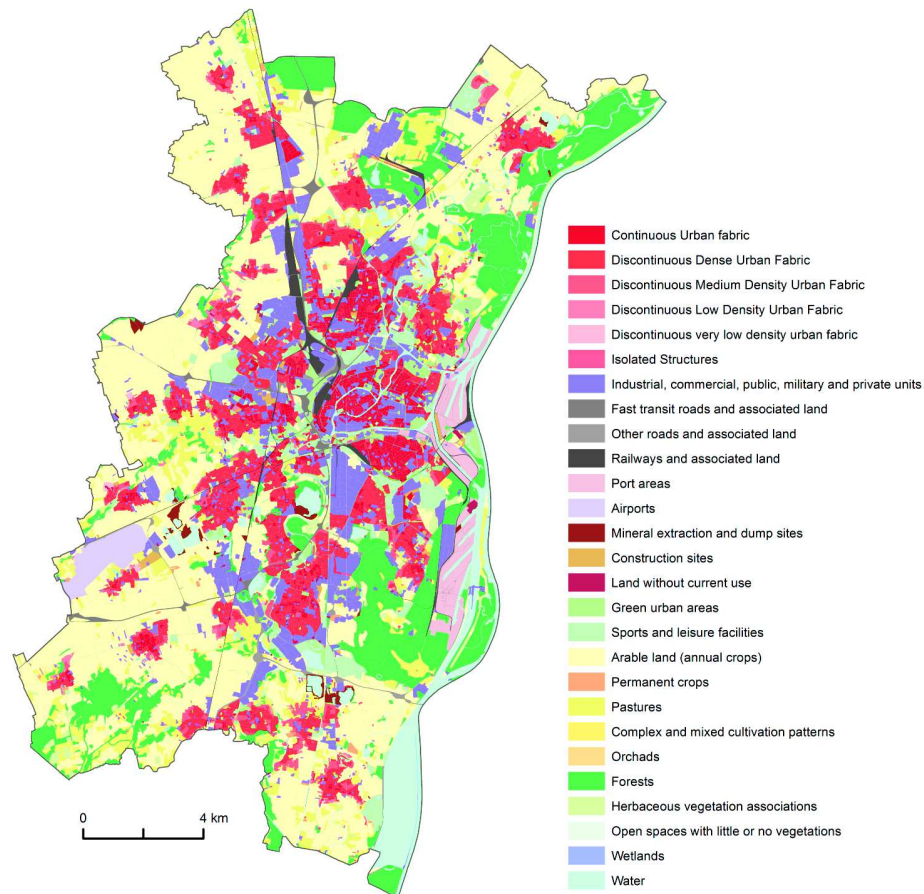


**Figure 2.5** – Extrait de la base de données Corine Land Cover sur l'Eurométropole de Strasbourg (source : CLC2012, MEDDE/CGDD/SOeS)

*Base de données Urban Atlas*

Urban Atlas est une deuxième base de données d'occupation des sols produite à l'échelle européenne. Elle est réalisée elle aussi par photo-interprétation (GMES, 2006). Son échelle d'utilisation est plus grande puisqu'elle est au 1/10 000<sup>e</sup> et sa précision géométrique est de 5 m. Cette base concerne uniquement les grandes aires urbaines de plus de 100 000 habitants, est disponible pour l'année 2006 et est en cours de réalisation pour l'année 2012. Pour ce millésime, les bases Urban Atlas sont disponibles pour les deux agglomérations de Strasbourg et de Rennes. Pour ces deux sites, les images utilisées pour la photo-interprétation sont des images SPOT5 avec une résolution spatiale de 2,5 m.

Sa nomenclature est aussi composée de quatre niveaux et comprend au total 20 classes (annexe B.2). La surface minimum de représentation est de 0,25 ha pour les classes urbaines, de 1 ha pour celles non urbaines et les routes dont la largeur est supérieure 10 m sont cartographiées. Urban Atlas détaille plus que CLC les tissus urbains discontinus qui sont divisés en quatre classes selon la densité de bâti (figure 2.6).



**Figure 2.6** – Extrait de la base de données Urban Atlas sur l’Eurométropole de Strasbourg (source : Urban Atlas 2012, Land Copernicus)

*Base de données d'occupation des sols d'Alsace – BDOCS*

La BDOCS est une base de données produite par la région Alsace et le réseau CIGAL (Coopération pour l'Information Géographique en Alsace) disponible librement sur leur site internet. Son utilisation est comme pour Urban Atlas conseillée au 1/10 000<sup>e</sup> et les surfaces minimums de représentation cartographique sont pour la plupart de 0,5 ha, mais peuvent être différentes selon les classes. Les routes représentées doivent par contre être au minimum de 12 m de large (annexe B.3). Cette base a une précision géométrique de 6 m et est déjà disponible pour trois dates : 2000, 2008 et 2012. Elle aussi est réalisée par photo-interprétation à partir de photographies aériennes ayant une résolution spatiale de 20 cm pour l'année 2012 (CIGAL, 2013).

La BDOCS se décompose elle aussi en quatre niveaux de nomenclature permettant des visualisations à différentes échelles. Le niveau 4 de cette nomenclature est beaucoup plus détaillé que les bases précédentes et comprend 55 classes. Pour les différentes classes urbaines, cette base est plus orientée utilisation du sol avec quatre classes décrivant l'habitat selon des critères de densité et le caractère collectif ou individuel des logements, mais surtout 23 classes décrivant les espaces urbains spécialisés et les grandes emprises. La figure 2.7 présente un extrait de la BDOCS pour l'agglomération de Strasbourg. La nomenclature a ici été simplifiée en 27 classes thématiques.

Le tableau 2.1 résume les principales caractéristiques des trois bases de données d'occupation du sol qui viennent d'être présentées.

**Tableau 2.1** – Caractéristiques des bases de données CLC, Urban Atlas et BDOCS

Caractéristiques	CLC	Urban Atlas	BDOCS
Couverture	Europe	Ville européenne de plus de 100 000 habitants	Alsace
Nomenclature	4 niveaux, 44 classes	4 niveaux, 20 classes	4 niveaux, 55 classes
Échelle	1/100 000 <sup>e</sup>	1/10 000 <sup>e</sup>	1/10 000 <sup>e</sup>
Surface minimum de représentation	25 ha	0,25 ha pour les classes urbaines 1 ha pour les classes non urbaines	Dépendante des classes d'occupation du sol (majoritairement 0,5 ha)
Précision	25 m	5 m	6 m
Données sources	IRS Resourcesat-1 et Resourcesat-2 (25 m), RapidEye (25 m)	Spot 5(2,5 m)	Photographies aériennes (20 cm)
Millésimes disponibles	1990, 2000, 2006, 2012	2006, 2012	2000, 2008 et 2012

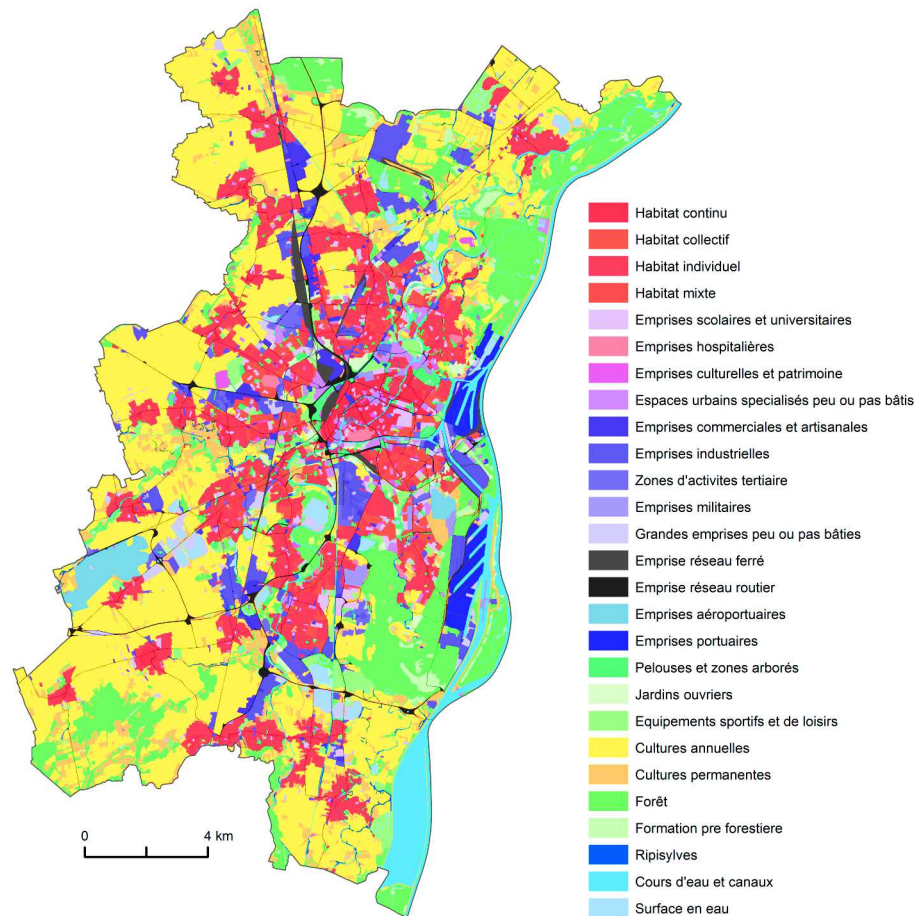


Figure 2.7 – Extrait de la BDOCS sur l'Eurométropole de Strasbourg (source : BDOCS2012 – CIGAL v2)

### 2.2.2.2 Bases de données géographiques non continues spatialement

À la différence des trois bases de données qui viennent d'être présentées, les suivantes ne sont pas continues sur l'ensemble de l'espace d'études. Ainsi des espaces sans aucune information sont présents, les polygones ne sont pas jointifs et peuvent, pour certaines bases de données, se superposer. Ces bases se concentrent sur des objets plus élémentaires et peuvent provenir de sources très différentes.

#### *La BD TOPO® de l'IGN*

La BD TOPO® est une des quatre composantes du Référentiel à Grande Échelle de l'IGN (RGE®). Il s'agit d'une base de données multi-thèmes qui en contient 10 : le réseau

roucier, le réseau ferroviaire, le réseau de transport d'énergie, le réseau hydrographique, les bâtiments, la végétation arborée, l'orographie, la structure administrative, les points d'activités ou d'intérêts et la toponymie. La mise à jour de cette base est régulière allant de six mois (routes importantes) à tous les trois ou cinq ans selon les dates de prises de vue aérienne (bâtiments, végétation). Plusieurs couches vectorielles sont disponibles pour chacun de ces thèmes qui utilisent une représentation surfacique, linéaire ou ponctuelle. Elle est réalisée par saisie photogrammétrique, à une précision métrique permettant son utilisation pour des échelles allant du 1/5000<sup>e</sup> au 1/50 000<sup>e</sup> (IGN, 2014b) et les informations sur la hauteur des bâtiments y sont intégrés.

Pour les bâtiments, la source est en cours de changement puisque depuis 2008 l'IGN utilise les informations provenant du cadastre. À terme ils ne seront donc plus issus d'un processus photogrammétrique permettant ainsi d'obtenir une représentation plus fine des bâtiments. Pour l'instant, les deux types de sources coexistent. Pour Rennes par exemple, la majorité des bâtiments provient du cadastre tandis qu'aucun bâtiment issu de cette source n'avait été intégré en 2015 pour l'agglomération de Strasbourg. Les bâtiments sont différenciés selon leur nature. Plusieurs couches sont ainsi disponibles : le bâti indifférencié principalement composés d'habitation, le bâti remarquable tel que les bâtiments religieux, administratifs ou sportifs et le bâti industriel.

La couche végétation est aussi en cours de changement puisqu'une couche multi-thèmes est en cours de réalisation dont l'objectif est de préciser la nature des surfaces arborées (feuillus, conifères, landes, etc.). La deuxième version de cette couche est disponible pour nos deux sites d'études. La surface minimum de représentation de cette couche est de 500 m<sup>2</sup> pour la végétation arborée et de 5000 m<sup>2</sup> pour les vignes et les vergers. Pour le thème hydrographie, les cours d'eau doivent avoir une largeur minimum de 7,5 m afin d'être intégré dans la BD TOPO® et les surfaces en eau doivent avoir une longueur minimum de 20 m. La figure 2.8 présente un extrait de cette base situé dans le centre-ville de Rennes.



Figure 2.8 – Extrait de la BD TOPO® du centre-ville de Rennes (source : BD TOPO® 2015, IGN)

### *la BD Parcellaire® de l'IGN*

Une base permettant de décrire les éléments gris est la BD Parcellaire®. Elle aussi est produite par l'IGN et constitue une seconde composante du RGE®. Cette base récemment créée fournit l'information cadastrale numérique à partir du géoréférencement du plan cadastral dématérialisé à une échelle qui dépend des relevés cadastraux, mais est au minimum équivalente à celle de la BD TOPO®. Elle représente ainsi les bâtiments, les parcelles, les divisions cadastrales et les divisions administratives selon deux modes de représentation, une forme image et une forme vecteur (IGN, 2010). L'ensemble du territoire français est d'ores et déjà existant sous forme image, mais celle vecteur issue de la numérisation de la forme image par les communes n'est pas encore terminée. Fin 2014, 82 % du territoire français est couvert par la BD Parcellaire® sous forme vecteur. Pour l'agglomération de Strasbourg, les données vecteurs sont encore manquantes et afin d'obtenir l'information cadastrale dans ce format, l'Eurométropole de Strasbourg nous a fourni la couche d'information décrivant les bâtiments et le parcellaire cadastral est disponible sous licence ouverte sur leur site internet. La figure 1.3 présentait plusieurs extraits de cette base. Pour le site de Rennes, l'ensemble des bâtiments et des parcelles a d'ores et déjà été numérisé.

### *La base de données OpenStreetMap*

Récemment sont apparues des bases de données géographiques collaboratives et libres dont la plus connue est OpenStreetMap (OSM). Cette base de données qui ciblait au départ les réseaux de transport, s'est fortement enrichie et comprend notamment en France l'empreinte au sol des bâtiments du cadastre, des points d'intérêt ainsi que des informations sur l'utilisation du sol. Les données présentes ne sont toutefois pas complètes et peuvent être très disparates selon les objets et les lieux. Les réseaux par exemple sont bien représentés dans l'OSM (Haklay, 2010) même si la qualité sémantique de leurs informations peut être hétérogène notamment en France (Girres et Touya, 2010). La représentation des bâtiments est par contre beaucoup moins précise et manque d'exhaustivité. En effet, tous ne sont pas représentés et leur géométrie n'est pas forcément bien modélisée (Hecht *et al.*, 2013b). Plus spécifiquement pour Rennes et pour Strasbourg, les données de l'OSM sont relativement complètes, autant les réseaux que les bâtiments, mais la qualité des informations attributaires est très hétérogène (Deniau, 2013 ; Vétillard, 2013).

### *La BD Forêt® de l'IGN*

Le référentiel géographique forestier (BD Forêt®) est une autre base thématique de l'IGN. Elle décrit les formations végétales forestières et naturelles par une description de la densité de couvert du peuplement, de sa composition et de l'essence dominante

permettant une nomenclature de 32 classes (IGN, 2014a). La surface minimale de représentation cartographique est 5000 m<sup>2</sup>. Elle est obtenue par photo-interprétation d'images en infrarouge couleur. De plus, la BD Forêt® est compatible avec la couche végétation de la BD TOPO® et sa mise à jour se fait tous les 3 à 5 ans selon la fréquence des prises de vue aérienne. La version 2 de cette base n'est pas encore disponible pour l'ensemble de la métropole, mais l'est pour les sites de Rennes et Strasbourg.

#### *Les couches High Resolution Layers*

L'Union Européenne et son agence spatiale ont récemment mis à disposition des cartes d'occupation du sol à haute résolution (*High Resolution Layers* – HRL). Ces différentes cartes sont composées de cinq couches : l'artificialisation, les forêts, les prairies permanentes, les zones humides et les surfaces en eau (Langanke, 2013). Elles sont issues du traitement automatique des mêmes images satellites que pour la base CLC avec une résolution spatiale de 20 m. Elle exprime pour chaque pixel de l'image le pourcentage de chacune des cinq classes d'occupation du sol considérée et est disponible pour 39 États. Une couche couvert forestier est aussi dérivée de cette analyse qui considère l'ensemble des pixels ayant une densité de végétation arborée supérieure à 10 % avec un seuil minimum de représentation de 5000 m<sup>2</sup>.

#### *Le Registre Parcellaire Graphique – RPG*

Le RPG est une base de données géographique permettant de décrire les îlots culturels au niveau européen. Cette base de donnée gratuite est disponible sur la plateforme ouverte des données publiques françaises ([data.gouv.fr](http://data.gouv.fr)), a été mise en place entre 2002 et 2005 et est administrée par l'Agence de Services et de Paiement (ASP). Elle décrit au 1/5000<sup>e</sup> les îlots agricoles et les types de culture associés. La quasi-totalité des parcelles y est représentée. Cette base est mise à jour annuellement par les agriculteurs. Toutefois, seuls les agriculteurs déclarant leurs îlots afin de bénéficier des aides de la Politique Agricole Commune sont concernés et cette base apparaît donc en partie incomplète. Les informations sur les cultures sont renseignées à l'îlot et ne correspondent donc pas nécessairement à la réalité. En effet la gestion des terres des exploitations se fait à une échelle inférieure, celle de la parcelle agricole. Lors de l'agrégation en îlot, la culture dont la surface est majoritaire est assignée.

En conclusion, les bases de données disponibles pour nos deux sites d'études ne proposent pas de cartographie à l'échelle des trames grises et vertes telle que nous les avons définies. Celles continues cartographient les grandes occupations et utilisations du sol, celles non continues les éléments urbains, agricoles ou naturels, mais aucune base ne se situe à une échelle intermédiaire qui est celle envisagée dans ce travail. En effet, la base CLC n'a pas une échelle adaptée à l'étude des tissus urbains qui sont plus de l'ordre du 5000<sup>e</sup>. Les deux autres ont une échelle qui s'en rapproche, mais souffrent d'autres défauts.

Urban Atlas pour sa description de la trame grise utilise une nomenclature qui se base sur la morphologie par la densité des emprises bâties. Elle reste donc dans l'opposition entre espaces construits et non construits qui nous semble insuffisante pour expliquer la forme des tissus urbains et pose toujours la question de la composition de ces espaces ouverts. La BDOCS est, elle, très précise, mais reste très attachée à l'utilisation du sol pour la plupart de ses classes urbaines et considère donc très peu la morphologie des tissus urbains. De plus cette base est locale et aucune base équivalente n'est disponible pour le site de Rennes. Cette analyse s'est fait qualitativement, mais d'autres, plus quantitatives, ont montré que ces différentes bases de données n'étaient pas adaptées pour cartographier la trame grise et la trame verte (Puissant, 2003 ; Hubert-Moy *et al.*, 2012 ; Combrette, 2016).

Les bases de données qui permettent de décrire les éléments constitutifs de la structure urbaine révèlent de fortes différences entre ceux gris et ceux verts. En effet, les bases concernant les éléments gris sont précises avec une échelle d'utilisation assez grande pour caractériser les tissus urbains. Pour les éléments verts par contre, les différentes bases sont beaucoup plus disparates. La couche végétation de la BD TOPO® est celle qui a la représentation la plus exhaustive de ces éléments, mais elle présente un taux de sous-détection assez important qui la rend difficilement exploitable à l'échelle communale et plus spécifiquement en milieu urbain où de nombreux arbres isolés de petite dimension sont omis (Hubert-Moy *et al.*, 2012). De plus, elle ne permet pas d'obtenir d'informations sur la répartition de la végétation herbacée. Les bases de données communales à grande échelle qui n'ont pas été présentées ici, car leurs conceptions sont très différentes selon les communes, s'attachent à identifier les arbres ou les espaces verts dont elles ont la gestion et ne sont donc pas exhaustives (annexe C). Les îlots agricoles issus du RPG par contre apparaissent être suffisamment bien renseignés pour constituer cette composante de la trame verte.

Ainsi pour la suite de cette recherche, il est nécessaire de créer une base de données végétation permettant la cartographie exhaustive des éléments arborés et herbacés. Cette section a montré le caractère privilégié des images pour la construction de carte d'occupation du sol. En milieu urbain et pour obtenir une cartographie de la végétation arborée exhaustive permettant donc d'identifier les arbres isolés, il est nécessaire d'utiliser des images avec une très bonne résolution spatiale puisque la résolution idéale pour détecter ces objets se situe entre 0,7 et 1 m (Tran *et al.*, 2011). La section suivante présente les méthodes permettant de cartographier la structure urbaine.

### **2.3 Cartographie de la structure urbaine : aspects méthodologiques**

Cette section fait état des méthodes actuellement mobilisées pour cartographier la structure urbaine. La section précédente a montré que la constitution de bases de données permettant de décrire la ville est généralement réalisée par photo-interprétation



d'images satellites ou aériennes. Cette méthode de production peut être très longue puisqu'elle peut atteindre 3 h/km<sup>2</sup> pour la digitalisation directe de l'image et 1 h/km<sup>2</sup> pour l'attribution de classe à des parcelles cadastrales (Robert et Autran, 2012). Elle présente donc un coût relativement important et peut être source de nombreuses erreurs liées à la lassitude qu'entraîne ce travail (Van Coillie *et al.*, 2014). Des méthodes plus automatiques ont été proposées afin d'extraire des informations à partir de données géographiques. Les méthodes d'analyse d'image et leurs évolutions sont tout d'abord présentées (§ 2.3.1). Comme la cartographie des tissus requiert un niveau d'abstraction plus grand que celle des éléments urbains puisqu'elle nécessite de prendre en compte l'organisation de ses éléments, la section 2.3.2 se focalise tout d'abord sur les méthodes proposées pour l'extraction de la végétation à partir d'images THRS et dans un deuxième temps les méthodes permettant de cartographier la trame grise à l'échelle du tissu urbain sont exposées dans la section 2.3.3.

### 2.3.1 Évolution des approches pour l'analyse d'image : du pixel à l'objet

Avec l'apparition des images satellites, des méthodes d'analyse d'images automatiques ou semi-automatiques ont été développées afin d'extraire de l'information de manière plus efficiente que par photo-interprétation. Ces méthodes issues des statistiques sont de deux types : les bases de règle et les méthodes de classification. Le principe du premier est de construire manuellement des règles, c'est-à-dire des seuillages successifs, à partir des caractéristiques des pixels jusqu'à obtenir les classes recherchées. Les méthodes de classification sont plus automatiques et ont pour objectif de regrouper les pixels appartenant à la même classe thématique telle que la végétation arborée ou les bâtiments. Elles peuvent être classées en deux grandes familles : les non supervisées (clustering) et les supervisées (Tan *et al.*, 2005). Le principe de la classification non supervisée est de regrouper les pixels dont les caractéristiques se ressemblent sans qu'aucune connaissance ne soit *a priori* utilisée pour constituer les groupes. C'est dans un deuxième temps qu'il est possible pour l'utilisateur de labéliser chacun des groupes créés pour obtenir des classes thématiques. À l'inverse, les méthodes de classification supervisée nécessitent que des exemples soient labélisés avec un minimum d'un par classe recherchée. Ces exemples composant le jeu d'entraînement vont permettre à la méthode d'apprendre et ainsi attribuer une classe aux exemples non labélisés selon les règles apprises qu'elle aura construites ou le degré de ressemblance aux exemples d'entraînement.

Traditionnellement ces méthodes d'analyse d'images se basaient uniquement sur les valeurs radiométriques, mais ces approches dites pixels ont été critiquées depuis la fin des années 90 (Fisher, 1997 ; Blaschke et Strobl, 2001). Avec l'amélioration de la résolution spatiale des images, les méthodes classiques utilisant uniquement les valeurs spectrales des pixels pour extraire des informations ont montré de nouvelles limites. Elle s'est accompagnée d'une diminution de la résolution spectrale des images dont les effets ont été remarqués très tôt, dès le passage de 80 à 30 m avec le lancement du satellite Landsat 4 (Irons *et al.*, 1985). L'arrivée des images avec une résolution métrique à partir

des années 2000 n'a fait qu'amplifier ce phénomène et les classifications ont eu tendance à être moins performantes (Key *et al.*, 2001). Cette dégradation de la résolution spectrale a donc entraîné une diminution de la variabilité inter-classes tandis que l'amélioration de la résolution spatiale a, elle, engendré une augmentation de la variabilité intra-classe (Herold *et al.*, 2003a). S'il devient possible de détecter des objets jusqu'alors invisibles en raison de leur taille, un objet se trouve composé de plusieurs pixels qui peuvent avoir des réponses spectrales très différentes, ce qui augmente ainsi fortement son hétérogénéité. De plus, le bruit au sein d'un pixel et de ses réponses spectrales est plus important du fait qu'il subit de manière importante l'influence de ses voisins (Blaschke et Strobl, 2001). Tant que l'objet géographique reste plus petit ou de dimension proche de la taille du pixel, il s'agit d'identification et l'approche pixel peut être justifiée, mais quand l'objet d'intérêt devient composé d'une multitude de pixels, analyser sa configuration spatiale peut être plus informatif ou tout du moins complémentaire à son analyse radiométrique (Puissant, 2014). Pour un photointerprète, les valeurs radiométriques ne sont pas les seules informations prises en compte, mais la texture, la forme, le contexte sont tout autant de clés pour analyser les images (Puissant, 2003).

Un changement de niveau est donc apparu pour l'analyse d'images qui ne prendrait plus le pixel comme unité élémentaire, mais un objet (ou segment) donnant ainsi naissance aux approches orientées-objet aussi appelées OBIA (*Object-Based Image Analysis* – Hay *et al.*, 1996 ; Baatz et Schäpe, 2000). Cette approche est devenue de plus en plus populaire pendant les années 2000 comme le souligne le nombre de publications utilisant ce type d'analyse (Blaschke, 2010). Une méthode orientée-objet peut être décomposée en trois étapes importantes (figure 2.9) : la segmentation de l'image, la caractérisation des segments, et la classification de ces segments basés sur leur vecteur descriptif (Puissant, 2003). Un objet, résultant de cette première étape, est un regroupement de pixels contigus qui sont homogènes. Cette homogénéité relative est définie par la méthode de segmentation avec la possibilité, pour chacune, de faire varier un paramètre permettant de contrôler la taille des objets. Cette étape de segmentation de l'image et les méthodes souvent utilisées en télédétection sont explicitées plus en détail dans le chapitre 3. Pour chacun des objets de l'image, il est possible de calculer un grand nombre d'attributs pour les caractériser qui peuvent être basés sur les réponses spectrales de l'image, sa texture, mais elle innove en permettant le calcul d'attributs de forme des objets ainsi que des attributs contextuels tels que des relations d'inclusion ou d'adjacence. L'un des intérêts de ce type d'approche est de mieux correspondre à la manière dont nous abordons le paysage : de grandes zones composées d'une mosaïque d'objets géographiques qu'il est possible d'obtenir en faisant varier la taille des segments (Hay et Castilla, 2008).

L'approche orientée-objet a montré une amélioration importante des performances pour la cartographie des espaces urbains par rapport à une approche pixel (Myint *et al.*, 2011 ; Pu *et al.*, 2011). Pour l'extraction de végétation à partir d'image THRS, cette amélioration a aussi pu être constatée (Cleve *et al.*, 2008) et cette approche est ainsi devenue de plus en plus populaire pour cette application (Pu et Landry, 2012). La section suivante présente les méthodes qui ont été proposées pour l'extraction de la végétation pas seulement

avec des approches orientées-objet afin d'apprécier les différentes solutions ayant pour objectif d'en améliorer les performances. Cet état de l'art se concentre sur les méthodes développées pour l'analyse d'image THRS, mais présente tout d'abord d'autres types de données qui peuvent être utilisées pour cartographier la végétation.

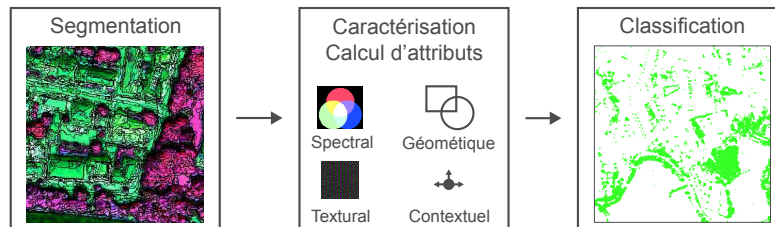


Figure 2.9 – Les étapes d'une démarche d'analyse d'image orientée-objet

### 2.3.2 Cartographie des éléments verts

De nombreuses méthodes ont été développées pour cartographier la végétation qui reposent sur une variété de types de données importante. Herrault (2015) propose une méthode d'extraction adaptée aux cartes anciennes. Ces cartes sont celles d'État Major datant du 19<sup>e</sup> siècle et sont comparées avec des cartes de végétation récentes afin de retracer l'évolution des forêts de Gascogne dans le sud de la France.

L'utilisation de capteurs actifs est une source de données qui peut être utilisée pour identifier la végétation. Betbeder *et al.* (2014) proposent ainsi une méthode pour identifier le réseau de haies à partir d'une image radar TerraSAR-X. Les performances obtenues pour l'extraction de la végétation sont bonnes et ce type de capteur permet aussi à ces auteurs de caractériser la structure interne des haies afin d'identifier la fragmentation de leur canopée. Les données obtenues grâce au Lidar sont aussi particulièrement bien adaptées pour l'analyse de la végétation tant dans des milieux naturels (Vannier et Hubert-Moy, 2014) que des milieux urbanisés (Höfle *et al.*, 2012). L'amélioration des technologies Lidar permet aujourd'hui d'enregistrer l'ensemble du retour de l'onde émise (*Full-Waveform Lidar*). En plus des différents retours qui permettent de mesurer la distance, l'intensité du signal est aussi enregistrée et apparaît être aujourd'hui une grande avancée pour l'analyse de la végétation puisqu'elle permet de réaliser des classifications par espèce (Reitberger *et al.*, 2008 ; Heinzl et Koch, 2011), mais aussi d'estimer la biomasse de chaque arbre (Popescu, 2007).

Les images issues de capteurs passifs de télédétection sont des données privilégiées pour l'identification de la végétation. Des méthodes d'extraction ont été proposées reposant sur des images aériennes ou satellites (Tansey *et al.*, 2009 ; Li et Shao, 2013), satellites (Sheeren *et al.*, 2009 ; Vannier et Hubert-Moy, 2014), hyperspectrales (Wania et Weber, 2007 ; Ghosh *et al.*, 2014 ; Dalponte *et al.*, 2014) ou des photographies terrestres (Li *et al.*, 2015). Pour ces données, la résolution spatiale des images apparaît être un facteur à

prendre en compte pour obtenir une cartographie précise. En effet, Vannier et Hubert-Moy (2014) ont comparé plusieurs images différentes pour l'identification des haies : une photographie aérienne (0,5 m), des images satellites KOMPSAT II (1 m), SPOT 5 (5 m et 10 m), ASTER VNIR (15 m) et IRS P6 LISS III (23 m). Leurs résultats montrent tout d'abord que la résolution a une influence importante puisque les performances les meilleures sont obtenues avec les images qui ont la meilleure résolution. Mais leurs résultats montrent aussi que la résolution spectrale est très importante puisqu'avec une résolution spatiale meilleure la photographie aérienne ne permet pas d'obtenir d'aussi bons résultats qu'avec l'image Kompsat II qui elle dispose d'une bande dans le proche infrarouge.

Le proche infrarouge est une longueur d'onde importante pour l'identification de la végétation. La chlorophylle a en effet des propriétés optiques originales puisqu'elle absorbe beaucoup la lumière dans les longueurs d'ondes du rouge au contraire du proche infrarouge ce qui permet une bonne différenciation entre les plantes et les autres matériaux. Ces propriétés ont permis la création de nombreux indices synthétiques de végétation qui combine ces deux longueurs d'onde (Bannari *et al.*, 1995) permettant une meilleure extraction de ces éléments (Van Delm et Gulinck, 2011). Mais le proche infrarouge n'est pas le seul domaine spectral permettant d'obtenir de meilleures classifications puisque de nombreuses recherches ont montré qu'une augmentation de la résolution spectrale permettait d'obtenir de meilleures performances notamment pour l'identification de différentes espèces de végétation arborée telle qu'avec l'utilisation d'images satellites WorldView-2 composée de huit bandes (Pu et Landry, 2012 ; Fernandes *et al.*, 2014 ; Sellin *et al.*, 2015) ou d'images hyperspectrales (Dalponte *et al.*, 2012).

Une des méthodes couramment utilisées afin d'identifier la végétation à partir d'une image est le recours à une base de règle. Afin de minimiser le nombre de règles, les indices de végétation sont couramment utilisés (Pham *et al.*, 2011 ; Van Delm et Gulinck, 2011). Si les méthodes à base de règles ont l'avantage d'être facilement compréhensibles et montrent de bons résultats dans plusieurs études, elles ont elles aussi leurs limites. Tout d'abord, les règles sont souvent trouvées de manière empirique, ce qui peut représenter un travail relativement long et soumis aux connaissances et à la subjectivité de l'utilisateur qui les crée. Seules certaines caractéristiques, par exemple spectrales, qu'il pense déterminantes seront utilisées (Belgiu *et al.*, 2014). De plus, il est difficile de transposer une base de règles à d'autres images, qu'elles proviennent ou non de capteurs différents, qu'elles se concentrent ou non sur de mêmes espaces géographiques, et il est donc nécessaire de l'adapter rendant ainsi difficilement reproductibles de telles méthodes (Tathiri *et al.*, 2014 ; Hussain et Shan, 2015). D'autres méthodes, plus reproductibles et plus automatisées, ont été développées : les méthodes de classification.

Pour l'extraction de la végétation, les méthodes de classifications supervisées sont celles le plus souvent utilisées, mais quelques recherches ont déjà employé des méthodes non supervisées. Nichol et Wong (2007) par exemple utilisent la méthode de clustering Isodata afin d'extraire la végétation urbaine en deux classes, arborée et herbacée, tandis que Youjing et Hengtong (2007) vont utiliser la méthode des k-moyennes pour discrétiser

six classes de végétation. Les approches utilisant des méthodes supervisées sont très variées pour extraire la végétation. Nous pouvons nommer la méthode des plus proches voisins (Mathieu *et al.*, 2007), celle des réseaux de neurones (Zhang et Xie, 2012), les arbres de décision (Tooke *et al.*, 2009 ; Zhang *et al.*, 2010), les ensembles de classifieurs (Feng *et al.*, 2015) ou encore les machines à vecteur de support (SVM – *Support Vector Machine*; Secord et Zakhor, 2007). Ces algorithmes ne sont pas explicités ici, mais certains d'entre eux sont présentés au cours de cette thèse et plus de détails peuvent être trouvés dans les ouvrages de Tan *et al.* (2005) ou de Cornuéjols et Miclet (2010).

L'arrivée des images THRS a révélé de nouveaux défis (§ 2.3.1) notamment pour la végétation arborée, car les réponses spectrales d'un arbre sont très influencées par les irrégularités de sa surface entraînant des différences d'illumination et des effets d'ombre qui nuisent beaucoup aux performances obtenues par des approches pixels (Quackenbush *et al.*, 2000). Afin d'améliorer les précisions de l'extraction de la végétation à partir d'images THRS plusieurs solutions ont été mises en pratique. Une première est d'intégrer des mesures de textures afin d'obtenir de nouvelles informations sur la structure de l'image et des objets d'intérêt. Cette méthode a permis d'améliorer les performances des classifications autant pour la cartographie de l'occupation du sol en milieu urbain que pour identifier les éléments arborés dans les espaces agricoles (Puissant *et al.*, 2005 ; Sheeren *et al.*, 2009). Pour extraire la texture, une des solutions est d'utiliser les matrices de cooccurrence en niveau de gris (GLCM – *Grey Level Coocurrence matrix*; Haralick *et al.*, 1973) pour obtenir de nouvelles bandes à ajouter à celles spectrales de l'image avant de réaliser la classification. Sheeren *et al.* (2009) par exemple, calculent 96 nouvelles bandes en faisant varier taille et orientation pour le calcul des GLCM et montrent que leur utilisation permet d'améliorer les performances pour l'identification des haies.

Afin d'obtenir des informations sur la structure des objets présents dans une image, le recours à des opérateurs issus de la morphologie mathématique est possible. Tuia *et al.* (2009a) utilisent différents filtres tels que l'ouverture, la fermeture, ouverture et fermeture par reconstruction, et montrent que les deux premiers permettent, à partir d'une image panchromatique Quickbird, une amélioration importante des performances des classifications réalisées avec SVM. Toujours sur le même type d'image, Dalla Mura *et al.* (2010) développent des profils d'attributs morphologiques qui permettent d'obtenir des informations sur la géométrie des objets de l'image.

Une autre solution, particulièrement adaptée à l'extraction de la végétation, est de ne plus réaliser les classifications sur une seule image, mais d'en utiliser plusieurs. L'objectif est ici de contrebalancer la faible résolution spectrale des images pour l'identification de la végétation en prenant en compte plusieurs stades phénologiques et ainsi permettre une meilleure discrimination des différentes formes végétales (Key *et al.*, 2001 ; Xie *et al.*, 2008). Cette méthode est très souvent utilisée pour l'identification des types de culture des parcelles agricoles (Guerschman *et al.*, 2003 ; Vaudour *et al.*, 2015), mais aussi pour identifier différentes classes de végétation arborée (Carleer *et al.*, 2005 ; Tigges *et al.*, 2013). Quel que soit le capteur utilisé, cette solution permet d'améliorer significativement les performances (Feret *et al.*, 2015).

L'approche orientée-objet a montré une amélioration importante des performances pour la cartographie de la végétation que ce soit pour des images multispectrales ou des photographies aériennes (Cleve *et al.*, 2008 ; Zhang *et al.*, 2010 ; Duro *et al.*, 2012a) même si des informations sur la texture étaient considérées (Tan et Wang, 2009). La segmentation est une étape très importante de cette analyse qui peut influencer fortement les performances des classifications (Stumpf et Kerle, 2011) et plusieurs stratégies peuvent être mises en place pour identifier la végétation à partir d'images.

Une première est d'utiliser une seule segmentation dont les paramètres sont souvent déterminés par une procédure d'essais et erreurs et de réaliser une classification sur la base des attributs calculés pour chaque segment (Walker et Briggs, 2007 ; Tan et Wang, 2009 ; Hofmann *et al.*, 2011). Zhang *et al.* (2010) utilisent aussi une segmentation unique, mais réalisent au préalable un seuillage au niveau du pixel sur un indice de végétation afin de segmenter dans un deuxième temps uniquement les pixels extraits. Mais le processus de segmentation permet d'obtenir une représentation de l'espace à plusieurs niveaux grâce à la réalisation d'une hiérarchie de segmentation qui peut permettre d'obtenir de meilleurs résultats (Youjing et Hengtong, 2007). Chaque niveau de segmentation peut être analysé simultanément en attribuant aux objets les plus petits les caractéristiques de ceux supérieurs (Duro *et al.*, 2012b), mais il est aussi possible d'analyser chacun des niveaux successivement afin d'obtenir des caractéristiques thématiques et non plus uniquement attributaires pour les niveaux suivants. Deux types de stratégie sont possibles : une ascendante où les segmentations les plus fines sont utilisées pour enrichir les plus grossières ou, à l'inverse, une stratégie descendante qui s'intéresse tout d'abord aux objets de niveau supérieur.

Avec une approche ascendante et trois niveaux de segmentation, Schöpfer *et al.* (2005) extraient tout d'abord les éléments verts puis affinent, avec les segmentations supérieures, les classes d'occupation du sol afin d'obtenir les arbres, la végétation herbacée et les parcelles agricoles. D'autres auteurs tels que Vannier et Hubert-Moy (2014) utilisent cette même stratégie, mais pour obtenir une typologie plus morphologique au niveau supérieur et identifient tout d'abord les arbres et les parcelles agricoles afin d'extraire les haies. Cet objectif d'obtenir des informations sur la forme des structures végétales est aussi considéré par Van Delm et Gulinck (2011) qui, avec une série de cinq segmentations, extraient sept classes : les haies de petites et grandes largeurs, les groupes d'arbres, les arbres isolés, la végétation mixte, herbacée et ouverte.

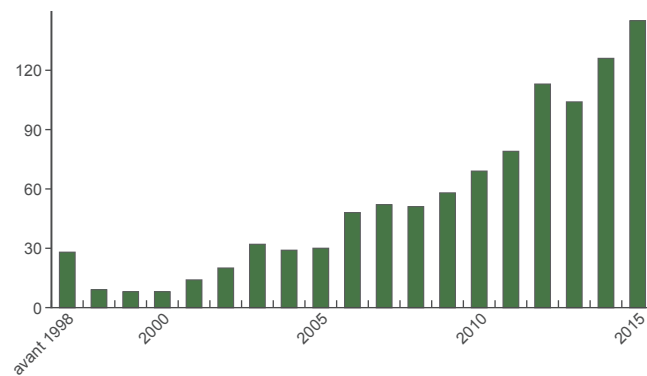
Mathieu *et al.* (2007) utilisent une approche descendante pour identifier plusieurs types de jardins privés. La première segmentation, celle produisant les objets les plus gros, leur permet d'extraire les zones d'habitats résidentielles et les segmentations plus fines ont pour objectifs de différencier les configurations des jardins pour ce type d'habitat. Plusieurs méthodologies utilisant une stratégie descendante ont aussi été développées pour extraire des objets qui deviennent identifiables au fur et à mesure que les segments sont plus petits (Youjing et Hengtong, 2007 ; Cleve *et al.*, 2008). Pham et He (2008) par exemple utilisent 3 niveaux de segmentation pour extraire au plus grossier les groupements d'arbres, au second les alignements et arbres isolés au troisième.

La diversité des sources et des types de données géographiques peuvent aussi être des avantages pour cartographier les éléments verts. Ainsi Schöpfer *et al.* (2005) utilisent une photographie aérienne couleur à laquelle est associée une image Aster avec une résolution spatiale plus faible, mais qui permet à ces auteurs de profiter d'une bande dans le proche infrarouge pour construire un indice de végétation. Peters *et al.* (2011) utilisent deux types de données différentes : des données optiques issues d'une photographie aérienne et des données radar provenant des capteurs de RAMSES SAR et TerraSAR-X. La hauteur est une information qui peut se révéler essentielle pour différencier végétation haute et basse. Obtenue par des relevés Lidar, elle a ainsi souvent été utilisée que ce soient en synergie avec des photographies aériennes (Secord et Zakhor, 2007), des images multispectrales ou hyperspectrales (Dalponte *et al.*, 2012). Toutefois, ces données issues de Lidar ne sont pas forcément disponibles. Iovan *et al.* (2007) ont ainsi proposé d'utiliser une méthode à partir de photographies aériennes permettant d'extraire la végétation et dans un second temps de construire un modèle numérique de surface par photogrammétrie et ainsi séparer végétation arborée et herbacée.

Cet état de l'art a montré la diversité des données et des méthodes qui peuvent être utilisées afin de cartographier la végétation qui, spécifiquement dans les espaces urbains, reste un domaine de recherche encore très traité (figure 2.10). Les images THRS sont des données qui permettent d'obtenir avec une échelle très grande la végétation urbaine notamment lorsqu'elles sont associées à des méthodes orientées-objet. Toutefois, ces méthodes ne sont pas sans limites. La segmentation est une étape qui a beaucoup d'influence sur les performances et qui est souvent réalisée par des procédures d'essais et erreurs et donc soumise aux représentations de l'utilisateur et peut être longue à mettre en place (Lefebvre *et al.*, 2011 ; Martha *et al.*, 2011). De plus, une grande partie des méthodes de cet état de l'art utilisent des classifications supervisées. Elles sont en effet très bien adaptées à l'identification d'objets spécifiques, mais nécessitent en contrepartie des exemples d'apprentissage qui permettent de considérer l'ensemble de la diversité des types d'objets recherchés (Chen et Stow, 2002 ; Foody et Mathur, 2004) dont la collecte, que ce soit par des relevés terrain ou par photo-interprétation, représente un coût important. De nombreux défis restent donc d'actualité afin de rendre plus automatiques et plus opérationnelles l'extraction de la végétation urbaine à partir d'images aériennes ou satellites.

### 2.3.3 Cartographie des tissus urbains

La cartographie des tissus est en général différente de celles de la végétation, car elle nécessite un niveau d'abstraction supérieur (Banzhaf et Höfer, 2008). En effet, c'est ici l'organisation des éléments entre eux qui est à prendre en compte contrairement à la cartographie des éléments verts qui sont eux de l'identification. Toutefois, certaines méthodologies tentent d'extraire directement les tissus urbains à partir d'images satellites HRS en passant par des classifications au niveau du pixel, par exemple avec des images Spot ou Landsat (Mesev *et al.*, 1995 ; Weber *et al.*, 2005 ; Van de Voorde *et al.*, 2011 ;



**Figure 2.10** – Évolution du nombre de publications traitant de la cartographie de la végétation urbaine (Données extraites de la base *Web of science*<sup>TM</sup> avec la recherche "urban" and "vegetation" and "mapping")

Bechtel *et al.*, 2015). Ridd (1995) a proposé une méthode basée sur la combinaison de trois indicateurs : la part de végétation, de surface artificialisée et de sol (*Vegetation – Impervious surface – Soil, VIS*). Ainsi à partir d’image HRS, des processus de démixage des pixels ont été appliqués afin d’obtenir la part de chacune de ces classes pour chaque pixel et d’utiliser ensuite la méthode de Rider pour obtenir une cartographie des espaces urbains (Phinn *et al.*, 2002 ; Setiawan *et al.*, 2006). Plus récemment Bechtel *et al.* (2015) cartographient les LCZ avec des images Landsat à partir d’une analyse pixel et d’une classification supervisée. Leur objectif est de proposer une plateforme permettant à chacun de réaliser cette cartographie pour des espaces urbains connus afin de créer une base de données mondiale des formes et fonctions de la ville (*wudapt.org*). Cette plateforme propose de construire par photo-interprétation les données d’entraînement à partir de Google Map pour pouvoir ensuite les utiliser pour la classification des images Landsat avec la méthode *Random Forest*. Bien que ce ne soit pas l’objectif de leur étude, Mathieu *et al.* (2007) utilise une image THRS qu’ils segmentent. Ils extraient ensuite l’utilisation du sol uniquement à partir des caractéristiques spectrales de l’image. Leur typologie est très simple (résidentiel, industriel et commercial, route et eau), mais ne passe pas par la description préalable des éléments constitutifs de la trame grise. L’ensemble des recherches qui viennent d’être présentées obtient une cartographie des espaces urbains directement à partir de l’image. Une grande majorité toutefois utilise les éléments et leurs configurations pour identifier les tissus urbains. Ces éléments peuvent provenir de différentes sources et peuvent être complétés par de nombreuses données complémentaires.

Les bases de données topographiques telles que la BD TOPO® sont des sources de données privilégiées pour la cartographie de la trame grise. Ainsi de nombreux auteurs utilisent l’empreinte au sol des bâtiments (Steiniger *et al.*, 2008 ; Colaninno *et al.*, 2011 ; Hamaina *et al.*, 2012) qui peut être complété par d’autres éléments issus des mêmes bases comme les routes (Puissant *et al.*, 2011 ; Hussain *et al.*, 2012) et la végétation (Long et Kergomard, 2005 ; Bernabé *et al.*, 2013). Toutefois ce type de données n’est pas



toujours disponible et Hecht *et al.* (2013a) proposent d'extraire les bâtiments à partir de cartes topographiques numérisées avant de cartographier les tissus.

L'image, notamment THRS, permet d'extraire les objets élémentaires ce qui constitue une première étape pour de nombreux auteurs pour identifier les tissus (Bauer et Steinnocher, 2001 ; Herold *et al.*, 2003b ; Novack et Kux, 2010 ; Walde *et al.*, 2014). Pour améliorer la description des éléments constitutifs du tissu, l'association de données provenant de sources différentes peut être une solution. Uniquement à partir d'images satellites, Wemmert *et al.* (2009) et Kurtz *et al.* (2012) profitent de différentes résolutions spatiales avec l'idée qu'à chaque résolution correspond un niveau sémantique qui permet d'obtenir des informations complémentaires. Chaque image permet donc d'enrichir les autres ce qui améliore ainsi l'identification des tissus.

La hauteur est une information essentielle pour différencier les tissus urbains et particulièrement celle des bâtiments (Hecht *et al.*, 2015). Cette information peut être disponible dans les bases de données topographiques comme dans la BD TOPO® (Hamaina *et al.*, 2012 ; Bonhomme, 2013), mais nécessitent souvent d'être obtenus avec d'autres données par exemple par des relevés Lidar (Barnsley *et al.*, 2003 ; Hermosilla *et al.*, 2014). La constitution de bases de données décrivant les éléments urbains constitue la première étape pour obtenir une cartographie des tissus urbains. La deuxième étape est donc la généralisation cartographique qui passe souvent par le choix d'une unité d'analyse pour étudier les relations entre ces éléments et leur organisation.

### **2.3.3.1 Unités d'analyse pour la cartographie des tissus urbains**

La cartographie des tissus nécessite généralement l'utilisation d'une unité d'analyse et plusieurs peuvent être utilisées. Elles sont issues de bases de données existantes ou créées à partir des données. Un état de l'art des découpages mobilisés pour cartographier les tissus urbains ou les méthodes permettant de les reconstruire est réalisé dans cette section.

#### *À partir de découpages existants*

En France, les IRIS (Îlots Regroupés pour l'Information Statistique) produits par l'INSEE (Institut National de la Statistique et des Études Économiques) peuvent être utilisés afin d'étudier la morphologie de la ville. Ils correspondent à un découpage du territoire infra-communal où sont diffusés les résultats du recensement de la population. Les études qui utilisent les IRIS ne recherchent pas à construire réellement les tissus urbains, mais plutôt à étudier à cette échelle les relations entre la forme urbaine et des données statistiques, dont l'IRIS est souvent la maille la plus fine disponible. Ainsi Rhein et Palibrk (2014) utilise ce découpage afin de déterminer des relations entre les formes urbaines, et plus spécifiquement celles de la végétation, et des caractéristiques sociodémographiques pour la ville de Paris. Leur objectif est de comprendre quelles sont

les formes les plus touchées par la gentrification et montrent que ce processus touche tout autant les IRIS où la part que représente la végétation est importante que celle où elle très faible. Ce même découpage est aussi utilisé par Tran *et al.* (2014) afin de vérifier l'influence de la morphologie urbaine sur la mobilité des habitants dans l'agglomération de Strasbourg. Cette recherche cherche à mettre en lumière l'influence sur les modes de transports et leur fréquence que peut avoir la morphologie. Les auteurs de ces deux études remarquent le caractère trop agrégé de ce découpage en IRIS, mais qui peut être difficilement affiné du fait de la disponibilité des indicateurs statistiques.

À une échelle plus grande, les blocs urbains quand ils sont disponibles permettent une cartographie assez fine de ces espaces. En Allemagne par exemple la base AKTIS® Basic DLM permet un découpage en bloc et a souvent été utilisée pour la classification des tissus (Banzhaf et Höfer, 2008 ; Lindner *et al.*, 2011 ; Walde *et al.*, 2014). Cette donnée contient aussi des informations sur l'utilisation du sol avec une nomenclature simple pour les blocs urbains : résidentiel, industriel et commercial, mixte et espace urbain spécialisé. Au travers de deux études, Hecht *et al.* (2013a, 2015) utilisent cette information qu'ils complètent par des indicateurs sur la forme des bâtiments afin d'affiner la typologie initiale de ces blocs. Une classification est tout d'abord réalisée à l'échelle des bâtiments, la classe attribuée au bloc est celle majoritaire des bâtiments.

Les données cadastrales sont des sources souvent mobilisées pour obtenir une unité d'analyse permettant la cartographie des tissus. Hu et Wang (2013) réalisent une classification au niveau de la parcelle cadastrale pour la cartographie de l'utilisation du sol du centre-ville d'Austin au Texas à partir des bâtiments et de leur hauteur ainsi que des autres surfaces artificialisées extraites d'une photographie aérienne couleur infrarouge. Une étude similaire menée par Wu *et al.* (2007) a montré que les indicateurs dérivés des surfaces artificialisées sont moins importants pour la séparabilité des tissus que ceux dérivés des bâtiments ou des parcelles cadastrales et soulignent le fait qu'il serait nécessaire de mieux caractériser ces surfaces pour connaître leur utilisation (parkings, routes, etc.) et ainsi améliorer la classification des parcelles.

Hermosilla *et al.* (2012) utilisent aussi les parcelles cadastrales afin de calculer une série d'indicateurs qui rendent compte de la forme des bâtiments et de la végétation ainsi que leurs relations topologiques. Ces parcelles sont ensuite agrégées à l'îlot et de nouveaux indicateurs sont construits à cette échelle pour modéliser les connectivités entre les parcelles. La cartographie des tissus est réalisée au niveau de l'îlot grâce à une classification supervisée par arbre de décision. Dans une étude plus récente, ces auteurs considèrent les mêmes données et indicateurs, mais utilisent aussi les espaces libres entre les îlots qui correspondent à la rue pour mesurer sa forme et ses relations avec les bâtiments et la végétation (Hermosilla *et al.*, 2014). L'îlot construit à partir du regroupement des parcelles cadastrales est aussi le découpage utilisé par l'APUR (2007) pour évaluer les besoins en énergie des bâtiments. Une limite à l'utilisation de ce découpage est qu'il ne permet pas d'obtenir une cartographie continue de la trame grise. Plumejeaud-Perreau *et al.* (2015) dont l'objectif est de cartographier les LCZ à partir de données topographique et cadastrale utilisent une tessellation de Voronoï à partir des

îlots pour remplir les vides qu'implique cette unité d'analyse. Les découpages existants ne sont pas toujours disponibles ou pas toujours adaptés à l'échelle de cartographie recherchée. De nombreuses recherches ont donc proposé des solutions diverses pour pouvoir étudier les tissus urbains.

### *Découpages réguliers de l'espace*

Quand aucun découpage n'est disponible, une première solution est de créer un maillage de l'espace régulier par un carroyage. Le carroyage est utilisé depuis les années 1980 afin de décrire et de modéliser l'espace (CERTU, 2011) dans de très nombreux domaines d'application pour quantifier les milieux naturels (Birch *et al.*, 2007) ou les espaces urbains (Long et Kergomard, 2005). Le carroyage correspond à une partition arbitraire de l'espace pour laquelle l'unité de base est le carreau. Cette représentation particulière de l'espace offre de nombreux avantages. Tout d'abord, il permet de s'affranchir des découpages existants, tels que les limites administratives ou statistiques (IRIS) qui, pour certains, peuvent être changeants dans le temps. Il permet ainsi de réaliser des analyses temporelles sur de longues périodes. Cette maille de base est simple et permet donc une intégration aisée dans de nombreux modèles, de climatologie urbaine par exemple (Bonhomme, 2013), d'évolution d'occupation des sols (Claggett *et al.*, 2004) ou de déplacements d'espèces (BenDor *et al.*, 2006).

Dans la littérature, différentes tailles de maille sont utilisées pour étudier et cartographier la trame grise. Antoni (2001) utilisent un maillage carré de 50 m afin de modéliser l'étalement urbain tandis que Bonhomme (2013) utilise une maille de 250 m pour proposer des scénarios prospectifs d'évolution de l'îlot de chaleur urbain pour Paris et Toulouse. Avec une unité d'analyse de 200 m, Long et Kergomard (2005) décrivent les tissus urbains de Marseille à partir de la BD TOPO® pour qu'ils puissent être réutilisés dans des applications de météorologie locale. La même taille de maille est utilisée par Poncet-Montanges *et al.* (2015) pour la cartographie de la structure urbaine de Munich à partir de données multi-sources : deux images à haute et très haute résolution spatiale, les routes et les points d'intérêts issus de l'OSM.

Toutefois, une des limites de ce type de représentation de l'espace est le choix d'une taille de maille adaptée (CERTU, 2011). Ce choix doit être guidé par trois questions : l'échelle du travail, les informations qu'il faut conserver et l'adéquation entre la nouvelle maille créée et celles déjà utilisées sur le même territoire. Le temps de traitements peut aussi devenir important lorsque l'étendue de la zone d'étude est grande. Pour la cartographie des tissus urbains, il est nécessaire de prendre en compte plusieurs effets liés à la taille du carreau. Elle peut en effet avoir une influence importante sur les indicateurs permettant de les identifier (Long et Kergomard, 2005 ; Bonhomme, 2013). Choisir une maille de trop petite dimension ne permet pas de décrire fidèlement un tissu urbain puisqu'il ne va en considérer qu'une partie et ainsi perdre des informations permettant de décrire l'arrangement spatial des objets entre eux. L'utilisation d'une maille trop

petite peut aussi, dans les espaces urbains denses, fractionner de manière importante les objets géographiques urbains comme les bâtiments et ainsi dégrader la qualité de leurs indicateurs. À l'inverse, une maille trop grande peut entraîner une agrégation trop importante des informations ou, dans le cas des tissus urbains, englober plusieurs structures urbaines ne permettant plus d'obtenir des tissus homogènes. Dans le cadre de l'analyse de la trame grise, les mailles peuvent être de tailles très différentes selon leur emplacement dans la ville et donc le type de tissu (§ 1.2.1.2) et sont généralement de formes plus complexes que les mailles (Barnsley et Barr, 2000) ce qui représente un facteur limitant à l'utilisation de ce type de découpage.

### *Découpages créés à partir des données géographiques*

L'îlot et la parcelle sont des unités d'analyse spécialement adaptée à l'analyse des tissus urbains, mais qui nécessite que les informations cadastrales soient disponibles. Pour pallier ce manque, plusieurs auteurs ont proposé de construire ces découpages à partir des données à leur disposition. Ainsi, Puissant *et al.* (2011) utilisent les informations sur le linéaire contenu dans la BD TOPO® pour construire les îlots urbains afin de cartographier les tissus de Strasbourg. Le réseau routier, le réseau ferré ainsi que le réseau hydrographique sont utilisés et lorsque ces différents réseaux forment un cycle fermé alors un îlot est créé. Cette construction permet d'obtenir une cartographie continue de l'espace puisque la rue y est intégrée, mais de nombreux îlots de petite dimension peuvent être créés notamment dans les zones où la densité des réseaux, quels qu'ils soient, est importante ou lorsque des petits objets sont entourés par un réseau, par exemple un rond-point. L'utilisation de cette unité d'analyse avec ce type de construction a aussi été utilisée par Hussain *et al.* (2012) qui prennent en compte uniquement le linéaire routier. Ces auteurs utilisent l'empreinte au sol des bâtiments pour cartographier les tissus urbains de Manchester en utilisant une classification supervisée. Bernabé *et al.* (2013) utilisent les mêmes données auxquelles ils ajoutent des informations sur la hauteur des bâtiments, mais s'appuient sur une classification non supervisée pour la cartographie des tissus urbains de Nantes. À partir de la classification d'une image Quickbird en dix classes d'occupation du sol, Novack et Kux (2010) utilisent ce découpage pour obtenir quatre types de tissus pour la ville de São Paulo au Brésil en construisant une base de règle.

Hamaina *et al.* (2012) vont construire des parcelles qu'ils appellent « morphologiques » pour cartographier la ville de Nantes à ce niveau d'analyse. Ces parcelles sont construites à partir de l'empreinte au sol des bâtiments et d'un diagramme de Voronoï. Cette tessellation permet de créer une partition qui représente en quelque sorte l'aire d'influence morphologique de chacun des bâtiments. Elles leur permettent de caractériser les espaces ouverts de chaque bâtiment et de réaliser une cartographie des tissus à partir d'une classification non supervisée prenant en compte la dimension spatiale des données.

L'ensemble des découpages créés à partir des données qui ont été présentés jusqu'ici utilise des bases de données vectorielles pour leur construction, mais il est toutefois possible de les construire directement à partir d'une image. Wu *et al.* (2006) proposent d'identifier plusieurs types de tissus en utilisant une image aérienne couleur infrarouge. Les bâtiments sont tout d'abord extraits avec une analyse pixel puis codés de façon binaire. Ces auteurs construisent ensuite plusieurs bandes de texture et par l'analyse d'un variogramme arrivent à obtenir des limites de types de tissus qu'ils classent après avoir testé différentes tailles de fenêtre pour calculer la texture.

Les récentes méthodes de segmentation d'images permettent de créer de grandes régions qui peuvent être utilisées comme des limites pour les tissus urbains. Stow *et al.* (2007) utilise ainsi une segmentation multi-échelles leur permettant, à partir d'une image Quickbird, d'identifier deux types d'habitats : l'habitat informel et non informel auxquels est ajoutée une classe non résidentielle. Ils testent ici deux stratégies. La première utilise uniquement les valeurs spectrales et la forme des régions pour les classer tandis que la deuxième passe par l'identification préalable des trois composantes des approches VIS présentées au début de cette section. Pour ces deux stratégies, les performances obtenues sont proches. Avec une typologie plus complexe en 10 classes, Lackner et Conway (2008) profitent de la possibilité des segmentations multi-échelles. Un premier niveau de segmentation fin est donc construit à partir d'une image Ikonos afin d'extraire 17 classes d'occupation du sol et un deuxième plus grossier est réalisé pour fixer les limites des types de tissus recherchés. Ces auteurs construisent ensuite une base de règles pour réaliser leur classification qui utilise sur les caractéristiques des objets d'occupation du sol tels que leur surface, leur présence, etc. Kurtz *et al.* (2012) proposent aussi une segmentation multi-échelles pour la classification des types urbains de Strasbourg. Toutefois à la différence des deux recherches précédente, celle-ci utilise une approche descendante et plusieurs images avec des résolutions spatiales différentes : une image Spot 5 (9,4 m), une image Quickbird (2,4 m) et une image Quickbird fusionnée (0,6 m).

Toujours à partir d'une image THRS, Comber *et al.* (2012) réalisent une classification orientée-objet pour extraire six classes d'occupation du sol. Pour construire les tissus urbains, ils n'utilisent pas un deuxième niveau de segmentation, mais créent dans un premier temps un graphe pondéré. Le graphe est construit selon l'adjacence des objets issus de la classification et la pondération se fait selon leur environnement. L'environnement est ici défini comme la part de chaque classe d'occupation du sol dans un rayon de 50 m et la dissimilarité de l'environnement entre deux objets voisins est utilisée comme pondération du graphe. Ces auteurs utilisent ensuite des méthodes de recherches de communautés dans les graphes afin de créer des sous-graphes fortement connectés et semblables. Ces communautés sont donc considérées comme des tissus homogènes. Plusieurs algorithmes de détection de communautés sont testés, mais ne permettent pas d'obtenir de bons résultats pour délimiter les tissus. Pour ces auteurs, ceci est principalement lié au fait que ces méthodes ne considèrent pas l'espace géographique

et nécessiterait donc de pouvoir intégrer cette dimension dans les méthodes de détection de communautés.

Enfin plusieurs auteurs proposent de classer les tissus urbains directement à partir du bâtiment. Toutefois le tissu urbain concerne l'organisation des éléments entre eux. Ainsi Colaninno *et al.* (2011) vont pour chaque bâtiment étudier l'organisation des éléments en réalisant des zones tampons de 200 m et l'associer à des informations sur la forme du bâtiment. Sept types sont ainsi identifiés pour la ville de Barcelone. Sahraoui *et al.* (2014) utilisent une approche similaire. Ils associent à la forme des bâtiments et à leur accessibilité à différentes aménités urbaines, des informations sur leur environnement dans un rayon visible du bâtiment (50 m) et un plus grand représentant les déplacements réguliers hors et vers le lieu de résidence (500 m). Cet environnement est caractérisé par la part de chaque classe d'occupation du sol et par un indice de diversité. Chaque bâtiment est ensuite classé grâce à une méthode non supervisée puis une généralisation cartographique par la technique des enveloppes est effectuée afin d'obtenir une typologie des espaces résidentiels de trois villes françaises.

Cette section a permis de montrer la diversité des méthodes qui peuvent être utilisées pour cartographier la trame grise urbaine. Le choix d'un découpage ou de la méthode pour le faire est une étape cruciale dans la mise en place d'une méthodologie d'identification des tissus. Ce découpage a en effet de nombreuses implications sur l'échelle de représentation qui sera obtenue ainsi que sur les types qu'il sera possible d'identifier. Quand les informations cadastrales sont disponibles, les îlots apparaissent être des unités adaptées. Ces îlots, quand ils sont construits à partir des données, permettent de prendre en compte les différentes tailles des tissus, mais quand le linéaire est dense de nombreux artefacts de petites dimensions peuvent être créés. De plus il est possible avec cette unité construite ou existante de créer des îlots de types mixtes. Le carroyage quant à lui ne permet pas de rendre compte des tailles différentes de tissus et peut lui aussi contenir des îlots mixtes. Il apparaît néanmoins intéressant notamment dans le cadre d'analyse multi-temporelles puisque les mêmes unités peuvent toujours être réutilisées et superposées ce qui ne serait pas possible avec les îlots du fait de la création régulière de nouveau. Réaliser ce découpage à partir des images apparaît intéressant, mais le choix d'une segmentation ou d'une taille de fenêtre reste une étape difficile qu'il est nécessaire de bien évaluer. Aucune étude à notre connaissance n'a tenté d'évaluer ces différentes unités d'analyses et il est donc difficile de dire laquelle est la plus adaptée à l'étude des tissus. La prochaine section s'intéresse à la place qui est accordée à la végétation pour l'identification de la trame grise.

### 2.3.3.2 La prise en compte de la végétation pour la cartographie des tissus

Traditionnellement, l'étude des tissus oppose les espaces construits et les espaces ouverts sans que la composition de ces espaces ouverts soit clairement mentionnée. La définition des tissus nous a montré que ses éléments constitutifs sont les voies, les parcelles et les bâtiments et que la végétation n'était pas directement prise en compte (§ 1.2.1.1).

Toutefois pour la cartographie des tissus, de nombreux auteurs ont utilisé la végétation pour permettre une meilleure discrimination.

Les applications bioclimatiques de l'étude des formes urbaines utilisent souvent la végétation pour caractériser les tissus du fait des effets qu'elle peut avoir sur l'îlot de chaleur urbain ou la modification des flux atmosphériques qu'elle entraîne (§ 1.3.3). Long et Kergomard (2005) et Bonhomme (2013) utilisent ainsi la couche de végétation de la BD TOPO® pour la classification des tissus. Dans ces deux recherches, la végétation arborée est considérée uniquement par la part qu'elle représente dans chaque maille. Cette même base de données et ce même indicateur est utilisé par Sahraoui *et al.* (2014) pour mesurer l'environnement de chaque bâtiment.

Lorsque les tissus urbains sont étudiés à partir d'une analyse d'image, la végétation est souvent extraite dans la première étape de classification d'occupation du sol. La végétation peut alors être considérée dans son ensemble sans distinction de classes (Herold *et al.*, 2003b ; Stow *et al.*, 2007), uniquement la végétation arborée (Hermosilla *et al.*, 2012) ou avec une distinction entre différentes strates (Banzhaf et Höfer, 2008 ; Comber *et al.*, 2012). Dans la plupart de ces recherches la description la végétation reste très souvent limitée comme précédemment à la part qu'elle occupe dans chacune des unités d'analyse. Cet indicateur est très important pour décrire les formes urbaines puisque c'est un de ceux que Herold *et al.* (2003b) ont déterminés comme le plus discriminant pour la cartographie de l'utilisation du sol, devant ceux dérivés des bâtiments. Mais peu d'auteurs ont considéré et évalué d'autres indicateurs permettant de mieux rendre compte de sa forme, de son organisation.

Hermosilla *et al.* (2012) considère en plus de sa densité, la hauteur moyenne de la végétation arborée et son écart-type pour chaque parcelle ainsi que ses caractéristiques spectrales grâce à un indice de végétation. Avec une classification en deux strates de la végétation (arbre et herbe), Bauer et Steinnocher (2001) et Walde *et al.* (2014) considèrent les relations topologiques entre tous les éléments qui compose un bloc urbain par l'intermédiaire d'un graphe d'adjacence leur permettant de calculer une série d'indicateurs tels que le nombre de voisins, le nombre d'arrête entre deux classes d'occupation du sol différentes, des mesures de centralité, etc.

Herold *et al.* (2003b) considèrent en plus de la part d'occupation du sol de la végétation, la densité en terme de nombre d'éléments par hectare ainsi que des indicateurs permettant la description de ces éléments tels que l'aire, la forme fractale, la cohésion et les distances au plus proche voisin de ces éléments. Cette recherche est à notre connaissance la seule à avoir analysé l'apport que représentent des indicateurs de formes de la végétation. Elle a montré que ces indicateurs permettaient d'obtenir une meilleure description des tissus urbains, mais ne considère toutefois pas plusieurs strates de végétation. Il apparaît donc intéressant de mieux apprécier les formes et les organisations de la végétation arborée et herbacée et d'évaluer leur contribution à l'analyse et la cartographie de la structure urbaine.

## Conclusion

Ce chapitre a tout d'abord permis de définir une donnée géographique, les modes de représentation des objets géographiques ainsi que les évolutions que ce secteur a connues. Il avait aussi pour objectifs de présenter les sites d'études qui ont été retenus ainsi que les bases de données qui peuvent être mobilisées pour l'analyse de la structure urbaine, sa trame grise et sa trame verte. L'analyse de ces différentes bases a montré qu'il existait un manque pour la cartographie de la ville à l'échelle du tissu puisque les données existantes décrivent soit les éléments gris et verts ou, à plus petite échelle, les grandes occupations ou utilisations du sol. De plus, cette analyse a montré que la cartographie de la végétation, dans les bases de données géographiques à grande échelle, n'était pas suffisante pour étudier les trames vertes en milieux urbains du fait de surfaces minimums de représentation trop grandes, mais aussi parce que la plupart ne considèrent que la végétation arborée.

Ce chapitre a permis de présenter les différentes méthodes existantes permettant d'extraire les éléments verts à partir d'images THRS puisque ces images sont particulièrement adaptées à l'identification d'objets dans les espaces urbains et, dans un deuxième temps, les méthodes dédiées à la cartographie des tissus urbains. Cet état de l'art des méthodes existantes a permis de montrer que pour l'analyse d'image THRS celles orientées-objet sont très bien adaptées, mais qu'il existe encore des défis pour rendre ces méthodes plus efficaces. Pour la cartographie des tissus, il a montré la diversité des découpages qui pouvaient être utilisés, mais aussi des données qui peuvent être mobilisées. De plus, il a montré que la végétation est prise en compte trop simplement, souvent uniquement par la part qu'elle occupe, dans la définition des tissus urbains. Ce chapitre a finalement permis de préciser les objectifs secondaires de cette thèse qui sont rappelés dans la conclusion générale de cette partie.



## Conclusion de la partie I

### *Positionnement de recherche*

Cette première partie a permis de définir le cadre conceptuel, thématique et méthodologique ainsi que de préciser les objectifs de cette thèse. L'objectif principal poursuivi dans cette thèse est de mieux comprendre l'articulation de la trame grise et la trame verte en milieu urbain et de proposer une méthodologie permettant de cartographier de manière semi-automatique la structure urbaine à partir de données surfaciques qui décrivent les éléments urbains. Le premier chapitre a permis de préciser ce que nous considérons comme la structure urbaine, la combinaison de différentes trames, la trame grise qui s'apparente à la notion de tissu urbain et la trame verte qui elle se décompose en de nombreuses sous-trames telles que celle forestière, agricole, etc. Ces deux trames sont souvent analysées en opposition l'une à l'autre, mais le fait que les limites de la ville sont souvent décrites comme floues montre bien l'interpénétration de ces deux trames.

Ce premier chapitre a aussi montré la diversité des typologies qui peuvent être utilisées pour cartographier la trame grise et la trame verte. Les typologies présentées dans ce chapitre étaient celles issues de la recherche, mais le chapitre 2 a révélé que dans les bases de données institutionnelles les nomenclatures sont aussi variées. Dans ce contexte nous proposons une typologie grise (figure A) ainsi qu'une typologie verte (figure B) adaptées pour la cartographie des espaces urbains et périurbains aux échelles du 1/5000<sup>e</sup> au 1/10 000<sup>e</sup>. Ces typologies ont bénéficié de travaux antérieurs (GeOpenSim<sup>2</sup> – Puissant *et al.*, 2011) et ont aussi été utilisées dans le cadre du projet de recherche VALI-URB dont l'objectif était de valider des méthodes de télédétection pour cartographier les espaces urbains. Ces deux typologies ont ainsi été validées par les utilisateurs finaux : CIGAL pour la trame grise et l'AUDIAR (agence d'urbanisme de Rennes) pour la trame verte.

Pour la trame grise, elle se décompose en deux niveaux. Le premier niveau est principalement morphologique et est composé de huit classes tandis que le second en 25 classes mêle morphologie, notamment pour quatre classes bâties, et utilisation du sol

---

2. Projet financé par l'Agence Nationale de la Recherche pour la création d'une plate-forme Géographique OpenSource de Simulation des évolutions des tissus urbains. Projet collaboratif entre le LIVE, l'équipe COGIT de l'IGN (Cartographie et Géomatique) et le LSIT (Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection)

pour les espaces urbains spécialisés qu'ils soient bâtis ou non. La typologie verte retenue comprend elle trois niveaux différents auxquels est associé le niveau 1A qui correspond à un niveau intermédiaire basé sur la morphologie des éléments verts. Ce niveau intermédiaire constitue le niveau recherché dans cette thèse.

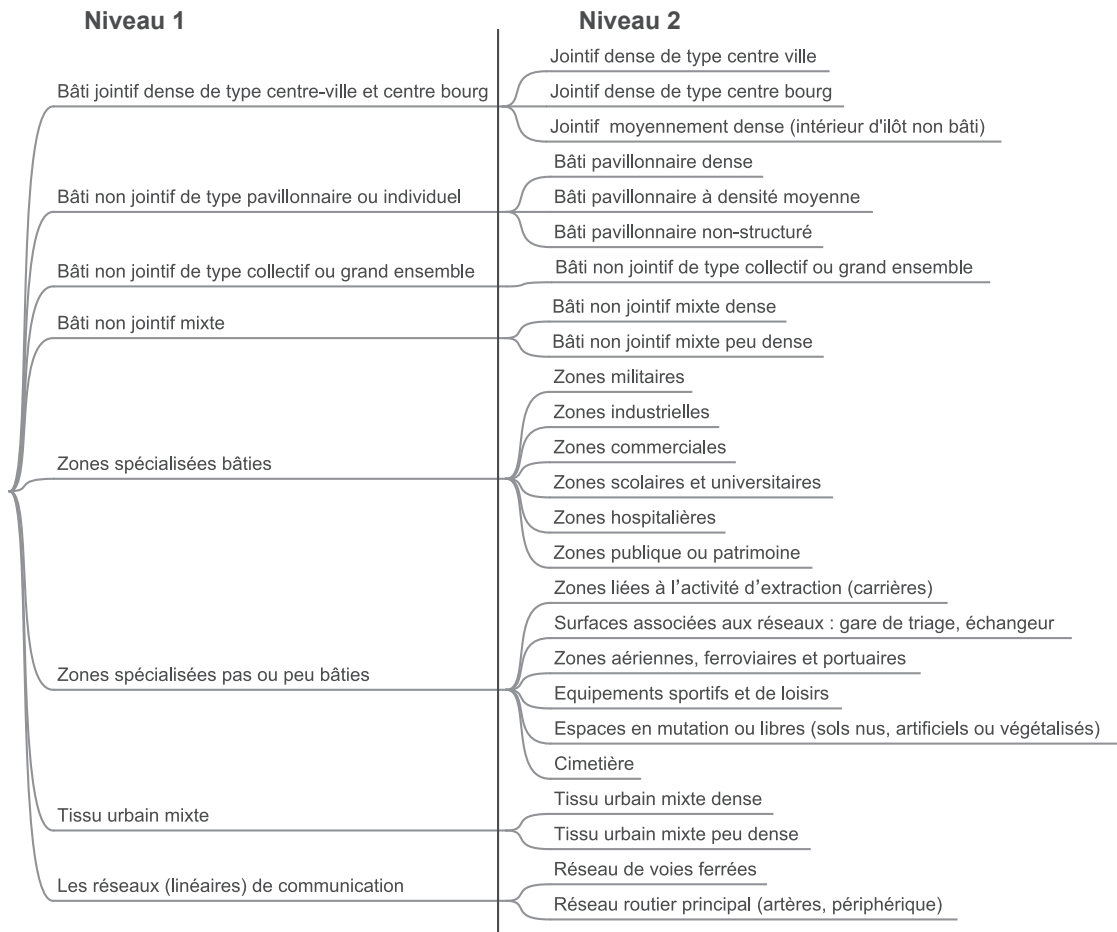


Figure A – Typologie de la trame grise à l'échelle des tissus urbains proposée

Le deuxième chapitre de cette thèse a précisé les notions liées aux données géographiques. Il a aussi présenté les deux sites d'études qui font l'objet de cette thèse, Strasbourg et Rennes. Il a montré que les bases de données existantes ne permettaient pas d'obtenir une représentation cartographique satisfaisante des tissus urbains. En effet celles continues sur l'espace urbain n'ont pas une échelle adaptée et leur mise à jour peut être difficile, notamment parce que ces bases sont réalisées par photo-interprétation d'images. Cette opération pour de grandes étendues peut être très longue et coûteuse et il est donc nécessaire de développer des méthodes plus automatiques pour cartographier la structure urbaine à l'échelle des tissus. Notre hypothèse est que l'occupation du sol au niveau des éléments permet de reconstruire la structure urbaine. Toutefois, les données

existantes qui cartographient les éléments verts ne sont pas exhaustives et n'ont pas une échelle de représentation suffisamment grande pour cette application à l'inverse des bases qui décrivent les éléments gris. Une première étape est donc de cartographier ces éléments verts.



Figure B – Typologie de la trame verte proposée

Les images THRS sont des sources de données qui permettent cette cartographie exhaustive. Ce type d'image nécessite des méthodes qui soient adaptées. Les méthodes orientées-objet sont apparues comme des solutions permettant d'extraire l'occupation du sol à partir de ces images. Elles ont souvent été associées à des classifications à base

de règles, mais qui peuvent aussi être longues à mettre en place et nécessitent d'être automatisées d'autant plus que le nombre d'images provenant de satellites d'Observation de la Terre ne cesse d'augmenter. Les méthodes de classification sont apparues efficaces et plus spécifiquement les supervisées qui permettent d'obtenir de bonnes performances. Mais si ces méthodes apparaissent adaptées, plusieurs défis font qu'elles ne sont pas pleinement opérationnelles. Tout d'abord la segmentation est une étape importante souvent réalisée par essais et erreurs et qui nécessite donc d'être automatisée. De plus les classifications supervisées nécessitent des données d'apprentissage, souvent nombreuses, et il est donc important de développer des méthodes qui permettent à l'utilisateur d'être guidé pour leur choix afin de minimiser l'effort pour les créer. Enfin avec l'amélioration de la résolution spatiale, les images sont de plus en plus lourdes à traiter et il est donc nécessaire de proposer des stratégies permettant d'en extraire des informations sur de grandes étendues. L'ensemble de ces défis correspond ainsi à des objectifs secondaires de cette thèse et font l'objet de sa deuxième partie. La méthodologie qui y est construite a pour objectif d'extraire la végétation urbaine à partir d'images THRS.

Ces éléments végétaux, associés à d'autres issus de bases de données existantes, nous permettent de cartographier les trames grise et verte (partie III). Le chapitre 2 a permis de montrer la diversité des méthodes qui peuvent être mobilisées pour cette application et que le choix d'un découpage est important notamment quand aucun n'est disponible. Il est donc nécessaire de définir une unité d'analyse adaptée pour cartographier le milieu urbain, mais aucune recherche à notre connaissance n'a tenté d'évaluer plusieurs découpages. La partie III permet donc de répondre à ce manque à partir de la base de données multi-sources qui y est constituée. À partir de cette base, des indicateurs sont dérivés pour décrire la morphologie et l'organisation spatiale des tissus. Ces indicateurs ont, comme nous l'avons montré, une grande importance pour répondre aux objectifs de la typologie recherchée. Nous faisons aussi l'hypothèse que la cartographie des tissus peut être améliorée grâce à d'autres indicateurs que ceux dérivés des surfaces des classes d'occupation du sol et nous en proposons donc de nouveaux afin de mieux décrire la structure urbaine. Les méthodes qui sont proposées sont tout d'abord développées et testées sur l'agglomération de Strasbourg puis validées grâce au site d'étude de Rennes.

Les encadrés ci-après résument les objectifs généraux et spécifiques identifiés et la figure C les différentes étapes de la méthodologie mise en place.

**Objectifs généraux :**

- Comprendre l'articulation de la trame grise et la trame verte pour mieux quantifier, analyser et représenter la structure urbaine.
- Proposer une méthodologie adaptée à la cartographie à une échelle du 1/5000<sup>e</sup> au 1/10 000<sup>e</sup> de la structure urbaine et de ses éléments à partir de données multi-sources.

**Objectifs spécifiques :**

- Sélection automatique d'une méthode de segmentation et de ses paramètres pour l'extraction de la végétation urbaine à partir d'images satellites THRS Pléiades.
- Sélection d'échantillons d'entraînement informatifs et divers par apprentissage actif pour réduire les coûts de labélisation.
- Adaptation des méthodes pour la classification d'image sur de grandes étendues.
- Création d'une base de données multi-sources pour décrire l'occupation du sol et les éléments urbains et périurbains à une échelle du 1/5000<sup>e</sup>.
- Choix d'une unité d'analyse adaptée pour la cartographie de la structure urbaine, de sa trame grise et de sa trame verte.
- Sélection d'indicateurs permettant de rendre compte de l'organisation des éléments gris et verts afin de cartographier la structure urbaine complétés par d'autres indicateurs pour élargir l'analyse quantitative de la structure urbaine.

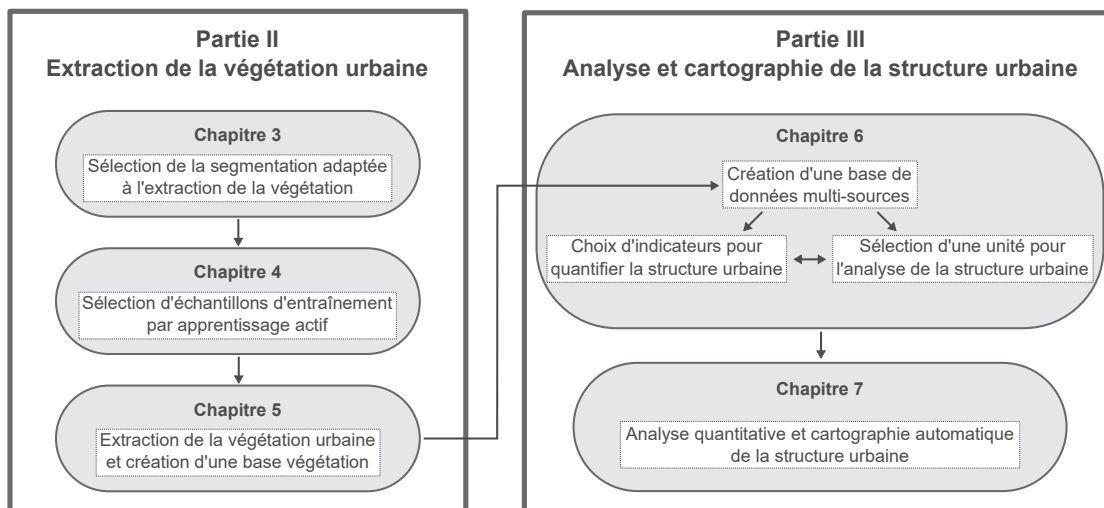


Figure C – Structuration de la démarche méthodologique de la thèse



## Partie II

### Extraction de la végétation urbaine à partir d'images satellites à Très Haute Résolution Spatiale

Introduction de la partie II.....	79
<b>CHAPITRE 3</b> : Sélection de la segmentation optimale pour l'extraction des éléments verts urbains.....	83
<b>CHAPITRE 4</b> : Comparaison de stratégies d'échantillonnage pour la classification d'images THRS.....	111
<b>CHAPITRE 5</b> : Généralisation de l'approche orientée-objet par apprentissage actif pour la classification de la végétation.....	137
Conclusion de la partie II.....	163





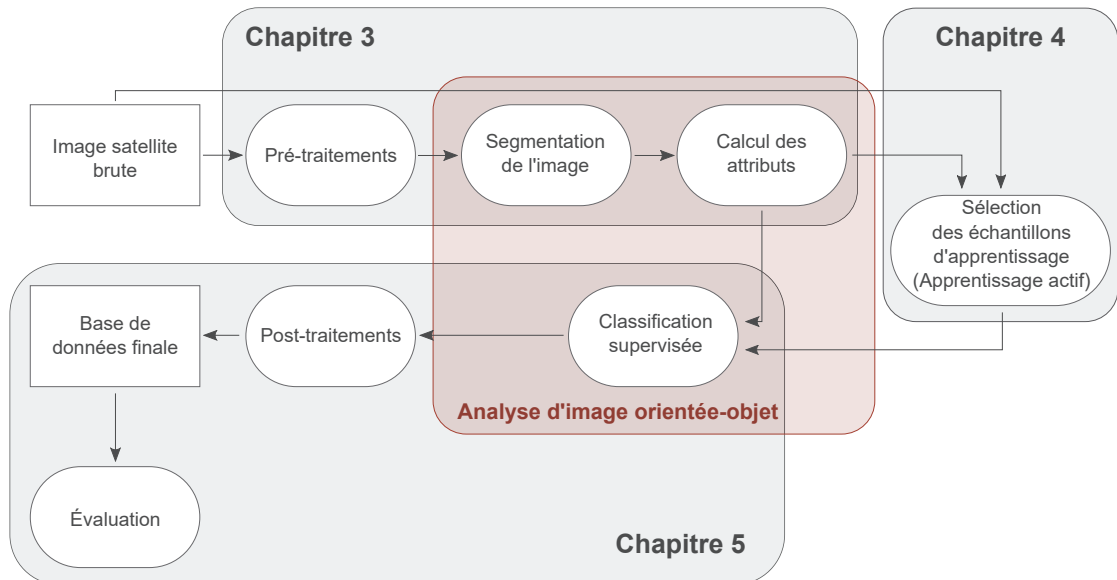
## Introduction de la partie II

Les images, tant satellitaires qu'aériennes, sont des sources de données privilégiées pour la cartographie des espaces urbains et périurbains notamment pour l'identification de la végétation qui constitue l'objectif général de cette partie. L'apparition des images satellites THRS a permis à ce type d'image d'être utilisé pour l'extraction des objets urbains en offrant une taille de pixel équivalente à celle des prises de vue aériennes tout en fournissant une plus grande résolution spectrale et une plus grande répétitivité. Avec la mise à disposition de ces images, de nouveaux défis méthodologiques sont toutefois apparus.

Une des approches proposées, devenue très populaire pour l'analyse d'images THRS, est l'approche orientée-objet (Blaschke, 2010). L'approche méthodologique retenue pour extraire la végétation arborée et herbacée urbaine en milieu urbain et périurbain se place dans ce paradigme (Blaschke *et al.*, 2014). Pour Puissant (2003), une démarche orientée-objet peut être décomposée en trois étapes. La première étape consiste à segmenter l'image en régions (ou segments) qui sont le résultat de l'agrégation de pixels homogènes grâce à un algorithme de segmentation. La segmentation est importante, car elle peut influencer fortement la qualité des attributs et donc les résultats de la classification. L'objectif de la deuxième étape d'une démarche orientée-objet est de caractériser ces régions à partir des valeurs spectrales de l'image, de leur texture, de la géométrie des objets ou de leur contexte. L'étape suivante est la classification de ces segments. L'objectif est ici de définir les objets selon leurs caractéristiques afin de leur attribuer une classe. Dans cette thèse, notre choix s'est porté sur l'utilisation de méthodes supervisées de classification qui permettent d'obtenir de bonnes performances lorsque les classes sont connues *a priori*. Le principe de ce type de classification est d'utiliser des exemples dont la classe est connue pour que la méthode apprenne puis prédise ces classes à l'ensemble des régions qui ne sont pas labélisées. Le choix des échantillons d'apprentissage a une influence très importante sur les performances puisqu'il doit permettre de saisir toute la diversité de chacune des classes. La labélisation de ces données d'entraînement représente ainsi un coût qui peut être important à cause du temps nécessaire à cette étape et la mise en place de méthodes d'échantillonnage adaptées à l'image reste aujourd'hui un défi.

La figure D reprend les différentes étapes de la démarche orientée-objet en y intégrant les autres étapes nécessaires de l'analyse d'image telles que les pré et post-traitements,

ainsi que l'évaluation des résultats, mais aussi la sélection d'échantillons adaptés pour la classification supervisée des segments. La structuration de notre démarche méthodologique est aussi représentée sur cette figure et correspond à la structure de cette partie.



**Figure D** – Processus d'analyse d'image orientée-objet par classification supervisée et structuration de la démarche méthodologique

Ainsi le chapitre 3 consiste à définir la segmentation optimale pour deux zones tests de nos sites d'étude de Strasbourg et de Rennes. Optimale dans le sens où cette segmentation permette de représenter cartographiquement au mieux la végétation urbaine. Différentes méthodes couramment utilisées pour évaluer les segmentations sont testées et comparées afin de vérifier si ces méthodes sont appropriées pour définir la segmentation optimale. Ce chapitre permet aussi de présenter les images satellites Pléiades utilisées, les prétraitements réalisés, ainsi que l'ensemble des attributs permettant de caractériser les régions issues du processus de segmentation.

Le second chapitre de cette partie (chapitre 4) propose différentes méthodes permettant de sélectionner des exemples pour l'entraînement des méthodes de classification. En effet la multiplication des images satellites nécessite de revoir les méthodes utilisées pour les classifications telles que celles à base de règles qui nécessitent de nombreux paramètres définis manuellement ou des méthodes supervisées qui ont besoin de nombreux échantillons d'apprentissage pour être performantes. L'avenir est donc dans des méthodes permettant de guider l'utilisateur dans le choix des échantillons. L'objectif est ici de réduire le nombre d'échantillons nécessaire à l'obtention de classifications performantes en ne sélectionnant que les plus divers et les plus informatifs afin de réduire l'effort de labélisation par le développement de méthodes d'apprentissage actif

multi-classes. L'ensemble de ces méthodes est appliqué sur les zones tests de Strasbourg et de Rennes afin de déterminer la méthode la plus efficiente.

Enfin, le chapitre 5 généralise à l'ensemble de nos deux zones d'études les méthodes proposées précédemment, la segmentation et la sélection d'exemples d'apprentissage, puis compare différents algorithmes de classification pour définir le plus performant pour l'extraction de la végétation de chacun des sites d'études. Le protocole de post-traitements ainsi que l'évaluation des cartographies obtenues font aussi l'objet de ce chapitre.



# Chapitre 3

## Sélection de la segmentation optimale pour l'extraction des éléments verts urbains

### Sommaire

---

Introduction . . . . .	84
3.1 Segmentation de l'image et évaluation : choix des méthodes . . . . .	84
3.1.1 Choix des méthodes de segmentation . . . . .	85
3.1.2 Évaluation de la qualité des segmentations . . . . .	87
3.2 Sélection des zones tests et données . . . . .	92
3.2.1 Les images Pléiades disponibles . . . . .	92
3.2.2 Choix des sites tests . . . . .	93
3.2.3 Prétraitements des images Pléiades et présentation des attributs calculés . . . . .	94
3.2.4 Créations des données d'entraînement et de validation . . . . .	95
3.3 Sélection des segmentations adaptées à l'extraction de la végétation urbaine . . . . .	97
3.3.1 Présentation des segmentations réalisées . . . . .	97
3.3.2 Évaluation des segmentations pour la cartographie des éléments verts . . . . .	99
3.3.3 Comparaison des méthodes d'évaluation de segmentation . . . . .	105
Conclusion . . . . .	108

---

## Introduction

Comme nous l'avons présenté dans la première partie de cette thèse (§ 2.3.2), les images, aériennes ou satellites, sont des sources de données mobilisables afin d'extraire la végétation en milieux urbains. Mais le manque de méthodes opérationnelles limite leur usage. Les méthodes orientées-objet, si elles paraissent assez bien adaptées à ce type d'images, ne sont pas sans limites. En effet, elles peuvent être assez longues à mettre en œuvre et la segmentation reste une étape critique qui peut avoir une influence importante sur les performances (Lefebvre *et al.*, 2011 ; Martha *et al.*, 2011). Choisir une segmentation adaptée est, comme pour une base de règles, un processus souvent réalisé par essais et erreurs afin d'obtenir une adéquation parfaite entre les bords des segments et ceux des objets recherchés. C'est en évaluant qualitativement, visuellement, les segmentations obtenues que le choix est fait ce qui remet donc en cause la reproductibilité de ce type d'approche ainsi que son objectivité. Néanmoins, de nombreuses méthodes existent afin de déterminer une segmentation adaptée, mais celles-ci ont rarement été comparées et évaluées.

Les objectifs de ce chapitre sont dès lors de vérifier les performances de plusieurs algorithmes de segmentation et de quelques méthodes permettant de les évaluer afin de sélectionner celle qui est la plus adaptée à l'extraction de la végétation en milieu urbain. Cette méthode vise aussi à être reproductible et est construite pour le site d'étude de Strasbourg puis validée sur celui de Rennes. La prochaine section présente plus en détail le principe de la segmentation d'image et les méthodes retenues, aussi bien celles de segmentation que celles permettant de les évaluer. Nous présentons ensuite dans la section 3.2 les images qui sont utilisées ainsi que les sites tests utilisés afin de tester et d'évaluer notre démarche. Cette section présente aussi les prétraitements qui ont été réalisés sur ces images et les données auxiliaires qui ont été créées. Enfin, les résultats sont présentés et analysés afin de déterminer la segmentation optimale pour extraire la végétation en milieu urbain (§ 3.3).

### 3.1 Segmentation de l'image et évaluation : choix des méthodes

Dans le cadre d'une démarche d'analyse d'image orientée-objet, la première étape de segmentation est critique puisqu'elle va avoir une influence importante sur la qualité des descripteurs et donc sur la performance de la classification (Zhang, 1995 ; Clinton *et al.*, 2010 ; Gao *et al.*, 2011). De nombreuses méthodes de segmentation ont été développées et appliquées pour l'analyse d'image de télédétection (Dey *et al.*, 2010) ainsi que de nombreux indicateurs pour les évaluer et déterminer celle la plus en adéquation avec les objets recherchés.

### 3.1.1 Choix des méthodes de segmentation

Les méthodes de segmentation d'images sont généralement classées en trois catégories : les approches basées sur l'analyse des histogrammes, sur la croissance des régions et enfin les méthodes basées frontières (Pal et Pal, 1993 ; Lefebvre, 2011 ; Kurtz, 2012). Comme nous l'avons mentionné, le processus de segmentation va regrouper des pixels voisins au sein d'une même région selon des critères dépendant de la méthode utilisée. L'objectif de la segmentation, pour que celle-ci soit optimale, est donc de créer un segment pour chaque objet géographique présent sur l'image, que tous les pixels décrivant un objet soient dans un seul et unique segment. De plus, les bords des segments et ceux des objets géographiques doivent être en parfaite adéquation. Toutefois, les objets recherchés peuvent être de tailles très différentes sur une image, et ce, au sein d'une même classe thématique, ce qui rend complexe le processus de segmentation. Deux phénomènes apparaissent ainsi conjointement : la sur et la sous-segmentation.

La sur-segmentation se traduit par un nombre trop important de segments. Visuellement, chaque objet d'intérêt devient donc divisé en plusieurs sous-objets qui ne correspondent donc plus réellement à celui recherché (figure 3.1b). À l'inverse, la sous-segmentation correspond à un nombre de segments trop faibles et plusieurs objets d'intérêt vont se trouver fusionnés au sein d'un même segment (figure 3.1c). Généralement, une sur-segmentation est préférée, car elle permet de fusionner les objets de même classe thématique si elle est suivie d'une classification tandis que la sous-segmentation rend plus difficile la séparation des objets mixtes. Dans le cadre d'une classification supervisée, les résultats obtenus apparaissent donc généralement plus performants lorsque l'image est sur-segmentée plutôt que sous-segmentée (Stumpf et Kerle, 2011 ; Smith, 2010).

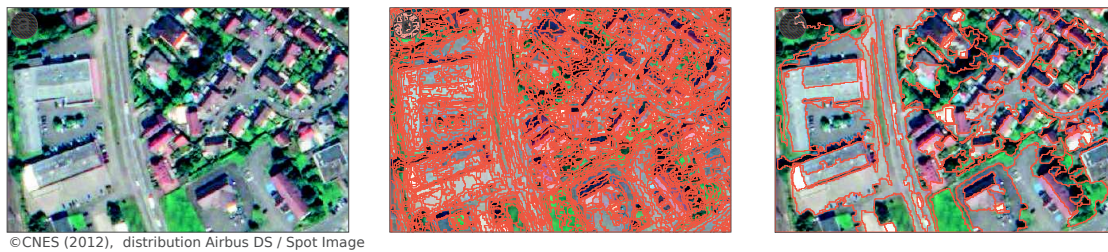


Figure 3.1 – Exemple d'une image (a) sur-segmentée (b) et sous-segmentée (c)

Deux algorithmes de segmentation sont testés et comparés dans cette recherche : Multi-Resolution et Mean Shift. Notre choix s'est porté sur ces deux méthodes pour plusieurs raisons. La première est que ces deux algorithmes représentent chacun une des catégories de méthodes de segmentation d'images. Multi-Resolution est une méthode basée sur la croissance des régions, Mean Shift sur l'analyse d'histogramme. De plus, Multi-Resolution est certainement la méthode de segmentation la plus utilisée actuellement dans les approches orientées-objet pour l'observation de la Terre (Blaschke, 2010). Elle est souvent présentée comme ayant de très bons résultats comparés à d'autres méthodes (Neubert *et al.*, 2008) et est déjà implémentée dans plusieurs logiciels, commerciaux ou

libres. Nous utilisons l'implémentation de cette méthode dans le logiciel commercial eCognition, spécialisé dans ces approches orientées-objet. Mean Shift quant à elle a déjà été utilisée pour identifier la végétation avec de bons résultats autant en milieu agricole (Ponti, 2013) que dans des espaces urbains (Tong *et al.*, 2014). De plus, elle a l'avantage d'être intégrée à la librairie libre Orfeo Toolbox développée par le CNES. Cette implémentation présente un autre avantage non négligeable, elle peut être utilisée sur de très grandes scènes (Michel *et al.*, 2015) ce qui n'est pas possible avec Multi-Resolution et eCognition sans procéder à un tuilage préalable.

### 3.1.1.1 La segmentation Multi-Resolution

Multi-Resolution est un algorithme qui utilise une approche ascendante. Il fusionne tout d'abord les pixels en régions puis agrège ensuite ces régions en fonction de leur homogénéité (Benz *et al.*, 2004). Ce critère d'homogénéité est défini par plusieurs paramètres.

Tout d'abord, le paramètre d'échelle est celui qui a une influence directe sur le nombre de segments et donc sur le niveau d'agrégation des pixels. Ainsi plus ce paramètre est petit, plus la taille des segments se rapprochera de celle du pixel. L'échelle est ensuite décomposée en deux paramètres, la réflectance et la forme. L'utilisateur peut ainsi donner un poids différent aux valeurs spectrales de l'image ou la forme des objets qui résulteront de la segmentation dans le calcul de l'hétérogénéité. Ce critère de forme est lui aussi décomposé en deux paramètres, un de compacité et un de lissage. Un lissage plus important favorisera une forme plus lisse et arrondie tandis que la compacité produira des formes plus compactes. Il est de plus possible pour cette méthode de donner un poids différent aux bandes spectrales de l'image.

### 3.1.1.2 La segmentation Mean Shift

Le deuxième algorithme de segmentation que nous utilisons est Mean Shift. Cette méthode, associée à celles basées sur l'analyse des histogrammes, est au départ une méthode de clustering de données proposée par Fukunaga et Hostetler (1975) qui a ensuite été adaptée à la segmentation d'image par Comaniciu et Meer (2002).

Mean Shift est un algorithme de segmentation non paramétrique et itératif de recherche de modes qui se décompose en deux étapes. La première étape de cette méthode est le filtrage de l'image. Cette étape consiste à trouver les modes locaux en considérant le voisinage de chaque pixel, le rayon spatial, et en considérant uniquement les pixels proches spectralement, ceux ne dépassant pas le rayon spectral. Ces deux paramètres sont fixés par l'utilisateur. Ce processus de filtrage est itéré jusqu'à ce que soit atteint un seuil de convergence ou un nombre maximal d'itérations, tous deux fixés *a priori*. La deuxième étape de Mean Shift regroupe les pixels proches, d'une distance inférieure au rayon spatial et spectral, dans un même cluster pour former les segments. Les segments



qui sont de petites tailles sont ensuite associés au segment connexe le plus proche si cette taille est inférieure à un seuil défini par l'utilisateur.

### 3.1.2 Évaluation de la qualité des segmentations

L'étape de segmentation est très souvent réalisée par un processus d'essais et erreurs en inspectant visuellement les résultats et leur adéquation avec les objets recherchés par l'utilisateur afin de déterminer les paramètres optimaux. De nombreuses méthodes ont été développées afin de déterminer une segmentation optimale de façon plus ou moins automatique et reproductible. Ces différentes méthodes peuvent être classées en trois grandes approches : non supervisées, supervisées et par classification. Les approches non supervisées sélectionnent la segmentation optimale à partir d'indicateurs statistiques calculés pour chaque segment sans qu'aucun objet de référence ne soit utilisé à la différence de celles supervisées qui vérifient l'adéquation entre les segments et une vérité-terrain. Évaluer les segmentations par classification consiste à réaliser plusieurs segmentations et à sélectionner celle qui permet d'obtenir les meilleures précisions soit pour un objet en particulier soit de manière plus globale. La définition d'une segmentation optimale diffère donc selon la méthode utilisée pour son évaluation. En effet pour une évaluation non supervisée c'est l'analyse des descripteurs qui permet de la déterminer. Dans le cadre d'une évaluation supervisée des segmentations, c'est l'adéquation entre les objets réels et les segments qui vont permettre de la définir tandis que, lorsqu'elles sont évaluées par classification, ce sont les performances pour la cartographie des objets d'intérêt qui sont évaluées. Nous considérons ici cette dernière définition qui correspond à notre objectif final d'extraction des éléments verts urbains. Néanmoins, cette méthode est la plus longue à réaliser et l'utilisation de méthodes supervisées et non supervisées pour l'évaluation des segmentations peut permettre de réduire le temps nécessaire à cette étape si elles s'avèrent performantes.

Peu de publications ont eu pour objectif de comparer ces différentes approches. Belgiu et Drăguț (2014) ont comparé méthodes supervisées et non supervisées pour la segmentation des bâtiments par l'algorithme Multi-Resolution et vérifié par classification, la performance de celles sélectionnées par chacune des méthodes. Quelle que soit l'approche d'évaluation utilisée, leurs résultats ont montré que les paramètres d'échelle sélectionnés sont assez proches tout comme les performances des classifications réalisées à partir de ces deux segmentations. Dans ce travail toutefois, la sélection de la segmentation en comparant les résultats des classifications réalisées sur l'ensemble des séries n'a pas été testée. Une bonne segmentation, d'un point de vue statistique, n'est pas nécessairement celle qui produira les meilleurs résultats d'un point de vue thématique. Räsänen *et al.* (2013) ont en effet montré que pour la cartographie de la forêt boréale en Finlande, les segmentations évaluées visuellement permettaient d'obtenir de meilleures classifications que celles sélectionnées par une évaluation supervisée.

L'objectif de ce chapitre est de trouver la segmentation la plus adaptée pour la cartographie de la végétation urbaine à partir de méthodes d'évaluation existantes et de vérifier si

les approches supervisées ou non peuvent être employée afin de la déterminer. L'objectif étant de cartographier précisément la végétation, l'évaluation des segmentations par classification est considérée comme la référence. Sa mise en œuvre est néanmoins la plus coûteuse en temps comme nous le précisons dans la suite de ce travail. L'évaluation de la performance des deux autres méthodes d'évaluation apparaît nécessaire puisque si elles s'avèrent performantes, elles permettraient de sélectionner de manière objective une segmentation adaptée sans que le coût ne soit trop important. Les méthodes utilisées dans cette recherche sont présentées dans les trois sections suivantes.

### 3.1.2.1 Méthodes d'évaluation non supervisées des segmentations

Les approches non supervisées pour définir la segmentation optimale présentent l'avantage de ne pas demander de vérité-terrain qui peut parfois être très variable selon l'interprète, la complexité des objets recherchés ou des facteurs extérieurs ou techniques (Van Coillie *et al.*, 2014). De nombreuses méthodes ont été développées (Zhang *et al.*, 2008), et afin de déterminer la segmentation la plus adaptée à l'extraction de la végétation urbaine, la méthode non supervisée sélectionnée est celle proposée par Drăguț *et al.* (2010), l'ESP (*Estimation of Scale Parameter*).

L'ESP se base sur la méthode de variance locale de Woodcock et Strahler (1987) dont l'objectif est de comprendre comment la structure spatiale de l'image change avec la taille des pixels. Ils ont déterminé que la variance des valeurs spectrales des images, calculée dans une fenêtre glissante 3×3, est faible lorsque la taille des pixels est plus petite que la taille des objets, qu'elle atteint son maximum quand la taille du pixel est égale à environ la moitié ou les trois quarts de la taille des objets d'une image puis diminue quand les pixels sont plus grands que les objets. Calculée sur toute l'image, cette méthode a ensuite été adaptée pour déterminer la résolution spatiale optimale d'une image pour des objets géographiques spécifiques, que ce soit en milieu forestier (Marceau *et al.*, 1994) ou en milieu urbain (Puissant, 2003).

La méthode de variance locale a été utilisée que récemment pour déterminer la segmentation optimale d'une image par Kim *et al.* (2008). Ces auteurs ont considéré la variance locale comme la moyenne de la variance des valeurs spectrales des segments et l'ont combinée à une mesure d'autocorrélation spatiale. De nombreuses méthodes ont utilisé des mesures d'hétérogénéité intra-segments, comme la variance, qu'il faut minimiser, et une mesure d'hétérogénéité inter-segments, comme l'autocorrélation spatiale, qu'il s'agit de maximiser, afin de déterminer la segmentation optimale (Espindola *et al.*, 2006 ; Zhang *et al.*, 2012 ; Yang *et al.*, 2015 ; Troya-Galvis *et al.*, 2015). La méthode ESP ne se base que sur la variance locale et a été intégrée dans le logiciel eCognition.

La variance locale (*LV*) dans cette méthode est calculée à partir de la moyenne de l'écart-type de la valeur spectrale de l'ensemble des objets de l'image. Une série de segmentations est réalisée et le taux de variation de la variance locale ( $ROC_{LV} - Rate$

*Of Change of Local Variance*) entre deux segmentations voisines en terme d'échelle est calculée grâce à l'équation :

$$ROC_{LV} = \frac{LV_n - LV_{n-1}}{LV_{n-1}} \times 100 \quad (3.1)$$

L'analyse de la courbe obtenue à partir du  $ROC_{LV}$  permet de déterminer la segmentation optimale à l'apparition du premier pic. Ce pic représente le moment où les segments correspondent aux types d'objets caractérisés par des degrés égaux d'homogénéité. Cette approche, initialement développée pour des images panchromatiques, a été récemment étendue à celles multispectrales en moyennant les variances locales obtenues pour chaque bande avant de calculer le taux de variation (ESP2 ; Drăguț *et al.*, 2014).

### 3.1.2.2 Méthodes d'évaluation supervisées des segmentations

Les méthodes d'évaluation supervisée des segmentations vérifient l'adéquation entre les segments et une vérité-terrain. Les mesures de Hoover *et al.* (1996) sont certainement les plus réputées et sont intégrées dans l'Orfeo Toolbox, mais de nombreuses autres ont été développées et utilisées (Zhang, 1996 ; Clinton *et al.*, 2010). Ces métriques peuvent être divisées en trois groupes : celles basées sur l'aire, celles basées sur la position et les mesures arithmétiques.

Les métriques basées sur l'aire vont comparer les objets de la vérité-terrain et ceux issus de la segmentation et mesurer leur différence en terme de taille par des jeux d'intersection et d'union entre les polygones. De nombreuses métriques ont été proposées dans la littérature, le principe étant de mesurer pour chacun des objets de la vérité-terrain la sur et la sous-segmentation (Clinton *et al.*, 2010). La seconde catégorie de mesures utilise la position relative des objets par rapport à leurs segments correspondants. Par exemple Zhan *et al.* (2005) calculent la distance entre les centroïdes des segments et ceux de la vérité-terrain correspondant tandis que Lucieer et Stein (2002) mesurent la distance entre les bordures des objets de la vérité-terrain et ceux de la segmentation. Les mesures arithmétiques consistent à comparer le nombre de segments issus d'une segmentation au nombre d'objets identifiés sur une image. Elles mettent ainsi en évidence des relations entre objets et segments du type un-à-plusieurs, un-à-un, plusieurs-à-un. Ces différentes mesures qui permettent de décrire l'adéquation et les relations géographiques et topologiques des objets et des segments sont très souvent combinées afin d'obtenir un indice synthétique permettant de sélectionner de manière objective la segmentation respectant au mieux ce critère.

Ces approches ont été critiquées par le passé du fait que les vérités-terrain utilisées pour évaluer les segmentations ne sont souvent pas assez précises, notamment dans les milieux urbains (Corcoran *et al.*, 2010). Toutefois, elles ont été largement utilisées pour trouver l'échelle de segmentation la plus adaptée (Belgiu et Drăguț, 2014 ; Witharana et Civco, 2014) ou pour comparer différentes méthodes (Neubert *et al.*, 2008 ; Carleer

et al., 2005) nous poussant ainsi à vérifier les performances de ce type d'approche pour déterminer la segmentation la plus adaptée à l'extraction des éléments verts urbains. Dans cette recherche, notre choix s'est porté vers les mesures proposées par Liu et al. (2012) qui permettent de prendre en compte les relations géométriques et arithmétiques entre les segments et la vérité terrain.

Deux mesures sont ici calculées puis combinées afin de fournir un indice synthétique. La première est la *Potential Segmentation Error* (PSE) qui est le ratio entre l'aire totale des segments et l'aire totale des objets vrais. Cette métrique mesure la sous-segmentation et est calculée par l'équation :

$$PSE = \frac{\sum |s_i - r_k|}{\sum r_k} \quad (3.2)$$

où  $r_k$  correspond à l'aire du polygone de référence et  $s_i$  à celles des segments associés. Plus la valeur de cette mesure est grande, plus l'effet de sous-segmentation est important.

La seconde mesure proposée par ces auteurs est arithmétique, c'est le *Number of Segment Ratio* (NSR) qui est un indicateur de sur et de sous-segmentation déterminée par :

$$NSR = \frac{|m - v|}{m} \quad (3.3)$$

avec  $m$  qui représente le nombre de polygones de référence et  $v$  le nombre de segments correspondant. Tout comme pour la PSE, il s'agit de minimiser cette mesure.

Enfin, ces deux mesures sont combinées afin d'obtenir un indice synthétique permettant de comparer les segmentations. Grâce à la PSE et au NSR, une distance bidimensionnelle est calculée par rapport au point d'origine (0,0) qui représente la segmentation parfaite, celle qui n'est aucunement sur-segmentée ni sous-segmentée. Ainsi plus cette distance est faible, meilleure est l'adéquation entre les objets de la vérité-terrain et ceux de la segmentation. Cette distance, l'*Euclidean Distance 2* (ED2), peut être formulé par :

$$ED2 = \sqrt{PSE^2 + NSR^2} \quad (3.4)$$

### 3.1.2.3 Méthodes d'évaluation des segmentations par classification

Le principe de la sélection de la segmentation optimale par classification supervisée est de classer une série de segmentations à partir de la même méthode, avec les mêmes attributs, et d'évaluer les résultats afin de sélectionner celle où les performances sont les meilleures. Tout comme les approches supervisées, celles-ci sont soumises à la subjectivité de l'utilisateur qui doit produire une vérité-terrain afin d'étiqueter les segments utilisés pour l'entraînement et l'évaluation des classifications. C'est l'approche la plus coûteuse en temps, mais qui présente l'avantage d'être directement en correspondance avec la qualité de la cartographie finale. Ce type d'approche a été appliqué avec succès pour la détection de glissement de terrain à partir d'un unique niveau de segmentation

(Stumpf et Kerle, 2011) ainsi que pour la cartographie de l'occupation du sol à partir d'image à haute résolution et de plusieurs niveaux de segmentation (Smith, 2010).

Dans ce chapitre, l'ensemble des classifications est réalisé grâce à une méthode d'apprentissage automatique *Random Forest* (RF; Breiman, 2001) implémenté par Liaw et Wiener (2002) dans le logiciel de statistique libre R (R Core Team, 2014). RF est un algorithme qui crée une multitude d'arbres de décision basés sur la méthode des CART (*Classification and Regression Tree*, Breiman *et al.*, 1984). Lors de la construction d'un CART, à chaque nœud de l'arbre, une partition binaire est créée afin de séparer au mieux les exemples d'apprentissage selon leurs classes. Toutes les variables sont considérées et un seuillage est réalisé. La variable et le seuil sont sélectionnés grâce à l'indice de Gini qui sera minimal lorsque les deux sous-ensembles résultants seront les plus homogènes. Cette opération est ensuite répétée jusqu'au moment où les feuilles de l'arbre sont pures. À la différence des CART, les arbres construits par RF sont considérés dans leur ensemble, sans qu'il soit nécessaire de les élaguer. Pour chacun de ces arbres, l'ensemble du jeu de donnée d'entraînement n'est pas utilisé et seulement une partie des exemples est sélectionnée aléatoirement sans remplacement. Les exemples qui ne sont pas sélectionnés pour la création de chaque arbre sont appelés l'*Out-Of-Bag* (OOB).

Une seconde randomisation est réalisée à chaque nœud. Toutes les variables ne sont pas considérées pour réaliser la meilleure division, mais seulement une partie choisie aléatoirement. Le nombre de variables par nœud est par défaut fixé à la racine carrée du nombre total de variables, mais peut être ajusté par l'utilisateur. Chaque arbre ainsi créé va être titulaire d'un vote, la classe attribuée à un exemple est donc celle qui recueille la majorité des suffrages des arbres.

Cet algorithme devient de plus en plus populaire dans des domaines très variés et connaît, depuis la fin des années 2000, un intérêt très fort pour des applications de télédétection, quel que soit le capteur ou l'approche utilisée. Cette méthode a par exemple été utilisée pour cartographier les espaces urbains à partir de données Lidar (Chehata *et al.*, 2009), d'espaces agricoles avec des images Spot (Duro *et al.*, 2012b) ou pour l'extraction de végétation à partir de données multi-sources comme la combinaison d'images multispectrales et radar (Peters *et al.*, 2011) ou multispectrales et modèle numérique de terrain (Juel *et al.*, 2015). Plusieurs avantages permettent à cet algorithme de rencontrer un tel succès.

Tout d'abord, il permet d'obtenir de très bons résultats de classification qui sont comparables à ceux obtenus avec une autre méthode très utilisée, les SVM (Boser *et al.*, 1992). Comparée à SVM, cette méthode ne nécessite que peu de paramètres à ajuster avec une influence assez limitée sur les performances tant qu'un nombre suffisamment important d'arbres est pris en compte (Rodriguez-Galiano *et al.*, 2012). C'est aussi une méthode non paramétrique qui est assez peu sensible au bruit ainsi qu'au sur-apprentissage et dont le temps d'exécution pour l'entraînement est relativement court, mais qui nécessite néanmoins une quantité de mémoire assez importante (Gislason *et al.*, 2006).

Les exemples non considérés pour la construction des arbres permettent d'obtenir une mesure intrinsèque des performances de la classification, l'erreur OOB. Cette mesure est néanmoins plus optimiste que la réalité et ne permet donc pas de s'affranchir d'une validation plus traditionnelle (Genuer, 2010). L'OOB permet aussi d'obtenir une évaluation de l'importance des variables en permutant les valeurs d'une variable sur les exemples OOB et en mesurant la différence de l'erreur OOB avant et après permutation. Même si ces mesures ont tendance à favoriser les variables corrélées (Strobl *et al.*, 2008), elles ont pu être utilisées en télédétection afin de supprimer les variables les moins importantes et éviter le sur-apprentissage (Stumpf et Kerle, 2011), mieux interpréter et analyser des classifications (Rodriguez-Galiano *et al.*, 2012), ou pour savoir quel capteur (Guo *et al.*, 2011 ; Peters *et al.*, 2011) ou quelle échelle de segmentation (Duro *et al.*, 2012b) est le plus approprié pour identifier un objet particulier. L'ensemble des classifications réalisées dans ce chapitre est évalué grâce à l'algorithme RF par validation croisé afin de déterminer celle qui est optimale. Cette évaluation est considérée comme celle de référence puisque c'est elle qui est le plus en adéquation avec l'objectif final de cette partie à savoir une cartographie précise de la végétation urbaine.

## 3.2 Sélection des zones tests et données

Cette section est de présente les images satellites disponibles puis les sites tests qui ont été retenus et les données qui leur sont associées. Deux extraits sont sélectionnés, représentatifs de l'ensemble de nos deux agglomérations, afin d'alléger les traitements dans le but d'obtenir la segmentation la plus adaptée. Elle permet aussi de préciser les attributs calculés pour chacun des segments et la méthode de sélection des échantillons nécessaires à l'entraînement et la validation de la méthode de classification retenue.

### 3.2.1 Les images Pléiades disponibles

Dans le cadre du programme RTU Pléiades et du projet de recherche VALI-URB, le CNES nous a fourni plusieurs images satellites. Ces images THRS sont issues du satellite Pléiades 1A. Pléiades est une constellation composée de deux satellites identiques dont le premier a été lancé en décembre 2011 et le second un an plus tard. Ils sont tous les deux sur la même orbite avec un positionnement à 180° l'un de l'autre, sont orientables permettant ainsi une revisite quotidienne. Ils présentent aussi l'avantage de pouvoir réaliser des prises de vue stéréoscopiques.

Deux images sont produites par ces satellites : une panchromatique et une multispectrale composée de quatre bandes spectrales (rouge, verte, bleue, proche infrarouge). Les images Pléiades sont caractérisées par une résolution spatiale de 70 cm pour l'image panchromatique et 2,4 m pour l'image multispectrale. Pour leur livraison, les résolutions spatiales sont rééchantillonnées à 50 cm pour l'image panchromatique et 2 m.

Pour le site de Rennes une image Pléiades a été acquise le 4 octobre 2012 et pour celui de Strasbourg trois datant du 14 août 2012, du 21 septembre 2012 et du 24 avril 2013. L'ensemble de ces images a été fourni géoréférencé et orthorectifié en Lambert 93 par l'IGN Espace. La figure 3.2 présente les emprises de ces images pour nos deux sites d'études.

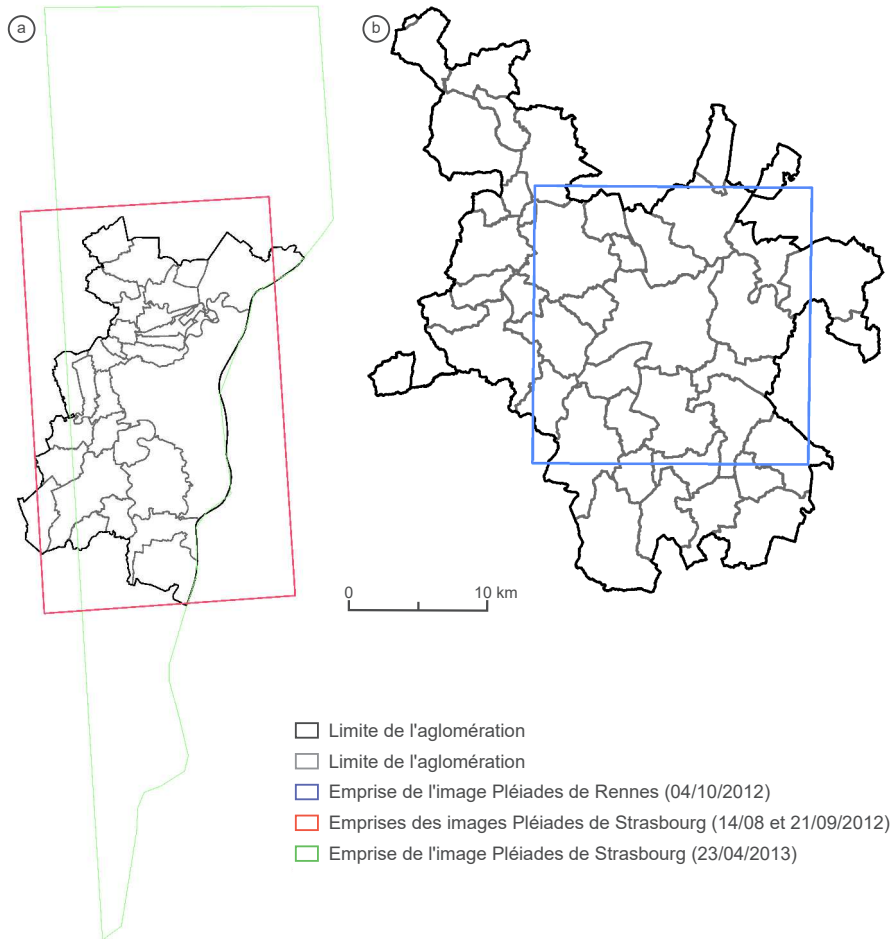
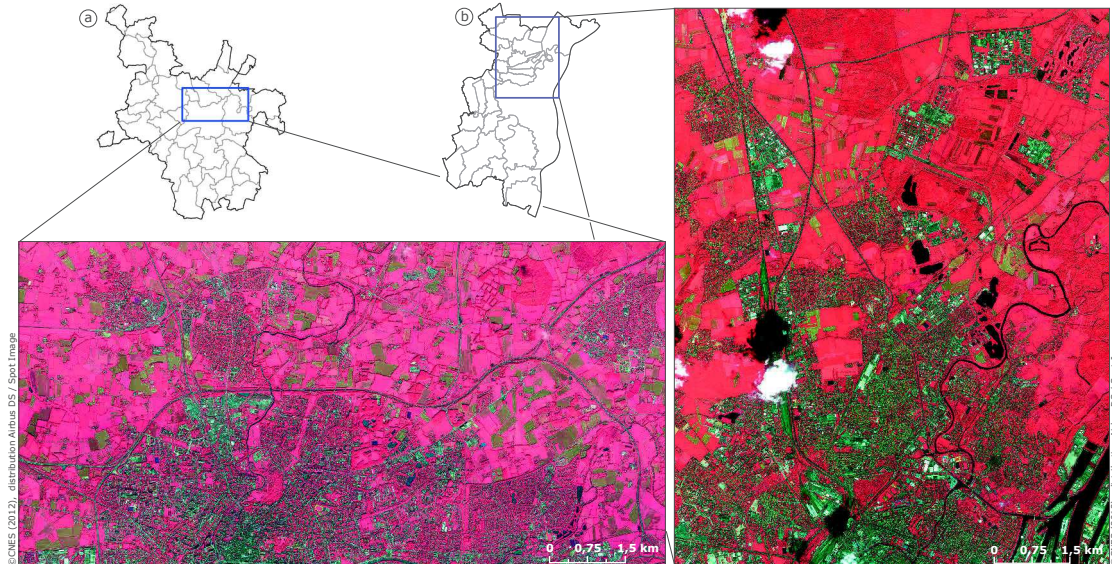


Figure 3.2 – Emprises des images Pléiades des sites de Strasbourg (a) et de Rennes (b)

### 3.2.2 Choix des sites tests

Les extraits des deux zones d'étude se situent au nord de la ville centre et s'étendent sur une surface d'environ 98 km<sup>2</sup> pour Strasbourg et 85 km<sup>2</sup> pour Rennes (figure 3.3). Leur partie centrale la plus au sud comprend le centre-ville, la plus dense, et plus on s'en éloigne moins les densités bâties apparaissent importantes. Les formes bâties sont très diverses et l'ensemble des tissus urbains des deux agglomérations, tels que les espaces résidentiels, dense à pavillonnaire, ou les zones d'activités est donc présent dans ces

sites tests. Il en va de même pour les formes végétales puisque celles très diversifiées des espaces urbains sont représentées, mais aussi celles qui caractérisent les espaces agricoles ou forestiers ainsi que les ripisylves, et ce, sur nos deux sites.



**Figure 3.3** – Localisation des zones tests dans les agglomérations de Rennes (a) et de Strasbourg (b) et extraits des images Pléiades associées

### 3.2.3 Prétraitements des images Pléiades et présentation des attributs calculés

Plusieurs prétraitements sont nécessaires afin de tester les méthodes de segmentation et les évaluer. Dans un premier temps, les fusions du canal panchromatique et des canaux multispectraux sont effectuées afin d'améliorer la résolution spatiale de ces deux images Pléiades. Afin de réaliser cette fusion, la méthode Gram-Schmidt intégrée dans le logiciel d'analyse d'image Envi 4.8 (Laben et Brower, 2000) est utilisée en recourant un rééchantillonnage par convolution cubique pour conserver au mieux la géométrie des objets. Un indice de végétation est aussi dérivé à partir des images, le MSAVI2 (*Modified Soil-Adjusted Vegetation Index 2*, Qi et al., 1994), calculé à partir de la formule :

$$MSAVI2 = \frac{2PIR + 1 - \sqrt{(2PIR + 1)^2 - 8(PIR - R)}}{2} \quad (3.5)$$

Le MSAVI est sélectionné, car Van Delm et Gulinck (2011) qui ont comparé plusieurs indices de végétation ont montré que cet indice est le plus adapté afin d'identifier différentes formes de végétation, arborée et herbacée, en milieu urbain.



Afin d'améliorer les résultats des segmentations et des classifications qui sont réalisés dans ce chapitre, l'utilisation de plusieurs prises de vue peut être une solution (§ 2.3.2). Deux images supplémentaires ont été acquises pour le site de Strasbourg le 21 septembre 2012 et le 24 avril 2013, mais aucune autre image n'était disponible pour la zone d'étude de Rennes. Ces deux images ont subi les mêmes prétraitements que les images précédentes. Dans le cadre d'une classification multi-temporelle, il est nécessaire de corriger les valeurs radiométriques des pixels des images, car celles-ci sont fortement influencées par différents facteurs tels que les conditions atmosphériques au moment de l'acquisition ou l'élévation solaire. Néanmoins, pour obtenir une bonne correction radiométrique, il est nécessaire que les images n'aient pas subi de transformations trop importantes. Or les images utilisées ici ont vu leurs valeurs spectrales fortement modifiées puisqu'elles ont été ré-échantillonnées et corrigées géométriquement. Il est toutefois possible de seulement normaliser radiométriquement les images dans le cadre où les caractéristiques physiques des objets ne sont pas recherchées (Schroeder *et al.*, 2006).

Dans cette recherche, cette normalisation est réalisée grâce à la méthode de Canty et Nielsen (2008), la *Iterative Re-weighted Multivariate Alteration Detection* (IR-MAD). Cette méthode a été développée pour des images à haute résolution spatiale et des images hyperspectrales. IR-MAD va tout d'abord identifier des paires de pixels invariants grâce à une analyse canonique des corrélations itérative entre l'image de référence et celle à corriger. Ces pixels sont ensuite utilisés afin de normaliser chaque bande spectrale selon un modèle de régression linéaire. L'image d'août est donc utilisée comme image de référence afin de normaliser les deux autres images Pléiades.

Un large éventail d'attributs est considéré afin d'évaluer la segmentation optimale grâce à la méthode de sélection par classification. Comme il a été dit, RF est un algorithme assez peu sensible au sur-apprentissage et permet donc de s'affranchir de la sélection préalable d'attributs pertinents. Ces attributs sont calculés grâce au logiciel eCognition qui est certainement le logiciel qui en permet le plus large choix. Ainsi des caractéristiques spectrales, de forme et de texture sont considérées pour chacun des segments (tableau 3.1).

### 3.2.4 Créations des données d'entraînement et de validation

En France, et en particulier pour Rennes et Strasbourg, les bases de données existantes ne permettent pas d'avoir une représentation suffisamment précise de la végétation permettant d'être utilisée comme une vérité-terrain pour la classification d'une image THRS. Il est donc nécessaire de la créer par photo-interprétation. Nous avons procédé à un échantillonnage afin de constituer ces données, car l'étendue des deux zones d'études ne permet pas leur numérisation exhaustive. Plusieurs méthodes sont régulièrement employées en télédétection afin d'obtenir ces exemples tels que l'utilisation d'un échantillonnage aléatoire, aléatoire stratifiée ou systématique (Congalton et Green, 2008).

**Tableau 3.1** – Vue d'ensemble des attributs calculés pour chaque segment

Type	Attributs	Bande spectrale ou indice
Attributs calculés pour chaque image		
Spectral	Moyenne des bandes spectrales	R, G, B, PIR, MSAVI
	Brillance	
	Écart-type des valeurs des pixels	R, G, B, PIR, MSAVI
	Valeur minimale des pixels	R, G, B, PIR, MSAVI
	Valeur maximale des pixels	R, G, B, PIR, MSAVI
	Ratio bande / brillance	R, G, B, PIR
	Différence moyenne aux voisins plus clairs	R, G, B, PIR, MSAVI
	Différence moyenne aux voisins plus foncés	R, G, B, PIR, MSAVI
Texture	Matrice de cooccurrence en niveau de gris	Homogénéité, contraste, entropie, 2 <sup>d</sup> moment angulaire, moyenne, écart-type, corrélation
Attributs de forme		
Géométrie	Aire, longueur, largeur, asymétrie, indice de bordure, compacité, densité, circularité, indice de forme	

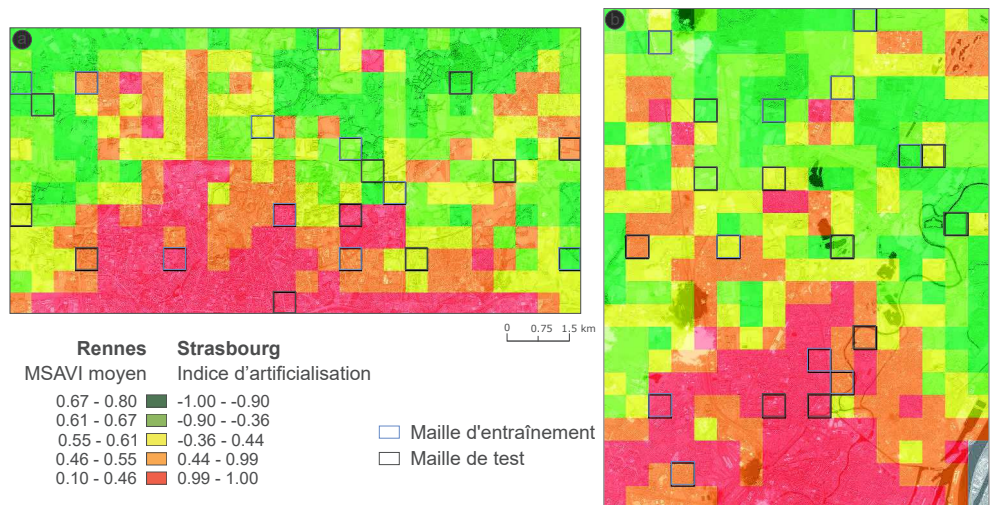
Notre choix s'est porté sur un échantillonnage aléatoire stratifié basé sur un maillage pour élaborer cet inventaire afin de saisir l'hétérogénéité de ces deux sites et ainsi obtenir des exemples plus diversifiés. Deux indices sont utilisés pour stratifier les deux zones d'études et sont calculés dans des mailles de 500×500 m. Pour Strasbourg, la stratification est réalisée à partir d'un indice d'urbanisation ( $I_U$ ) construit avec la base de données d'occupation des sols de la région Alsace grâce à la formule :

$$I_U = \frac{A_A - A_{NA}}{A_T} \quad (3.6)$$

où  $A_A$  représente les surfaces artificialisées,  $A_{NA}$  les surfaces non artificialisées et  $A_T$  la surface totale des mailles.

Pour le site de Rennes, aucune base de données d'occupation des sols n'est disponible à cette échelle pour construire cette stratification et la moyenne du MSAVI pour chaque maille est calculée afin d'être utilisée comme carte de stratification. Ces deux indices sont ensuite classés en cinq strates d'effectifs égaux pour que chaque strate soit également représentée dans l'espace (figure 3.4.a et b). Deux mailles par strate d'urbanisation sont sélectionnées de manière aléatoire afin de constituer les données d'entraînement et la végétation arborée et herbacée photo-interprétée exhaustivement pour chacune de ces mailles. La végétation arbustive n'est pas considérée, car il était impossible de distinguer visuellement les arbres des arbustes à partir de ces images. Deux mailles par strate sont aussi sélectionnées afin de constituer celles de validation. La numérisation de l'ensemble de ces mailles a été effectuée au 1/500<sup>e</sup> et les parcelles agricoles, qu'elles soient en culture ou en prairie, n'ont pas été considérées comme de la végétation herbacée. Les informations sur le parcellaire agricole sont déjà intégrées dans certaines bases de données comme le RPG avec une précision et une fréquence de mise à jour satisfaisante

(§ 2.2.2). La surface minimum de numérisation est fixée à  $6,25 \text{ m}^2$ , ce qui correspond sur l'image à un objet de  $5 \times 5$  pixels.



**Figure 3.4** – Stratification des sites de Rennes (a) par le MSAVI moyen et de Strasbourg (b) par l'indice  $I_U$  et localisation des mailles d'entraînement et de test

### 3.3 Sélection des segmentations adaptées à l'extraction de la végétation urbaine

Cette section présente dans un premier temps les paramètres sélectionnés et résultats des segmentations réalisées avec les algorithmes Multi-Resolution et Mean Shift (§ 3.3.1). La section 3.3.2 présente les résultats obtenus par chacune des méthodes d'évaluation et la section 3.3.3 compare ces méthodes et les segmentations que chacune a sélectionnées.

#### 3.3.1 Présentation des segmentations réalisées

Afin de déterminer la méthode et les paramètres de la segmentation optimale pour extraire les éléments verts urbains, plusieurs séries faisant varier un seul paramètre sont réalisées, l'échelle pour l'algorithme Multi-Resolution et le rayon spectral pour Mean Shift. Pour la première méthode le poids donné à la forme a été fixé 0,2 et celui de compacité à 0,5. Pour Mean Shift, le rayon spatial sélectionné est de 10 en utilisant une taille minimum des objets de quatre pixels.

Ces différentes séries sont composées chacune de 50 segmentations partant d'une sur à une sous-segmentation marquée visuellement. Afin de pouvoir comparer les méthodes Mean Shift et Multi-Resolution sur les sites de Strasbourg et de Rennes, le nombre d'objets de la plus fine et de la plus grossière segmentation, uniquement pour les mailles

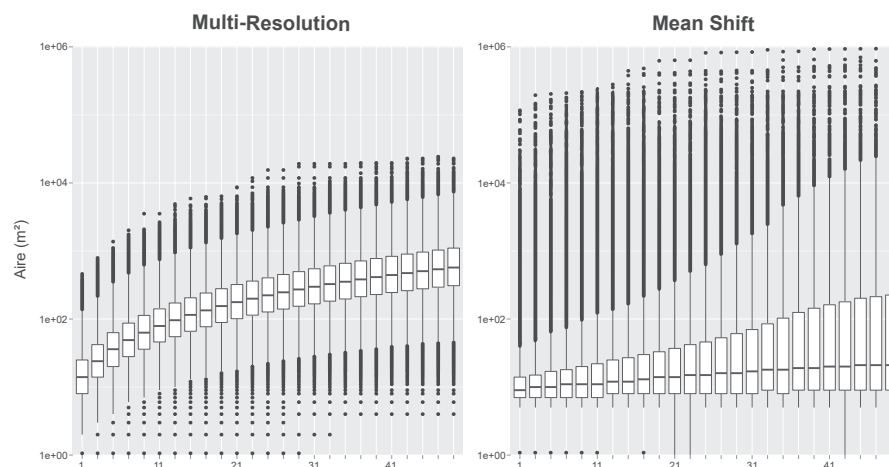
d'entraînement et de validation, est considéré. Ainsi pour chacune des séries, la segmentation la plus fine est composée de  $1\,000\,000 \pm 5\%$  de segments et la plus grossière de  $20\,000 \pm 5\%$ . Les paramètres entre ces deux extrémités ont pu être déterminés de manière empirique et ceux intermédiaires obtenus par interpolation linéaire.

De plus comme nous disposons de plusieurs images prises à des dates différentes pour le site d'étude de Strasbourg, nous avons voulu tester l'apport que peut représenter l'utilisation de multiples images dans cette première étape d'une démarche orientée-objet. Comme précédemment, les paramètres de la plus fine et de la plus grossière segmentation sont déterminés à partir du nombre d'objets. Le tableau 3.2 reprend les paramètres d'échelle et de rayon spectral minimum et maximum utilisés pour les différentes séries de segmentations.

**Tableau 3.2** – Vue d'ensemble des paramètres de segmentation utilisés pour Strasbourg et Rennes

Site d'étude	Segmentation	Image utilisée	Échelle ou rayon spectral minimum / maximum
Strasbourg	Multi-Resolution	Août	15 / 150
	Mean Shift	Août	95 / 1220
	Multi-Resolution	Multi-dates	16 / 140
	Mean Shift	Multi-dates	250 / 1920
Rennes	Multi-Resolution	Octobre	13 / 130
	Mean Shift	Octobre	80 / 1010

Les segmentations obtenues montrent de réelles différences entre les deux méthodes, quel que soit le site d'étude ou le nombre d'images considérées. Si la moyenne de l'aire des objets est assez similaire entre Mean Shift et Multi-Resolution pour un même indice de segmentation, sa distribution est quant à elle très différente. La figure 3.5 présente deux de ces distributions, celles du site de Rennes, mais celles de Strasbourg leur sont très semblables.



**Figure 3.5** – Distribution pour Multi-Resolution et Mean Shift de l'aire des segments sur le site d'étude de Rennes

Pour l'algorithme Multi-Resolution les distributions se ressemblent, quelles que soient les échelles de segmentation considérées, avec une augmentation continue de la taille des segments. Tandis que pour la méthode Mean Shift, la taille des objets est beaucoup plus diverse avec un nombre de valeurs aberrantes assez important notamment pour ceux de grandes surfaces et ce dès les segmentations les plus fines. En effet, la médiane reste pour l'ensemble des échelles relativement stable et c'est principalement l'écart entre les premiers et troisièmes quartiles qui augmente fortement. Ce sont donc principalement les objets de grandes tailles qui se trouvent encore plus imposants au fur et à mesure de l'augmentation du rayon spectral ; à l'inverse, la moitié la plus petite reste quant à elle plutôt stable. Ainsi, si pour Multi-Resolution les écarts entre les plus petits et les plus grands objets restent maîtrisés, ceux de Mean Shift sont très importants.

Les différences observées sur la figure 3.5 se traduisent assez bien visuellement (figure 3.6). Pour les quatre types d'utilisation du sol présentés ici, le nombre d'objets entre les deux méthodes et pour un même indice est comparable. Pourtant l'impression visuelle ne va pas forcément dans ce sens. En effet, le nombre d'objets apparaît moins important pour Mean Shift, dès la segmentation la plus fine pour les espaces agricoles et pour tous les autres espaces urbanisés à une échelle plus grossière. Pour cet algorithme, de grands objets ont été créés notamment au niveau des toits des bâtiments, des routes et de la végétation, mais de nombreux segments de très petite dimension restent présents aux limites de ces objets ou dans les zones les plus réfléchissantes. Pour Multi-Resolution par contre, la segmentation la plus grossière présentée apparaît toujours sur-segmentée, mais peu de segments de très petites dimensions sont encore présents.

### 3.3.2 Évaluation des segmentations pour la cartographie des éléments verts

Les résultats de l'évaluation des segmentations par les méthodes existantes sont présentés dans cette section. Nous revenons tout d'abord sur l'évaluation non supervisée (§ 3.3.2.1) puis, dans un deuxième temps sur celle supervisée (§ 3.3.2.2 et enfin présentons les résultats de l'évaluation par classification (§ 3.3.2.3).

#### 3.3.2.1 Sélection de la segmentation optimale par une approche non supervisée

L'utilisation de l'ESP2 a été appliquée à l'ensemble des séries afin de déterminer l'échelle ou le rayon spectral optimal pour la segmentation de la végétation urbaine. L'analyse du taux de variation ( $ROC_{LV}$ ) de la variance locale permet de le définir. À l'apparition du premier pic, la segmentation précédente est donc sélectionnée.

De manière générale, la segmentation choisie par l'ESP2 est plus fine avec la méthode Mean Shift qu'avec la méthode Multi-Resolution, que ce soit pour Rennes ou pour Strasbourg avec l'utilisation d'une ou de plusieurs images pour réaliser la segmentation (figure 3.7). Toutefois une limite est qu'il n'est pas possible de comparer les différentes

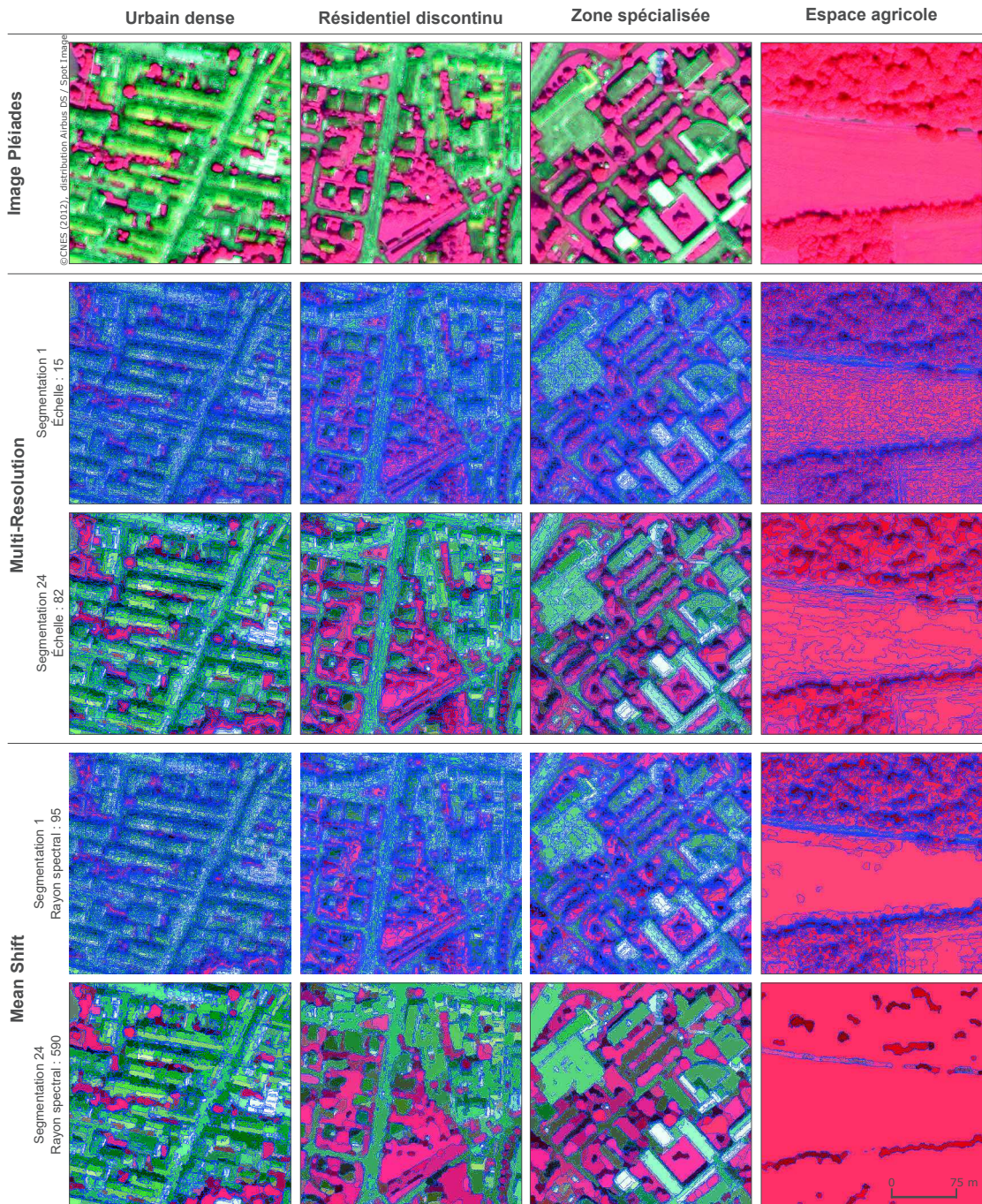


Figure 3.6 – Exemple de segmentations réalisées pour Strasbourg à deux échelles avec Multi-Resolution et Mean Shift pour quatre types d'utilisation du sol

séries afin de savoir laquelle des méthodes de segmentation est la plus adaptée à l'identification de nos objets d'intérêt. En effet, l'analyse du  $ROC_{LV}$  permet de sélectionner la segmentation lorsque la taille des objets augmente et non de comparer différentes méthodes ou différents paramétrages tels que le rayon spatial pour Mean Shift ou le poids de la forme pour Multi-Resolution. Dans un deuxième temps, il serait néanmoins possible de coupler à cette première évaluation une seconde, par une approche supervisée ou par classification, afin de déterminer la segmentation optimale à partir de celles sélectionnées pour chaque série.

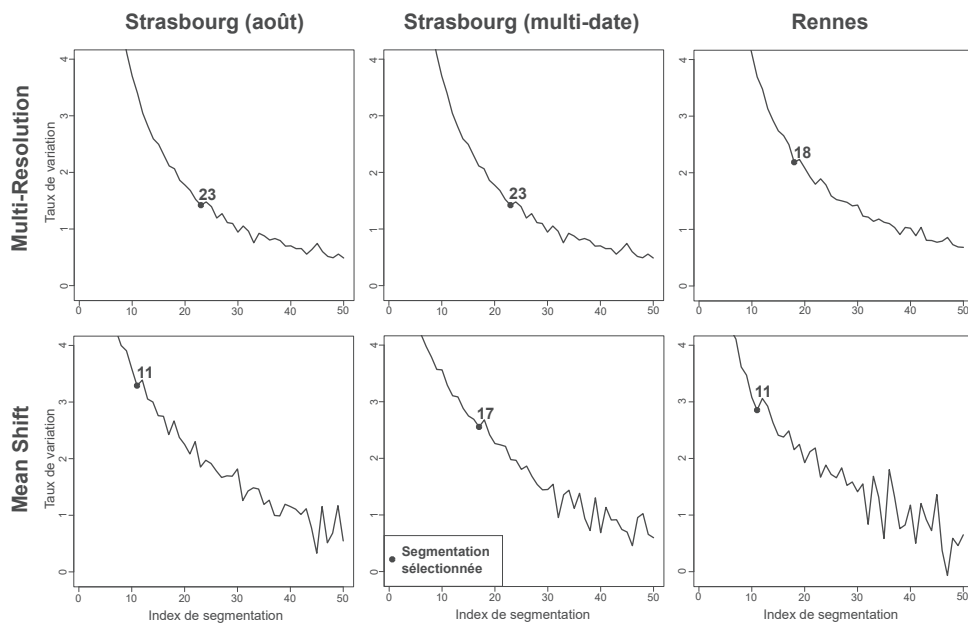


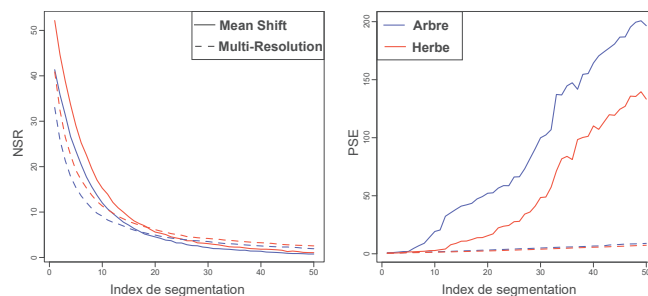
Figure 3.7 – Sélection de la segmentation optimale par l'ESP2 pour l'ensemble des stratégies

### 3.3.2.2 Sélection de la segmentation optimale par une approche supervisée

Pour évaluer des segmentations par l'intermédiaire de cette méthode il est nécessaire d'obtenir un nombre d'exemples suffisamment important pour être représentatif de la morphologie des objets d'intérêt. 100 exemples de chaque classe de végétation ont été sélectionnés. Ces exemples sont tirés de la vérité-terrain déjà construite pour chacun des deux sites tests et ont dû remplir un certain nombre de critères. Tout d'abord, seuls les objets ne touchant pas la bordure des mailles d'entraînement et de validation ont été considérés afin que la totalité de l'objet soit intégrée et non seulement une partie. Ce critère a eu tendance à supprimer ceux qui sont les plus grands, tel que les patches forestiers par exemple. De plus 5 % des plus grands et 5 % des plus petits restants ont aussi été supprimés afin de ne pas trop biaiser les mesures pour le calcul des indicateurs de validité des segmentations. Tous ces critères intégrés, 100 exemples

de végétation arborée et 100 exemples de végétation herbacée ont ainsi été sélectionnés de manière aléatoire pour chacun des sites tests afin de constituer notre vérité-terrain. Les segmentations sont évaluées dans cette recherche par classe, mais toutes les classes pourraient très bien être considérées dans leur ensemble.

Les résultats des segmentations que nous avons constatés visuellement (figure 3.6) apparaissent assez clairement sur les graphiques présentant la sur (NSR) et la sous-segmentation (PSE) (figure 3.8). L'effet de sur-segmentation apparaît très clairement pour les deux algorithmes lors des segmentations les plus fines et diminue assez rapidement lorsque les segmentations deviennent plus grossières. L'indicateur de sous-segmentation montre, lui, des résultats très différents selon les deux méthodes. Il reste en effet assez stable pour le Multi-Resolution tandis que pour Mean Shift la sous-segmentation devient rapidement très marquée, quelle que soit la classe considérée. Ainsi pour cette dernière, les segmentations moyennes se trouvent fortement sur et sous-segmentées.



**Figure 3.8** – Résultats de NSR et de PSE pour les séries de segmentation réalisées pour Strasbourg avec l'image d'août

Le calcul de l'ED2 permet d'obtenir un indice synthétique combinant ceux de sur et de sous-segmentation. Comme avec l'ESP2 les segmentations sélectionnées par cette méthode sont plus grossières avec Multi-Resolution qu'avec Mean Shift (figure 3.9). Pour chacune des séries réalisées, l'effet de sous segmentation a une importance très forte dans le calcul de l'ED2 qui apparaît beaucoup moins visible pour celle réalisée avec Multi-Resolution. De plus, cet indicateur permet de mettre en évidence que la végétation arborée nécessite une segmentation plus fine que celle herbacée. Seules les séries réalisées avec Mean Shift sur le site de Rennes montrent que la segmentation sélectionnée est la même pour les deux classes.

Il est aussi possible de comparer les valeurs de l'ED2 afin de sélectionner la segmentation la meilleure pour chaque classe est chaque site d'études. Le tableau 3.3 résume les meilleurs résultats obtenus pour chaque série, par site et par classe. La segmentation ayant un ED2 le plus faible est donc celle optimale. Ainsi pour Rennes et pour Strasbourg la méthode qui apparaît la plus adaptée à la segmentation de la végétation arborée et herbacée et la Multi-Resolution. Pour Strasbourg celle utilisant les images prises à différentes dates apparaît la plus performante en utilisant un paramètre d'échelle de 95



pour la classe arbre et de 103 pour l'herbe. Pour Rennes ce paramètre est de 109 pour la végétation arborée et de 121 pour celle herbacée.

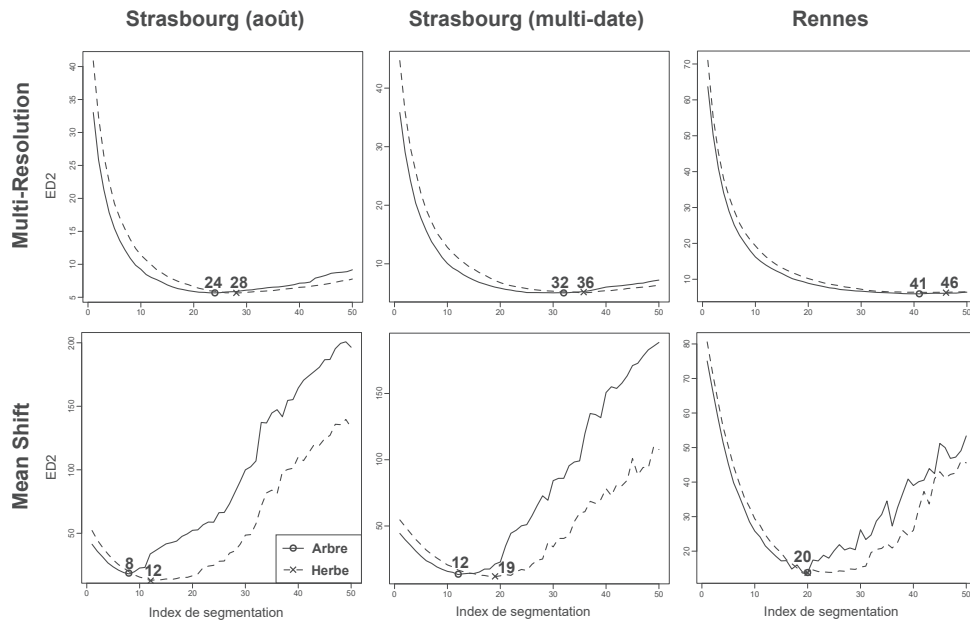


Figure 3.9 – Sélection de la segmentation optimale par l'ED2 pour l'ensemble des stratégies

Tableau 3.3 – Sélection de la segmentation optimale par série et par classe de végétation avec l'ED2 (en gris la meilleure segmentation par ville et par type de végétation)

Ville	Segmentation	Image	Indice de segmentation	Classe	PSE	NSR	ED2
Strasbourg	Multi-Resolution	Août	24	Arbre	3,91	4,11	5,67
			28	Herbe	3,61	4,35	5,65
	Mean Shift	Août	8	Arbre	9,05	15,58	18,02
			12	Herbe	4,08	12,14	12,81
	Multi-Resolution	Multi-dates	32	Arbre	3,7	3,41	5,03
			35	Herbe	3,45	3,72	5,08
Mean Shift	Multi-dates	12	Arbre	3,57	12,86	13,35	
Rennes	Multi-Resolution	Octobre	41	Arbre	4,17	4,23	5,94
			46	Herbe	4,58	4,37	6,33
	Mean Shift	Octobre	20	Arbre	9,51	10,02	13,81
			20	Herbe	5,53	12,18	13,38

### 3.3.2.3 Sélection de la segmentation optimale par classification

Dans cette section, l'ensemble des segmentations n'est pas testé avec la méthode de classification RF en raison du temps nécessaire à la réalisation de cette stratégie. En effet en plus du temps nécessaire à l'entraînement et la prédiction, il faut aussi considérer le

calcul des attributs pour l'ensemble des segmentations. Seulement une segmentation sur deux a donc été testée.

Pour l'ensemble des classifications, les mêmes paramètres ont été utilisés pour RF. Le nombre d'arbres ainsi que le nombre de variables aléatoires considérées à chaque nœud ont été laissés par défaut, soit 500 arbres et la racine carrée du nombre total de variables. De plus, sur l'ensemble des mailles d'entraînement et de validation, les deux classes de végétation sont fortement minoritaires par rapport à la classe autre. La végétation arborée représente 21 % pour les deux sites et celle herbacée 10 % pour Rennes et 13 % pour Strasbourg. Ce déséquilibre entre les classes peut entraîner un biais important dans les résultats de la classification et diminuer les performances. Beaucoup de méthodes d'apprentissage supervisé ont tendance à favoriser les classes surreprésentées et sous-estiment donc celles qui sont minoritaires (He et Garcia, 2009). Cet effet peut-être important avec RF (Stumpf et Kerle, 2011). RF permet, pour corriger le déséquilibre entre les classes et cet effet, de réaliser des échantillonnages bootstrap stratifiés (Chen *et al.*, 2004). Dans ce travail, un même nombre d'exemples de chaque classe a donc été utilisé pour la construction de chaque arbre.

Afin de mieux tester les performances de la segmentation et de son effet sur celles de la classification, l'attribution des classes aux segments s'est faite de manière différenciée selon leur appartenance aux données d'entraînement ou de validation. Pour les données d'entraînement, les segments ont été labélisés selon la classe majoritaire en termes de surface tandis que pour les données de validation, les résultats de la segmentation ont été intersectés avec la vérité-terrain. L'objectif de cette différenciation est d'éviter que l'entraînement de RF ne soit trop fortement bruité par le fait que des exemples avec exactement les mêmes attributs appartiennent à plusieurs classes différentes. Pour les segments de validation, l'intersection permet, par l'évaluation des performances, de vérifier l'adéquation entre la segmentation et la vérité-terrain. Pour évaluer les performances des classifications, la F-mesure par classe a été calculée, car cette mesure de précision synthétique pour chaque classe permet de s'affranchir du déséquilibre important entre elles.

L'évaluation des segmentations par classification montre des résultats intéressants (figure 3.10). Tout d'abord, à la différence de l'évaluation par la méthode supervisée, une segmentation plus fine de la végétation arborée comparée à celle herbacée ne paraît pas toujours comme adaptée. Il apparaît tout d'abord assez nettement que l'algorithme Multi-Resolution obtient des performances relativement constantes, quelle que soit l'échelle de segmentation choisie. En effet pour chaque série les écarts des F-mesures entre la meilleure et la moins bonne segmentation sont de 1,8 à 2,8 % pour la végétation arborée et de 2,8 à 6,3 % pour la végétation herbacée. Avec Mean Shift par contre, ces écarts explosent et vont de 26,4 à 43,7 % pour les arbres et de 22,6 à 36,7 % pour l'herbe. Pour Strasbourg, les meilleurs résultats sont obtenus avec une segmentation multi-dates et la méthode Multi-Resolution en utilisant un paramètre d'échelle de 33 et de 38 respectivement pour les classes arbre et herbe. Pour Rennes c'est avec le Multi-Resolution

et une échelle de 25 que les performances sont les meilleures pour la végétation arborée et Mean Shift avec un rayon spectral de 80 pour celle herbacée.

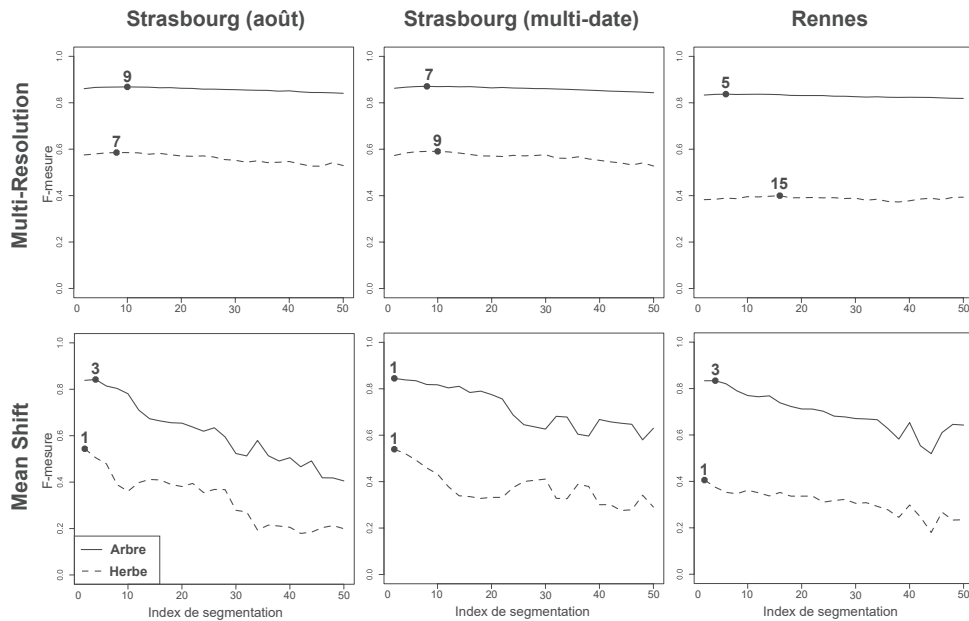


Figure 3.10 – Sélection de la segmentation optimale par classification avec RF pour l'ensemble des stratégies

### 3.3.3 Comparaison des méthodes d'évaluation de segmentation

Tout d'abord, les paramètres des segmentations sélectionnées avec les méthodes supervisée et non supervisée et avec les deux algorithmes pour le site de Strasbourg sont relativement équivalents approchant donc des conclusions de Belgiu et Drăguț (2014). En effet sur ce site, celles sélectionnées par l'ESP2 sont situées entre les deux sélectionnées par classes par l'ED2 excepté pour Multi-Resolution effectuée sur plusieurs images, où la segmentation sélectionnée par la méthode non supervisée est la moins grossière. L'évaluation des segmentations sélectionnées ici montre donc des résultats assez similaires lorsque celles-ci sont ensuite classées avec RF (tableau 3.4).

Tableau 3.4 – F-mesures obtenues pour les segmentations sélectionnées avec l'ESP2, l'ED2 et RF pour le site de Strasbourg (en gris les meilleurs résultats obtenus)

	Arbre			Herbe		
	ED2	ESP	RF	ED2	ESP	RF
Multi-Resolution – août	0,86	0,86	0,87	0,57	0,57	0,59
Multi-Resolution – multi-dates	0,86	0,87	<b>0,87</b>	0,56	0,57	<b>0,59</b>
Mean Shift – août	0,81	0,78	0,84	0,36	0,36	0,54
Mean Shift – multi-dates	0,82	0,78	0,84	0,33	0,34	0,54

Pour Rennes par contre, ces deux méthodes de sélection divergent fortement. En effet, l'utilisation de la métrique ED2 fait ressortir préférentiellement des segmentations plus grossières qu'avec l'ESP2 et ce particulièrement pour la méthode Multi-Resolution où l'écart du paramètre d'échelle est très important. Il en est de même, mais dans une moindre mesure, pour la méthode Mean Shift. Ces écarts vont ainsi se répercuter assez fortement sur les performances obtenues par classification à partir de ces segmentations (tableau 3.5). Si les différences sont assez faibles entre les résultats obtenus avec les segmentations Multi-Resolution malgré un écart d'échelle important, elles commencent à être marquées pour Mean Shift à cause de la chute rapide des performances lorsque le rayon spectral augmente (figure 3.10).

**Tableau 3.5** – F-mesures obtenues pour les segmentations sélectionnées avec l'ESP2, l'ED2 et RF pour le site de Rennes (en gris les meilleurs résultats obtenus)

	Arbre			Herbe		
	ED2	ESP	RF	ED2	ESP	RF
Multi-Resolution	0,82	0,83	<b>0,84</b>	0,39	0,40	0,40
Mean Shift	0,72	0,77	0,83	0,34	0,36	<b>0,41</b>

Les segmentations sélectionnées avec RF, que l'on peut considérer comme les plus précises puisqu'elles répondent le mieux à notre objectif thématique, montrent des résultats clairement différents. En effet pour chaque série de segmentations, celles qui permettent d'obtenir les meilleures performances font partie des segmentations plus fines que celles détectées par les deux autres stratégies de sélection. Pour obtenir les meilleurs résultats après classification, il apparaît important que les objets ne soient pas trop sur-segmentés. C'est ce qui apparaît en comparant les courbes résultantes de la sélection par l'ED2 et par RF pour Multi-Resolution, mais surtout que les objets ne soient pas sous-segmentés. En effet les résultats obtenus pour Mean Shift montrent que dès qu'une sous-segmentation apparaît, et ce dès les segmentations les plus fines avec cet algorithme, les résultats se trouvent fortement dégradés. Cela constitue donc une limite importante liée à l'utilisation des méthodes de sélection de segmentation supervisée. Le défaut de ce type de méthodes et qu'elle cherche à trouver un équilibre entre sur et sous-segmentations, ce qui ne pose pas de problèmes quand ces deux effets ne sont pas de manière trop importante concomitants, mais qui entraîne une très mauvaise sélection quand ces deux effets sont marqués simultanément comme dans le cas de Mean Shift et des paramètres sélectionnés ici. Il est donc préférable dans le cadre d'une classification avec l'algorithme RF d'avoir une sur-segmentation assez marquée afin d'éviter toute sous-segmentation qui, elle, nuit fortement aux résultats.

D'autres limites à cette méthode, mais aussi aux deux autres sont à mentionner ou à rappeler, tout comme les avantages respectifs de chacune. Concernant l'ESP2 il est très rapide à mettre en œuvre notamment quand le choix de la segmentation se porte sur Multi-Resolution puisqu'il est directement intégré à eCognition. Toutefois, il ne permet pas de comparer d'autres paramètres que le changement d'échelle et il est très sensible au choix de la valeur utilisé pour l'incrémentation. En effet si celle-ci est trop grande, alors il est possible qu'aucun pic n'apparaisse, ce qui dans notre cas, s'est produit si seulement

25 et non 50 segmentations étaient réalisées entre la valeur minimum et maximum du paramètre d'échelle ou du rayon spectral. Pour la méthode supervisée utilisée ici, elle reste relativement rapide puisqu'avec quelques exemples elle permet d'obtenir une mesure de sur et une de sous-segmentation. De plus, elle permet de comparer différentes méthodes de segmentation et d'obtenir des résultats pour chaque type d'objet qui est recherché dans l'image. Toutefois, le compromis fait entre les effets de sur et de sous-segmentations peut engendrer la sélection d'une segmentation non adaptée. Enfin, réaliser cette évaluation en ayant recourt à une méthode de classification permet d'avoir les mêmes avantages que celle supervisée. Elle est de plus directement reliée à l'objectif de cartographie précise des éléments arborés et herbacés urbains. Toutefois, cette méthode est très longue à mettre en place. Il est en effet nécessaire de réaliser la caractérisation des segments de chaque segmentation pour ensuite les classer et les évaluer. Cette étape pourrait être accélérée par la prise en compte unique de l'erreur OOB, même si cette erreur est optimiste, sans qu'il soit donc nécessaire de le faire par validation croisée. Mais l'utilisation d'une méthode de classification supervisée nécessite un nombre d'exemples pour l'entraînement assez conséquent, sauf si ces exemples sont réellement informatifs. L'utilisation et le développement de méthodes pour la sélection de ce type d'exemples pourraient donc être une alternative intéressante.

Un objectif de ce chapitre, en plus de comparer les différentes stratégies de sélection de segmentation, était de vérifier l'apport que peut représenter l'utilisation d'images multi-temporelles pour la segmentation. Même si les meilleurs résultats sont obtenus avec cette stratégie, ils sont très proches de ceux obtenus en réalisant la segmentation sur une seule image (tableau 3.4) pour une complexité fortement accrue. Il a tout d'abord été nécessaire de réaliser une normalisation des images de septembre et d'avril afin que les différences trop importantes au niveau de la valeur des pixels ne puissent pas perturber cette étape. La segmentation à proprement parler est elle aussi rendue plus complexe et donc plus longue pour chacune des méthodes. Finalement, cet apport reste donc, pour l'identification de la végétation arborée et herbacée en milieu urbain, assez limitée.

Il s'agissait aussi de comparer les performances de chaque méthode de segmentation pour l'extraction de la végétation. Les segmentations réalisées avec Mean Shift apparaissent pour Strasbourg légèrement moins performantes que celles réalisées avec Multi-Resolution. Elle arrive néanmoins à la surpasser pour la végétation herbacée de Rennes. Toutefois, son principal défaut est qu'elle apparaît beaucoup moins stable que la segmentation Multi-Resolution rendant ainsi beaucoup moins facile son utilisation dans un contexte opérationnel. Il est nécessaire de bien comprendre cette méthode et de bien mieux la paramétrer que ce que nous avons fait dans le cadre de cette recherche puisque nous avons décidé de jouer principalement sur un seul paramètre ce qui n'est assurément pas suffisant pour cette méthode.

Les meilleures performances obtenues pour la végétation arborée sont très convenables pour nos deux sites tests, mais celles de la végétation herbacée sont au contraire peu satisfaisantes. Pour Rennes notamment, beaucoup de confusion sont présentes dans

les espaces agricoles où de grandes parcelles ont pu être assignées à la classe herbe, impactant ainsi de manière importante les résultats. Cet effet se fait moins sentir sur le site de Strasbourg. Des tests non présentés ici ont été réalisés afin de vérifier l'importance de l'utilisation de plusieurs images dans le processus de classification qui ont montré qu'avec une seule image, les performances obtenues pour la végétation herbacée de Strasbourg étaient du même ordre que celles obtenues sur Rennes (Rougier et Puissant, 2014). C'est l'utilisation de caractéristiques multi-temporelles qui a permis de limiter ces erreurs, notamment lorsque les parcelles ont vu leur occupation du sol changer, d'une culture à un sol nu par exemple. Des tests plus approfondis sont réalisés dans le prochain chapitre de cette thèse afin de vérifier l'amélioration des performances qu'apportent ces caractéristiques pour l'extraction de la végétation de Strasbourg (chapitre 4).

## Conclusion

La première étape d'une démarche orientée-objet nécessite une segmentation de l'image en régions homogènes. De nombreuses méthodes existent et deux ont été comparées dans ce chapitre, Multi-Resolution et Mean Shift, afin de déterminer la segmentation optimale, celle qui permettait d'obtenir les classifications de la végétation arborée et herbacée les plus performantes. Trois méthodes d'évaluation ont permis d'évaluer les résultats de ces deux algorithmes, une non supervisée, une supervisée et une par classification en considérant cette dernière comme la référence. Les méthodes non supervisée (ESP) et supervisée (ED2) testées ont montré qu'elles permettaient de sélectionner des segmentations dont les performances sont proches de celles optimales sélectionnées par la méthode par classification pour l'algorithme Multi-Resolution, mais que les segmentations sélectionnées pour Mean Shift sont beaucoup moins performantes. En effet, cet algorithme a tendance à fortement sur et sous-segmenter l'image conjointement. Or les résultats ont montré qu'une sur-segmentation, même marquée, ne dégrade que peu les performances des classifications tandis qu'une sous-segmentation va fortement les impacter. Multi-Resolution permet d'obtenir des segments de tailles relativement homogènes et va ainsi révéler des résultats plus stables permettant une utilisation plus aisée de cette méthode pour la segmentation d'image. De plus, les performances les meilleures sont majoritairement obtenues avec cette méthode.

La comparaison des méthodes d'évaluation a montré que celles supervisée et non supervisée sont assez limitées et peu précises même si elles sont beaucoup plus rapides à mettre en place que l'évaluation par classification. En effet, la méthode non supervisée utilisée ne permet pas de comparer plusieurs méthodes de segmentation ni de définir quels paramètres sont les plus appropriés pour un type d'objet. L'évaluation supervisée, quant à elle, cherche un compromis entre sur et sous-segmentations, ce qui, lorsque les objets sont de tailles très diverses comme ceux de végétation, entraîne la sélection d'une segmentation ne permettant pas d'obtenir de bonnes performances lors de la classification.

De plus pour le site de Strasbourg qui dispose de plusieurs images, la réalisation de segmentations multi-temporelles ont montré des performances meilleures que celles mono-dates. Toutefois, cet apport reste très limité et complexifie fortement le processus de segmentation, ce qui rend, au final, peu efficiente l'utilisation de cette stratégie pour l'extraction de la végétation arborée et herbacée urbaine.

Dans ce chapitre, de nombreux échantillons ont été utilisés pour permettre l'entraînement de la méthode de classification supervisée *Random Forest*. Ce travail est une tâche longue et fastidieuse alors que l'utilisation d'exemples informatifs qui saisisaient toute la diversité des classes d'intérêt permettrait de réduire l'effort nécessaire à cette tâche et ainsi rendre plus opérationnelle l'utilisation d'images satellites pour l'extraction de la végétation urbaine. Le prochain chapitre de cette thèse a donc pour objectifs de présenter les méthodes qui peuvent être utilisées pour la création de ces données d'entraînement et de proposer de nouvelles méthodes d'apprentissage actif.

*Valorisations scientifiques associées à ce chapitre :*

**Rougier S.**, Puissant A., 2014, « Improvements of urban vegetation segmentation and classification using multi-temporal Pléiades images », *Proceedings of the 5th International Conference on Geographic Object-Based Image Analysis*, Thessalonique, Grèce, 21–24 Mai 2014

Puissant A., **Rougier S.**, Stumpf A., 2014, « Object-oriented mapping of urban trees using Random Forest classifiers », *International Journal of Applied Earth Observation and Geoinformation*, vol. 26, pp. 235–245

**Rougier S.**, Puissant A., 2013, « Analyse, caractérisation et suivi des espaces urbains et péri-urbains à partir de données multi-sources », *Colloque biennal des Zones Ateliers*, Paris, France, 12–13 décembre 2013





# Chapitre 4

## Comparaison de stratégies d'échantillonnage pour la classification d'images THRS

### Sommaire

---

Introduction . . . . .	112
4.1 La sélection d'échantillons pour la classification d'images . . . . .	112
4.2 Méthodes proposées pour la sélection d'échantillons d'apprentissage . . . . .	116
4.2.1 Apprentissage actif spatial multi-classes (MCSAL) . . . . .	118
4.2.2 Apprentissage actif spatial multi-classes prenant en compte la diversité (MCSAL <sub>D</sub> ) . . . . .	118
4.2.3 Apprentissage actif spatial multi-classes stratifié (MCSAL <sub>S</sub> ) . . . . .	119
4.2.4 Échantillonnage aléatoire stratifié (SRGS) . . . . .	120
4.3 Préparation des données . . . . .	120
4.4 Protocole de tests . . . . .	121
4.5 Évaluation des méthodes d'apprentissage actif proposées . . . . .	123
4.5.1 Effet de la normalisation de l'entropie . . . . .	123
4.5.2 Sensibilité des paramètres du critère de diversité . . . . .	124
4.5.3 Choix de la taille des régions et comparaison des méthodes d'apprentissage actif avec l'aléatoire stratifié . . . . .	126
4.5.4 Apport des données auxiliaires pour l'amélioration des performances . . . . .	130
4.5.5 Évaluation des erreurs selon l'aire et la pureté de la classe . . . . .	131
Conclusion . . . . .	135

---

## Introduction

Le chapitre précédent a montré la reproductibilité des méthodes de classification supervisée dans le cadre d'une démarche orientée-objet afin de cartographier la végétation de Rennes et de Strasbourg. Toutefois, pour réaliser l'entraînement de *Random Forest*, un nombre conséquent d'exemples a été utilisé afin d'obtenir des résultats qui restent peu performants notamment pour la végétation herbacée. En effet, pour obtenir de bons résultats avec une classification d'image supervisée, il est nécessaire que le jeu d'entraînement utilisé soit de grande taille, et ce particulièrement pour les classes les plus hétérogènes (Chen et Stow, 2002) ou que les données d'entraînement soient représentatives de l'espace étudié (Foody et Mathur, 2004). La collecte de ces données d'entraînement, que ce soit par photo-interprétation ou par des relevés sur le terrain, est comme nous l'avons déjà dit, un travail long, fastidieux et qui représente donc un coût très important. Avec la photo-interprétation, la qualité des exemples utilisés peut-être différents en fonction de l'interprète, mais aussi de sa lassitude induite par le temps passé à effectuer cette tâche (Van Coillie *et al.*, 2014). Il est donc nécessaire, afin que l'utilisation d'image satellite pour la cartographie soit réellement opérationnelle, de réduire l'effort pour la labélisation d'exemples.

L'objectif général de ce chapitre est donc de proposer une méthode qui permet de réduire de manière importante le nombre d'exemples à annoter tout en conservant une identification performante de la végétation urbaine. Plusieurs stratégies de sélection d'échantillons sont testées et évaluées afin de déterminer celle qui est la plus performante. Tout d'abord, les différentes méthodes régulièrement utilisées en télédétection pour sélectionner des échantillons représentatifs des objets présents sur les images sont introduites (§ 4.1). Dans un deuxième temps, les méthodes d'apprentissage actif développées dans cette recherche et celle de référence pour les évaluer sont explicitées (§ 4.2), les données utilisées (§ 4.3) ainsi que le protocole de test suivi afin d'évaluer ces différentes méthodes (§ 4.4). Enfin, la dernière section de ce chapitre présente les résultats des différents tests réalisés et permet de mieux comprendre la distribution des erreurs et leurs causes.

### 4.1 La sélection d'échantillons pour la classification d'images

Plusieurs stratégies sont possibles afin de réduire la taille du jeu d'entraînement pour la réalisation d'une classification supervisée dans le cadre d'applications de télédétection. Traditionnellement, un échantillonnage stratifié est souvent utilisé, mais nécessite d'avoir des données au préalable afin de créer cette stratification (Congalton et Green, 2008). Plus récemment, différentes autres méthodes ont été proposées pour réduire ce coût telles que l'apprentissage semi-supervisé ou l'apprentissage actif (Persello et Bruzzone, 2014).

Le principe des méthodes d'apprentissage semi-supervisées est, avec un nombre limité d'exemples d'entraînement, de profiter des exemples non labélisés qui sont généralement plus nombreux. Ces méthodes tentent de profiter des avantages des deux grandes familles de classifications, supervisées et non supervisées (Cornuéjols et Miclet, 2010). L'idée de ces approches est ainsi d'exploiter les informations dans l'espace des données de ces exemples non labélisés afin de mieux modéliser la distribution des classes recherchées. L'objectif est donc d'améliorer les règles de classification obtenues à partir des données étiquetées grâce à ces nouvelles informations (Persello et Bruzzone, 2014). De nombreuses méthodes ont été développées et appliquées à des applications de télédétection et ont montré une amélioration importante des performances des classifications comparée à un apprentissage supervisé classique (Chi et Bruzzone, 2007 ; Dopido *et al.*, 2013).

Dans cette recherche, nous nous concentrons sur les méthodes d'apprentissage actif. L'apprentissage actif est une approche qui vient de la pédagogie. Le principe est d'améliorer l'apprentissage des élèves en leur donnant un rôle actif c'est-à-dire en les mettant en situation afin d'augmenter les échanges au sein du groupe. L'idée est de les amener à atteindre les objectifs pédagogiques le plus rapidement possible en leur permettant de construire leurs propres connaissances (Bondu et Lemaire, 2008). Cette approche a été mise en parallèle avec l'informatique et les méthodes d'apprentissage automatique.

Dans le chapitre 3, nous avons utilisé une méthode d'apprentissage passif, c'est-à-dire que nous avons donné au modèle statistique un ensemble de données labélisées et caractérisées afin qu'il apprenne et qu'il puisse ensuite le généraliser à l'ensemble des données sans label. Aucune interaction n'intervenait. Lors d'une classification par apprentissage actif par contre, il existe une interaction entre le modèle statistique, l'élève, et l'utilisateur ou l'oracle, le professeur. Par un jeu de questions-réponses, c'est-à-dire un processus itératif, le modèle statistique va demander à l'oracle de labéliser les exemples qui seront pour lui le plus informatif (Settles, 2010). L'objectif est ainsi de réduire le nombre d'exemples nécessaire au modèle pour son apprentissage et donc réduire le coût humain indispensable à la réalisation d'une classification de qualité. Les méthodes d'apprentissage actif ont vu le jour en informatique à partir des années 1990 (Zhang et Veenker, 1991) et ont déjà démontré qu'elles permettaient généralement d'obtenir de meilleurs résultats lorsqu'elles sont comparées à un simple échantillonnage aléatoire (Settles, 2010 ; Tuia *et al.*, 2009b). Ces méthodes ont été récemment utilisées avec de bons résultats aussi bien pour des classifications d'images multispectrales et d'images hyperspectrales (Tuia *et al.*, 2009b ; Demir *et al.*, 2011) que de données multi-sources, lidar et hyperspectrale (Zhang *et al.*, 2015). Tuia *et al.* (2011b) identifient trois grandes familles de stratégies qui sont couramment utilisées pour la sélection de nouveaux exemples selon qu'elles sont basées sur la distance à la séparatrice (Campbell *et al.*, 2000), le désaccord d'un comité de modèle (Seung *et al.*, 1992) et la mesure de probabilités *a posteriori* (Lewis et Gale, 1994).

Les stratégies basées sur la distance à la séparatrice sont les plus utilisées pour des applications de télédétection et se basent sur l'algorithme de classification SVM (Tuia

*et al.*, 2011b ; Demir *et al.*, 2011 ; Pasolli *et al.*, 2014). Les distances entre chaque exemple non labélisé et l'hyperplan sont calculées et les exemples qui sont considérés comme les plus incertains et donc les plus informatifs sont ceux qui se situent le plus proche de cette séparatrice. Le ou les candidats sélectionnés vont ensuite être ajoutés afin de devenir des vecteurs support dans l'itération suivante.

Les stratégies par comité consistent à construire différents modèles de classification à partir de données légèrement différentes et de mesurer le désaccord entre les modèles pour chaque exemple non labélisé. Les exemples ajoutés sont ceux où ce désaccord est le plus fort. Cette stratégie peut être réalisée avec n'importe quel classifieur. Pour créer des modèles différents, il est possible par exemple de recourir à des méthodes bootstrap afin de créer des modèles sur des exemples différents (Tuia *et al.*, 2009b ; Copa *et al.*, 2010 ; Stumpf *et al.*, 2014). Une autre possibilité est de les construire en utilisant seulement une partie des caractéristiques. Chaque modèle peut ainsi être construit avec seulement une partie des bandes spectrales (Di et Crawford, 2012) ou en séparant les données issues de capteurs différents (Zhang *et al.*, 2015).

La troisième stratégie d'apprentissage actif est la mesure de probabilité *a posteriori* pour sélectionner les exemples les plus informatifs pour le classifieur. Ces stratégies peuvent se rapprocher de celles à comité de modèles sauf qu'un seul modèle est entraîné et qu'à la place du désaccord ce sont les probabilités d'appartenance aux classes considérées qui sont utilisées pour déterminer les candidats. Elles peuvent aussi prendre en compte, non pas la probabilité d'appartenance d'un candidat, mais considérer le changement de ces probabilités pour l'ensemble du jeu de données quand un nouvel exemple est ajouté au jeu d'entraînement. Dans ce cas, l'exemple ajouté est celui qui permet de maximiser ces changements. Cette stratégie peut être utilisée avec un grand nombre de méthodes de classification. Tuia *et al.* (2011a), à partir d'une analyse discriminante linéaire, ajoutent au jeu d'entraînement les exemples dont la probabilité d'appartenance à une classe est la plus ambiguë. Rajan *et al.* (2008) utilisent avec cette stratégie l'algorithme du maximum de vraisemblance. Ici, chaque exemple non labélisé est ajouté au jeu d'entraînement avec la classe dont la probabilité d'appartenance est la plus grande et la divergence de Kullback-Liebert est calculée entre les probabilités des modèles avec et sans le nouvel exemple. L'exemple finalement ajouté est celui qui permettra de maximiser cette divergence.

Un problème des stratégies d'apprentissage est le temps d'exécution nécessaire à la réalisation d'une classification du fait du nombre d'itérations qui, parfois, peut être important. Plusieurs solutions ont été proposées afin de réduire le nombre d'itérations nécessaire pour obtenir une bonne classification. Une première est de, non plus sélectionner une seule instance à chaque itération, mais d'en sélectionner un lot (Demir *et al.*, 2011) ou, plus spécifiquement pour l'analyse d'image, une région (Stumpf *et al.*, 2014).

D'autres méthodes ont été proposées afin d'intégrer un critère de diversité pour éviter que les informations apportées par les nouveaux exemples ne soient redondantes. Le but est donc d'obtenir plus rapidement un ensemble d'exemples qui représente toute

la complexité de l'image étudiée. Ainsi Demir *et al.* (2011) vont utiliser deux mesures afin de maximiser la diversité entre les exemples déjà labélisés et les candidats dont l'incertitude est grande. La première va mesurer le cosinus de l'angle formé dans l'espace des données entre ces deux types d'exemples tandis que la deuxième est basée sur un clustering. Cette deuxième réalise un clustering avec k-means, en un nombre de classes défini par l'utilisateur sur les exemples sélectionnés comme les plus incertains, et l'exemple le plus proche du centre de chaque cluster, le médoïde, est ajouté au jeu d'entraînement. L'utilisation d'une méthode de clustering afin de sélectionner des instances diversifiées a aussi été employée par Volpi *et al.* (2012) en utilisant une méthode hiérarchique basée aussi sur l'algorithme k-means. Dans cette recherche, chaque exemple candidat est associé à un cluster et si tous sont dans un cluster où des exemples du jeu d'entraînement sont aussi présents alors ils sont considérés comme redondants. Dans ce cas, les clusters sont scindés en sous-clusters jusqu'à ce que des exemples candidats soient considérés comme non redondants. Stumpf *et al.* (2014), quant à eux, utilisent deux mesures afin d'évaluer la diversité, une basée sur la distance euclidienne entre les exemples candidats et ceux labélisés et une basée sur l'entropie croisée. L'ensemble des publications intégrant un critère de diversité a montré qu'intégrer une telle mesure permettait une amélioration importante des performances dans les premières itérations et réduisait donc le nombre d'échantillons d'entraînement pour obtenir une classification performante.

Afin de réduire le nombre d'exemples à labéliser, Wan *et al.* (2015) proposent de combiner apprentissage semi-supervisé et apprentissage actif. Dans cette publication, un SVM est entraîné à partir d'un nombre d'exemples limité et des pseudo-labels sont attribués aux exemples s'ils ont un haut niveau de confiance. L'apprentissage actif vise ici à vérifier si les pseudo-labels sont justes. Pour cela un nouvel entraînement est réalisé à partir des exemples labélisés et pseudo labélisés, et les exemples interrogés dans la partie active de la méthode sont ceux situés à la marge. Un autre SVM est entraîné uniquement sur les données labélisées et les exemples qui ne changent pas de classe entre ces deux classifications sont ajoutés aux pseudo-labels. Cette méthode a montré de bonnes performances comparées à d'autres méthodes d'apprentissage actif pour la classification d'images hyperspectrales.

La dimension spatiale est importante pour les applications de télédétection et les méthodes d'apprentissage actif sont encore peu nombreuses et pourrait néanmoins tirer avantage à utiliser les informations spatiales (Crawford *et al.*, 2013). Différentes méthodes ont été proposées dans la littérature afin de prendre en compte cette dimension. Xu *et al.* (2014) utilisent des patchs de différentes tailles afin de prendre en compte les pixels voisins dans le processus d'apprentissage actif et ainsi réduire le temps d'exécution. Stumpf *et al.* (2014) ont proposé un apprentissage actif basé régions intégrant un critère de diversité. Leurs résultats ont montré que l'interrogation de régions plutôt que de lots d'exemples permettait de réduire le temps d'exécution ainsi que celui de labélisation. Les images satellites sont caractérisées par une autocorrélation spatiale importante (Wulder et Boots, 1998). Pasolli *et al.* (2014) ont ainsi intégré au critère

d'incertitude, un critère de distance aux exemples utilisés pour l'entraînement qu'il s'agit de maximiser. L'objectif est d'obtenir de nouveaux exemples moins corrélés et donc plus diversifiés. Cette caractéristique des images satellites a aussi été utilisée par Jun *et al.* (2009) afin de proposer une amélioration de la méthode de classification par maximum de vraisemblance qui s'adapte aux variations spatiales. Afin de minimiser le coût dans le cas où les nouveaux exemples seraient labélisés par des relevés de terrain, Liu *et al.* (2009) ainsi que Persello *et al.* (2014) ont proposé plusieurs stratégies permettant de minimiser les distances de déplacements pour la labélisation d'un lot de nouvelles instances. Toutefois, ces études ne prennent pas en compte les spécificités du terrain et une étude plus récente a proposé d'intégrer des informations sur la topographie ainsi que le réseau routier Demir *et al.* (2014).

Ainsi l'objectif de ce chapitre est d'extraire à partir d'images THRS la végétation urbaine, arborée et herbacée, en étendant aux classifications multi-classes la méthode d'apprentissage actif initialement proposée par Stumpf *et al.* (2014). Cette méthode a été développée pour la détection des glissements de terrain et donc pour des problèmes de classifications binaires. Elle se base sur un comité de modèle afin de sélectionner les régions les plus informatives dans le cadre d'une analyse orientée-objet. Dans de nombreuses publications dédiées à l'apprentissage actif multi-classes, toutes les classes recherchées doivent généralement être présentes dans le jeu de données d'entraînement initial. La méthode développée dans cette recherche ne nécessite pas que les classes soient connues *a priori* et elles peuvent ainsi être découvertes au fur et à mesure des itérations. Trois méthodes d'apprentissage actif sont proposées dans ce chapitre afin de sélectionner les nouvelles régions à labéliser.

De plus, la plupart des applications d'apprentissage actif en télédétection comparent ces méthodes à un simple échantillonnage aléatoire. Toutefois dans ce domaine d'autres stratégies de sélection d'échantillons sont très régulièrement utilisées et ont montré qu'elles permettaient d'obtenir de meilleurs résultats. C'est par exemple le cas de l'échantillonnage aléatoire stratifié. Ainsi un autre objectif de ce chapitre est de comparer les performances des méthodes d'apprentissage actif proposées à cette stratégie en stratifiant les sites d'études par des grilles, celles que nous avons présentées lors du chapitre précédent (§ 3.2.4).

## 4.2 Méthodes proposées pour la sélection d'échantillons d'apprentissage

Les différentes méthodes d'apprentissage actif développées dans ce chapitre utilisent une stratégie par comité de modèle pour la sélection de nouvelles instances. Pour cette stratégie, le désaccord du comité est mesuré pour chaque exemple n'étant pas labélisé et ceux qui sont sélectionnés sont ceux où cette mesure est maximum. Le comité de modèle est ici construit avec la méthode RF, qui, grâce à son échantillonnage bootstrap pour la construction de chaque arbre, permet de créer une multitude de modèles sur des

sous-ensembles disjoints (Bondu et Lemaire, 2008). Cette stratégie pour la construction des modèles appelée *Query-by-bagging* a été initialement proposée par Abe et Mamitsuka (1998). Elle a montré de bons résultats avec l'utilisation de forêts aléatoires (Borisov *et al.*, 2011) et a été adaptée récemment à la classification de données de télédétection (Stumpf *et al.*, 2014). Comme pour le chapitre 3, les forêts ont été construites avec 500 arbres et un échantillonnage bootstrap stratifié afin de corriger le déséquilibre entre les classes, ce qui permet, dans un processus d'apprentissage actif, d'améliorer les performances et d'équilibrer précisions pour l'utilisateur et pour le producteur (Stumpf *et al.*, 2014).

Afin de mesurer le désaccord du comité de modèle, l'entropie  $H$  est calculée grâce à l'équation :

$$H(x_i) = - \sum_1^{N_i} p_i \log p_i \quad (4.1)$$

où  $N_i$  représente le nombre de classes prédites par le comité pour un exemple  $x_i$  et  $p_i$  la part du vote des arbres pour une classe. L'entropie ainsi mesurée va être fortement influencée par le nombre de classes  $N_i$  prédit par le comité. En effet, plus ce nombre est important, plus l'entropie sera élevée. Cela peut donc entraîner un échantillonnage déséquilibré qui va favoriser les domaines d'incertitude où le nombre de classes est important et ne permet donc pas d'obtenir des exemples diversifiés. Copa *et al.* (2010) recommandent ainsi d'utiliser l'entropie normalisée  $nH$  selon le nombre de classes présentes dans la prédiction par l'équation :

$$nH(x_i) = \frac{H(x_i)}{\log N_i} \quad (4.2)$$

Les différents algorithmes d'apprentissage actif présentés dans cette étude sélectionnent à chaque itération une ou plusieurs régions. En effet, le coût de labélisation dans un processus d'apprentissage actif n'est pas le même pour chaque exemple et n'est pas linéaire au nombre d'instances sélectionnées à chaque itération (Settles *et al.*, 2008). Stumpf *et al.* (2014) ont ainsi montré que sélectionner des régions plutôt que des exemples, même en lots de différentes tailles, permet d'obtenir de meilleurs résultats plus rapidement en termes de temps de calcul ainsi qu'en temps de labélisation.

Trois stratégies sont proposées dans cette recherche afin de déterminer les régions qui sont interrogées à chaque itération. La première va sélectionner la région la plus incertaine, celle où le comité est le plus en désaccord (§ 4.2.1 ; *Multi-Class Spatial Active Learning* - MCSAL). La deuxième stratégie proposée intègre, en plus du critère d'incertitude, un critère de diversité afin que les nouvelles régions sectionnées soient moins redondantes avec celles déjà sélectionnées et donc qu'elles soient plus informatives (§ 4.2.2 ; *Multi-Class Spatial Active Learning considering Diversity* - MCSAL<sub>D</sub>). Ces deux premières méthodes ont été adaptées de Stumpf *et al.* (2014). De plus, une nouvelle stratégie est introduite dans cette recherche. Elle permet à l'utilisateur d'intégrer des connaissances du terrain afin de guider la sélection de nouvelles régions. En effet dans

un processus d'apprentissage actif, seul l'algorithme décide des nouveaux exemples à labéliser et l'utilisateur n'a en général aucune possibilité d'influer sur la sélection de nouveaux exemples. Ainsi cette stratégie va sélectionner plusieurs régions à chaque itération en fonction d'une carte fournie préalablement permettant de stratifier l'espace d'étude (§ 4.2.3 ; *Multi-Class Spatial Active Learning integrating Stratification – MCSAL<sub>S</sub>*). Afin de vérifier les performances de ces stratégies, elles sont comparées à une méthode de référence très souvent utilisée en télédétection pour la sélection des données d'entraînement qui se base sur un échantillonnage aléatoire stratifié à partir d'un maillage (§ 4.2.4 ; *Stratified Random Grid Sampling - SRGS*).

#### 4.2.1 Apprentissage actif spatial multi-classes (MCSAL)

Pour trouver la région la plus incertaine  $W_X$  de taille  $w$  dans le cadre d'une démarche orientée-objet, les segments sont tout d'abord représentés par leur centre de gravité et leur valeur d'entropie est projetée sur un raster  $G$  dont la résolution spatiale est  $g$ . Pour chacun des pixels, la moyenne de l'entropie des segments leur est ainsi attribuée ce qui permet d'obtenir une carte d'entropie  $M_H$  pour l'ensemble des régions qui ne sont pas utilisées pour la construction du comité de modèle. Dans un second temps, une fenêtre mobile  $W$  est appliquée à  $M_H$  pour calculer l'entropie moyenne locale  $\mu_H$  à chaque position  $(x, y)$  du raster  $G$  grâce à l'équation :

$$\mu_{H(x,y)} = \frac{1}{w^2} \sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} M_H(x+i, y+j) \quad (4.3)$$

où  $i$  et  $j$  indiquent les indices d'une grille régulière. Cette méthode peut aisément être adaptée à une approche pixel sans avoir à projeter les valeurs d'entropie des segments préalablement. Pour cette première stratégie, la région  $W^X$  sélectionnée par l'algorithme à chaque itération pour labélisation est celle maximisant l'entropie moyenne locale  $\mu_H$ . La fonction de requête peut donc être formulée par :

$$W^X = \operatorname{argmax}_{W \in M_H} \mu_H(W) \quad (4.4)$$

Ainsi tous les exemples dont le centre de gravité est contenu dans la région sélectionnée sont interrogés pour labélisation.

#### 4.2.2 Apprentissage actif spatial multi-classes prenant en compte la diversité (MCSAL<sub>D</sub>)

Dans un processus itératif d'apprentissage actif, les nouveaux exemples demandés par la méthode sont les plus incertains pour un comité de modèle ou les plus proches de l'hyperplan pour celles basées sur les marges. Toutefois, les informations obtenues



grâce à ces nouveaux exemples peuvent être fortement redondantes et ne sont donc pas réellement informatives.

Stumpf *et al.* (2014) ont proposé deux mesures permettant de prendre en compte la diversité. La première est basée sur la mesure de l'écart-type de la distance euclidienne dans l'espace des données entre chaque exemple de la région sélectionnée et son voisin le plus proche dans les données d'entraînement. La seconde mesure est basée sur l'entropie croisée entre les exemples de la région et ceux du jeu d'entraînement. Dans cette recherche, seule la mesure basée sur l'entropie croisée est utilisée puisqu'elle permet une diminution plus importante du temps de labélisation, mais augmente néanmoins le temps d'exécution. Dans un premier temps, les  $m$  régions ne se chevauchant pas et maximisant l'entropie sont sélectionnées et la diversité est mesurée entre chaque région candidate et les exemples déjà labélisés. Le calcul de l'entropie croisée  $H^X$  dépend de la distance logarithmique moyenne  $\rho_k(X, c)$  entre les exemples candidats ( $c \in W^m$ ) à leurs  $k$  voisins les plus proches dans le jeu d'entraînement ( $s \in X$ ) et est défini par :

$$H^X(W^m, X) = \log(v_D|X|) - \Psi(k) + \frac{D}{|W^m|} \sum_{c \in W^m} \log \rho_k(X, c) \quad (4.5)$$

où  $v_D$  est le volume de l'hyperboule dans  $R^D$ ,  $D$  est le nombre d'attributs,  $\Psi$  est la fonction digamma et  $|X|$  désigne la cardinalité du jeu d'entraînement (Boltz *et al.*, 2007). Tous les attributs utilisés pour la classification ne sont pas nécessairement utiles pour séparer les différentes classes considérées et il peut être intéressant de ne sélectionner qu'une partie des attributs pour le calcul de l'entropie croisée. La méthode de classification RF présente l'avantage d'avoir une mesure intrinsèque de l'importance des variables calculées grâce aux exemples non sélectionnés lors de l'échantillonnage bootstrap (Breiman, 2001). Ainsi, un seuil  $t_i$  sur la diminution moyenne de précision donnée par RF peut être utilisé pour déterminer les variables utilisées pour calculer  $H^X$ . Pour chaque expérimentation menée dans cette recherche, le nombre de voisins considéré pour le calcul de l'entropie croisée est fixé à  $k = 10$ . Le nombre de voisins  $k$  a peu d'effet sur les performances (Boltz *et al.*, 2007) et 10 est la valeur par défaut de l'entropie croisée implémentée dans le package FNN (Beygelzimer *et al.*, 2013). La fonction de requête pour la sélection de nouveaux exemples est formulée par :

$$W^X = \operatorname{argmax}_{W^m \in M_H} H^X(W^m) \quad (4.6)$$

### 4.2.3 Apprentissage actif spatial multi-classes stratifié (MCSAL<sub>S</sub>)

Pour la classification supervisée d'image de télédétection, plusieurs méthodes d'échantillonnage sont couramment utilisées, telles que l'échantillonnage stratifié ou systématique (Congalton et Green, 2008). L'intégration d'une stratification permet souvent d'obtenir de meilleurs résultats qu'un simple échantillonnage aléatoire. Il permet en effet

d'obtenir des exemples représentatifs de chaque classe considérée ou une représentation de la complexité d'une scène.

Il est donc apparu intéressant de pouvoir intégrer dans le processus d'apprentissage actif des données fournies par l'utilisateur afin de contraindre le choix des nouveaux exemples avec pour objectif d'obtenir des exemples plus diversifiés. Ainsi au lieu de ne sélectionner qu'une seule région à chaque itération comme dans les deux méthodes précédentes, cette troisième stratégie d'apprentissage actif va sélectionner une région par strate de la carte de stratification fournie par l'utilisateur. Pour chaque strate, la région sélectionnée de taille  $w$  est celle où la moyenne locale de l'entropie  $\mu^H$  est maximale.

#### 4.2.4 Échantillonnage aléatoire stratifié (SRGS)

Dans de nombreuses publications, les méthodes d'apprentissage actif sont comparées à un simple échantillonnage aléatoire. Comme il a été mentionné précédemment, en télédétection, un échantillonnage aléatoire stratifié ou même systématique est plus couramment utilisé. Dans cette étude, les trois méthodes d'apprentissage actif sont comparées à un échantillonnage aléatoire stratifié. La carte de stratification utilisée ici est la même que celle utilisée pour MCSAL<sub>S</sub>. Dans un premier temps, une grille *grd* est créée à partir de la carte de stratification. Cette grille est réalisée en découpant chacune des zones issues de la carte de stratification en sous-maille de taille  $w$ , équivalente à celle utilisée pour les classifications par apprentissage actif. Comme pour MCSAL<sub>S</sub> une sous-maille par strate est intégrée à chaque itération, mais ici de manière aléatoire.

### 4.3 Préparation des données

Les données, les zones tests et les vérités-terrain utilisées afin de tester les méthodes développées dans ce chapitre sont les mêmes que celles définies dans la section 3.2. La différence est que toutes les zones qui ont été photo-interprétées ont la possibilité d'être utilisées pour l'entraînement et que les classes ont été attribuées selon la majorité, quelle que soit la zone considérée. De plus, des données supplémentaires ont pu être acquises pour le site Strasbourg qui n'étaient pas encore disponibles lors de la réalisation du chapitre précédent. Celles-ci concernent la hauteur par le biais d'un Modèle Numérique de Surface (MNS) à une résolution spatiale de 1 m ainsi qu'un Modèle Numérique de Terrain (MNT) avec une résolution de 2 m fournis dans le cadre du partenariat CIGAL. Tous deux ont été obtenus par photogrammétrie à partir de photographies aériennes acquises entre le 27 juin et le 30 septembre 2011. Un Modèle Numérique d'Élévation (MNE) a été dérivé par soustraction du MNS par le MNT.

La même échelle de segmentation a été utilisée pour les deux sites d'études. Le chapitre 3 a montré qu'une sur-segmentation de l'image permettait d'obtenir de meilleurs résultats dans le cadre d'une classification avec RF et que, lorsque l'algorithme Multi-Resolution

était utilisé, les performances sont relativement stables, quel que soit le paramètre d'échelle utilisé. Ainsi, une segmentation avec un paramètre d'échelle de 30 a été appliquée sur nos deux sites d'étude, en utilisant uniquement l'image d'août pour celui de Strasbourg. À ce premier niveau de segmentation, un deuxième, plus large, a été réalisé. L'utilisation d'information provenant de segmentations multi-échelles a montré que les performances des classifications pouvaient être fortement améliorées (Smith, 2010). Les paramètres de ce deuxième niveau de segmentation ont été déterminés par une procédure d'essais et erreurs avec l'objectif de mieux délimiter les parcelles agricoles. En effet, les parcelles agricoles apparaissent assez difficiles à différencier de la végétation herbacée (§ 3.3.3) et des attributs calculés à partir d'une segmentation assez large peuvent être intéressants pour mieux les identifier (Duro *et al.*, 2012b). Cette deuxième échelle de segmentation est réalisée avec Multi-Resolution et un paramètre d'échelle de 75 combinée à SDA (*Spectral Difference Algorithm*) avec un maximum de différence spectrale de 30. SDA est une segmentation qui fusionne les objets voisins selon la moyenne de leurs valeurs spectrales si celles-ci sont en dessous du seuil maximum fixé par l'utilisateur.

Pour chacune des images Pléiades et pour le premier niveau de segmentation, les attributs calculés avec eCognition sont les mêmes que ceux utilisés précédemment (tableau 3.1). Ceux calculés à partir des informations de hauteur, seulement pour Strasbourg, et ceux du deuxième niveau de segmentation sont résumés dans le tableau 4.1. Ceux du deuxième niveau ont été attribués par projection aux segments correspondants les plus fins.

**Tableau 4.1** – Vue d'ensemble des attributs calculés à partir des informations de hauteur et du deuxième niveau de segmentation

Type	Attributs	Bande spectrale ou indice	
Attributs de hauteur (1 <sup>er</sup> niveau de segmentation)			
Spectral	Moyenne des bandes spectrales	MNS, MNT, MNE	
	Valeur minimale des pixels	MNS, MNT, MNE	
	Valeur maximale des pixels	MNS, MNT, MNE	
	Différence moyenne aux voisins plus clairs	MNS, MNT, MNE	
	Différence moyenne aux voisins plus foncés	MNS, MNT, MNE	
Attributs du 2 <sup>e</sup> niveau de segmentation			
Géométrie	Aire, périmètre, indice de bordure, compacité, densité, forme rectangulaire, indice de forme		
Texture	Matrice de cooccurrence en niveau de gris	Homogénéité,	moyenne, écart-type

## 4.4 Protocole de tests

Pour évaluer les résultats ou pour la sélection de modèles d'apprentissage, Kohavi (1995) recommande d'utiliser une validation croisée avec un nombre limité de plis (entre et 10 et 20) et de préférence stratifiés. Les expérimentations sont réalisées pour les deux

sites d'étude à partir d'une validation croisée stratifiée de 16 plis construits à partir des indices présentés précédemment (§ 3.2.4) afin d'évaluer la précision globale moyenne et l'écart-type des différents algorithmes d'apprentissage actifs sur l'ensemble des données. Chacune des mailles photo-interprétées qui constituent notre vérité terrain a été divisée en quatre afin d'obtenir 16 sous-mailles par strate d'urbanisation. Ainsi une sous-maille par strate est utilisée pour constituer les données de test.

Afin de produire les données d'initialisation pour chaque pli de la validation croisée, un segment par strate d'urbanisation est sélectionné aléatoirement et une fenêtre de taille  $w$  centrée sur ce point est créée permettant de considérer tous les segments dont le centre de gravité  $y$  est compris. Pour chaque test et pour chaque méthode, les mêmes données d'initialisation sont considérées.

Les expérimentations sont toutes effectuées à partir de forêts composées de 500 arbres ainsi qu'un échantillonnage bootstrap stratifié afin de corriger le déséquilibre entre les classes. De plus, la taille des pixels  $g$  du raster créé pour représenter l'entropie a été fixée à 5 m. Cette résolution est apparue *a priori* un bon compromis entre la précision de localisation des segments et le temps de calcul, et permet de pouvoir considérer de petites tailles de fenêtres glissantes pour le calcul de l'entropie locale. Plusieurs tests sont réalisés pour vérifier l'influence des autres paramètres pour les trois méthodes d'apprentissage actif.

Le premier test réalisé pour les trois méthodes actives porte sur l'influence que peut avoir la normalisation de l'entropie sur les performances des classifications. Pour les trois méthodes, la taille des fenêtres reste constante ( $w = 15$  m) et les paramètres de  $m$  et  $t_i$  pour MCSAL<sub>D</sub> sont fixés à 8 et 0,01 pour les deux jeux de données.

D'autres tests sont effectués afin de vérifier l'influence des paramètres pour la méthode d'apprentissage actif MCSAL<sub>D</sub>. Pour chacun de ces tests, une fenêtre de 15 m est utilisée. Le premier concerne le nombre de régions candidates sélectionnées. Trois valeurs de  $m$  sont testées 3,5 et 8 en conservant  $t_i$  à 0,01. La seconde série d'expérimentations est dédiée à vérifier l'influence du paramètre qui contrôle l'importance minimum des variables dans le calcul de l'entropie croisée pour la sélection finale de la région ajoutée au jeu d'entraînement. Ainsi, quatre paramètres de  $t_i$  sont expérimentés : 0, 0,01, 0,03, 0,05 ; en laissant constant  $m$  à 8.

Pour les deux jeux de données et les trois algorithmes d'apprentissage actif, une autre série d'expérimentations concerne la taille des fenêtres de sélection des nouveaux exemples du jeu d'entraînement. Trois tailles sont testées,  $w = 15, 30, 50$  m en laissant inchangé les paramètres  $m$  et  $t_i$  de MCSAL<sub>D</sub> à 8 et 0,01. Pour ces trois tailles de fenêtres, une comparaison avec SRGS est effectuée afin de vérifier l'intérêt que peuvent représenter les méthodes d'apprentissage actif pour l'extraction des éléments arborés et herbacés en milieu urbain. Pour les différents tests présentés précédemment, l'ensemble des attributs des deux zones d'études est utilisé. Un autre test de cette recherche consiste donc à vérifier l'influence que peut avoir la combinaison de différentes sources de données sur

la performance de l'extraction de la végétation. Ces tests sont uniquement réalisés pour le site d'étude de Strasbourg qui est le seul site combinant de nombreuses données.

Enfin, le dernier test consiste à évaluer les performances obtenues sans prendre en compte les plis de la validation croisée afin de mieux discerner les causes des erreurs de classification et en évaluant spatialement ces confusions. Un premier essai est toutefois réalisé en les prenant en compte afin de vérifier la stabilité des mesures de précision entre les deux procédures d'évaluation.

## **4.5 Évaluation des méthodes d'apprentissage actif proposées**

Afin d'évaluer les performances des différentes méthodes proposées dans ce travail, deux mesures de précision sont considérées. La première est la précision globale. Cette mesure est celle qui est la plus utilisée dans les publications concernant l'apprentissage actif. Cette mesure présente l'avantage d'être synthétique, mais peut être fortement biaisée lorsqu'une classe est très fortement majoritaire comme c'est le cas ici pour nos deux sites d'études où la classe autre est très fortement dominante. Ainsi la F-mesure est aussi utilisée afin d'évaluer les performances des classifications pour nos deux classes de végétation. Pour l'ensemble des tests présentés, ces deux mesures sont calculées en termes de segments. Seule la section 4.5.5 compare les performances obtenues à partir du nombre de segments à ceux obtenus à partir de leur aire.

### **4.5.1 Effet de la normalisation de l'entropie**

Le premier test réalisé porte sur l'effet de la normalisation de l'entropie sur les performances obtenues pour les trois méthodes d'apprentissage actif développées dans ce chapitre. La figure 4.1 présente la moyenne des précisions globales obtenues ainsi que leurs écarts-types à partir de la validation croisée. Normaliser l'entropie pour ces deux zones d'étude ne permet pas nécessairement d'améliorer les résultats obtenus dans les premières itérations en intégrant plus de diversité dans le choix des nouvelles fenêtres. Seul MCSAL<sub>D</sub> pour le site de Strasbourg atteint plus rapidement un plateau quand l'entropie est normalisée. À l'inverse, MCSAL avec le jeu de données de Rennes voit lui ses performances améliorées dans les premières itérations quand l'entropie n'est pas normalisée. Pour les quatre autres cas présentés, les deux méthodes de calcul de l'entropie semblent peu influencer sur les résultats. L'analyse pour les deux classes de végétation, non figurée ici, présente des résultats similaires à ceux qui sont représentés par la figure 4.1.

La normalisation de l'entropie modifie finalement peu la sélection finale des fenêtres. En effet, seules deux ou trois classes peuvent être présentes dans la prédiction, ce qui influence donc assez peu l'ordre des valeurs de l'entropie lorsqu'elle est normalisée. De plus, ces valeurs sont ensuite par deux fois moyennées, une première fois lors de la

projection des valeurs attribuées aux segments sur le raster puis par le passage d'une fenêtre glissante, lissant ainsi les valeurs obtenues. Il pourrait ainsi être intéressant d'utiliser ces méthodes avec plus de classes afin d'évaluer plus finement l'effet de la normalisation de l'entropie sur les résultats obtenus. Pour la suite des tests, l'entropie normalisée est toujours utilisée pour l'ensemble des algorithmes d'apprentissage actif et des sites d'étude.

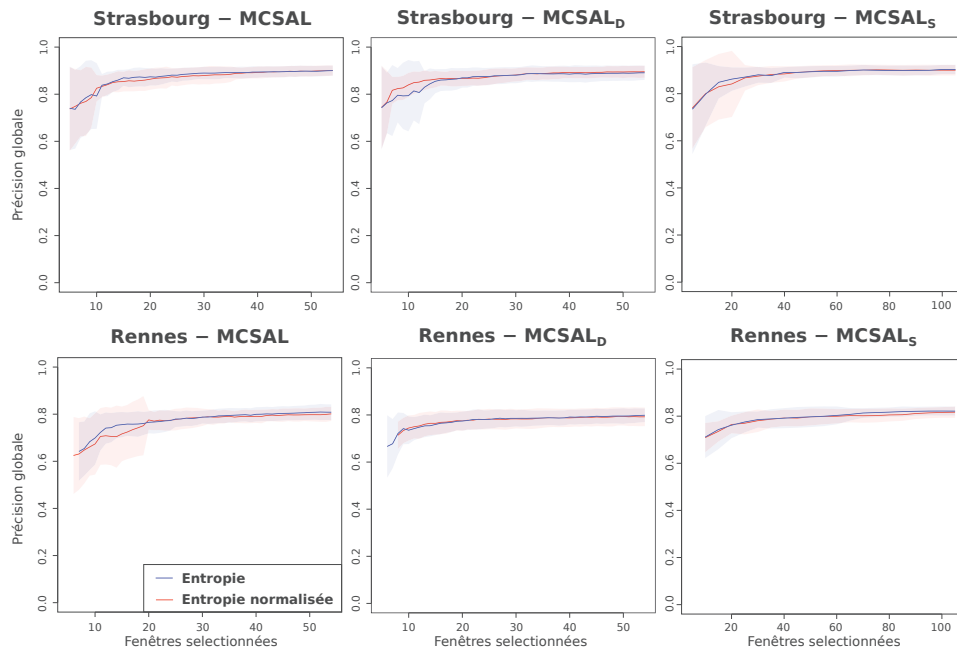
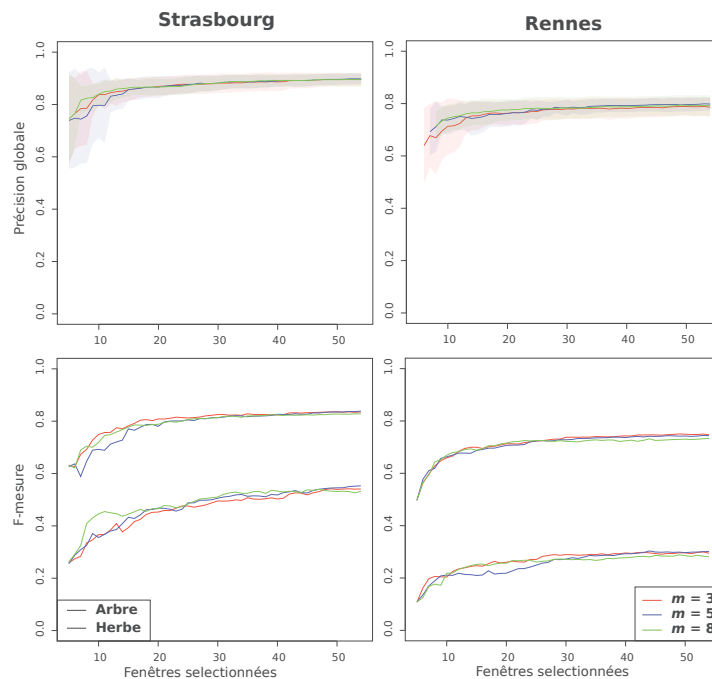


Figure 4.1 – Effet de la normalisation de l'entropie sur les trois stratégies d'apprentissage actif

#### 4.5.2 Sensibilité des paramètres du critère de diversité

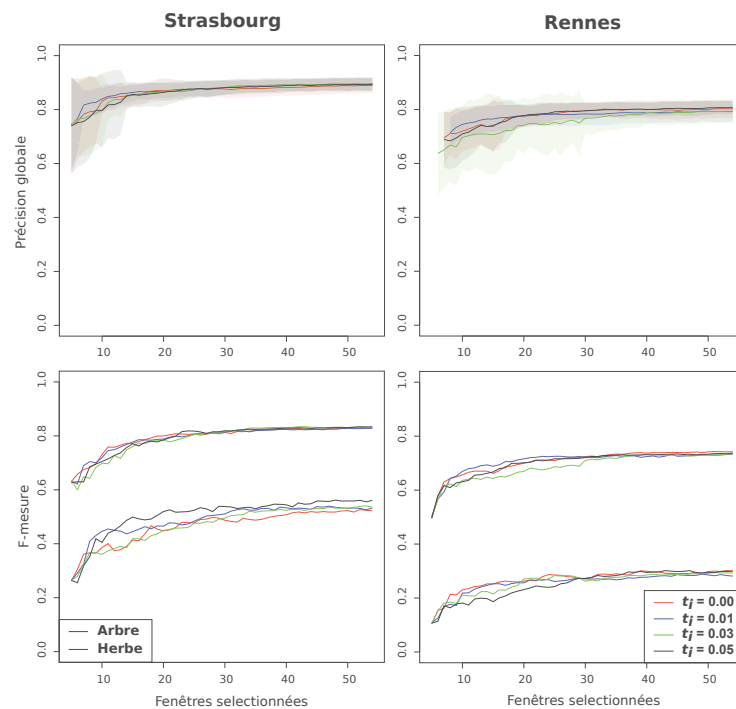
Dans cette partie, les deux paramètres qui composent le critère de diversité intégrée à  $MCSAL_D$ ,  $m$  et  $t_i$ , sont testés. Pour le site de Strasbourg, les résultats apparaissent meilleurs lorsque le nombre de régions candidates est important ( $m = 8$ ) dans les premières itérations (figure 4.2). Dans les itérations suivantes, les précisions globales, mais aussi par classe, convergent assez rapidement, quel que soit le paramètre de  $m$  choisi. Une importante amélioration de la F-mesure pour la végétation herbacée pour Strasbourg est obtenue quand  $m$  est grand. Cet effet est moins important pour la végétation arborée. Les résultats sur le jeu de données de Rennes sont similaires à ceux obtenus sur Strasbourg même si les différences apparaissent moins nettes sur les F-mesures obtenues par classe. Prendre en considération plus de régions dans le processus de sélection permet donc d'obtenir plus de diversité dans les exemples labélisés et ainsi d'améliorer les résultats dans les premières itérations.

L'effet du paramètre contrôlant les variables prises en compte dans le calcul de l'entropie croisée  $t_i$  n'apparaît pas très clair (figure 4.3). En effet, les résultats sont très différents selon les deux sites d'études. Pour Strasbourg le meilleur résultat global est obtenu avec  $t_i$  égal à 0,01 dans les premières itérations et pour la végétation herbacée avec une valeur de  $t_i$  à 0,05. Pour Rennes, lorsque ce paramètre est égal à 0,01, les premières itérations présentent de meilleurs résultats, mais seront ensuite moins bons qu'avec d'autres valeurs après plus d'itérations.



**Figure 4.2** – Performances globales et par classes de  $MCSAL_D$  selon le nombre de fenêtres candidates  $m$  pour Strasbourg et Rennes

Bien que fixer  $m$  à 8 permet d'obtenir de meilleurs résultats dans les premières itérations pour les deux sites, le nombre limité de valeurs de  $m$  testée ne permet pas d'extrapoler cette tendance pour un nombre croissant de régions. En effet avec un nombre plus important de régions candidates, le critère de diversité pourrait prendre une place trop majoritaire vis-à-vis de celui d'incertitude et donc empêcher la sélection de nouveaux échantillons qui seraient informatifs. De nouvelles recherches pourraient être menées afin de déterminer le meilleur compromis entre incertitude et diversité pour la sélection de nouveaux échantillons. Considérant les expériences menées dans cette section, les valeurs de  $m$  et de  $t_i$  sont fixés à 8 et 0,01 pour la suite des tests réalisés.



**Figure 4.3** – Performances globales et par classes de  $MCSAL_D$  selon le seuil minimum d'importance des variables  $t_i$  considéré dans le calcul de l'entropie croisée pour Strasbourg et Rennes

### 4.5.3 Choix de la taille des régions et comparaison des méthodes d'apprentissage actif avec l'aléatoire stratifié

La figure 4.4 présente les résultats qui comparent les différentes stratégies d'apprentissage actif et la méthode de référence basée sur un échantillonnage aléatoire stratifié en nombre de régions interrogées. Trois tailles de fenêtres sont testées (15, 30, 50 m). Les précisions globales obtenues pour les deux sites apparaissent bien meilleures avec les méthodes d'apprentissage actif quand une petite taille de fenêtre est utilisée ( $w = 15$  m). Pour des tailles de régions plus grandes, l'apport des méthodes proposées est plus limité même si les précisions globales restent légèrement plus élevées.

Les performances obtenues par classes de végétation et évaluées grâce à la F-mesure sont par contre beaucoup plus contrastées. En effet si les performances obtenues pour la végétation arborée sont assez similaires à ce qui a été constaté précédemment avec les précisions globales, celles obtenues pour la végétation herbacée sont très différentes. Seules les méthodes  $MCSAL$  et  $MCSAL_S$  pour Strasbourg avec un paramètre  $w$  égal à 15 m permettent d'obtenir de meilleurs résultats comparés à SRGS. Pour les autres tailles de fenêtres de ce jeu de données ou pour toutes celles du jeu de données de Rennes, les F-mesures obtenues sont soit équivalentes entre les différentes méthodes, soit meilleures, quand un échantillonnage aléatoire stratifié est utilisé. Plusieurs causes peuvent être évoquées.



La première est que les différentes méthodes d'apprentissage actif ont toutes tendance à équilibrer le nombre d'exemples sélectionnés pour chaque classe et ne rendent plus compte du déséquilibre existant pour les deux sites d'étude. Ainsi la végétation herbacée se trouve fortement surreprésentée par rapport à la réalité et les résultats de la classification de cette classe peuvent être impactés. De plus, en utilisant une taille de raster  $g$  de 5 m, certains pixels ne contiennent aucun centre de gravité appartenant à un segment. C'est, pour les deux zones d'étude, uniquement le cas dans les zones agricoles où les segments sont de plus grandes tailles que dans les zones urbanisées. Ainsi quand la fenêtre glissante est utilisée pour déterminer les régions les plus incertaines, les valeurs d'entropie attribuées sont fortement diminuées quand un pixel vide est présent puisque la valeur de l'entropie est considérée comme égale à 0. Ces valeurs sont donc souvent minimisées dans les espaces agricoles ce qui entraîne une sous-représentation de ces zones avec les méthodes d'apprentissage actif, au contraire de SRGS, alors qu'elles peuvent être difficiles à différencier de la végétation herbacée. Plus la fenêtre est de grande taille, plus le nombre de pixels impactés par cet effet est important. D'autres tests devraient être réalisés afin de vérifier si cet effet pourrait être annulé en ne considérant pas ces pixels dans le calcul de l'entropie locale afin de voir si cela n'impacterait pas les performances des classifications. Une autre solution serait de ne plus considérer une représentation matricielle en projetant le centre de gravité des segments pour la création de cette carte d'entropie, mais de passer à une représentation vectorielle permettant de s'affranchir de cette étape de projection des segments.

L'utilisation de fenêtres plus grandes permet d'atteindre assez rapidement un plateau pour la végétation arborée. Par contre pour la végétation herbacée, celui-ci est atteint plus vite avec une fenêtre de taille réduite. Les résultats obtenus après la labélisation d'un nombre conséquent de fenêtres (supérieur à 40) sont assez proches, quelle que soit la taille de fenêtre sélectionnée. En effet pour un même nombre de fenêtres labélisées, seule la végétation herbacée présente de meilleures performances (1 à 2 % pour Strasbourg et 1 à 5 % pour Rennes) lorsque  $w$  est fixé à 30 ou 50 m et sont équivalente entre ces deux tailles de fenêtres pour la végétation arborée. Il apparaît donc plus intéressant de choisir dans ce cas une taille de fenêtre de 15 ou de 30 m puisque le temps de labélisation serait alors moins important qu'avec des régions de 50 m.

La comparaison des temps d'exécution pour les trois stratégies d'apprentissage actif donne avantage à MCSAL<sub>S</sub> (figure 4.5). Pour les deux sites d'étude, un temps important à chaque itération de ces trois méthodes est consacré à la création de la carte d'entropie. MCSAL<sub>S</sub> sélectionne à chaque itération cinq régions correspondant au nombre de strates de la stratification utilisée. Le temps d'exécution est donc fortement diminué. Bien que MCSAL<sub>S</sub> interroge plusieurs régions à chaque itération les performances en termes de nombre de régions sélectionnées ne sont pas très différentes qu'avec les deux autres méthodes d'apprentissage actif (figure 4.4). Il apparaît intéressant d'intégrer cette stratification dans le processus d'apprentissage actif puisqu'elle permet de diminuer le temps d'exécution tout en n'augmentant pas le nombre de régions interrogées et donc le temps de labélisation nécessaire pour la classification.

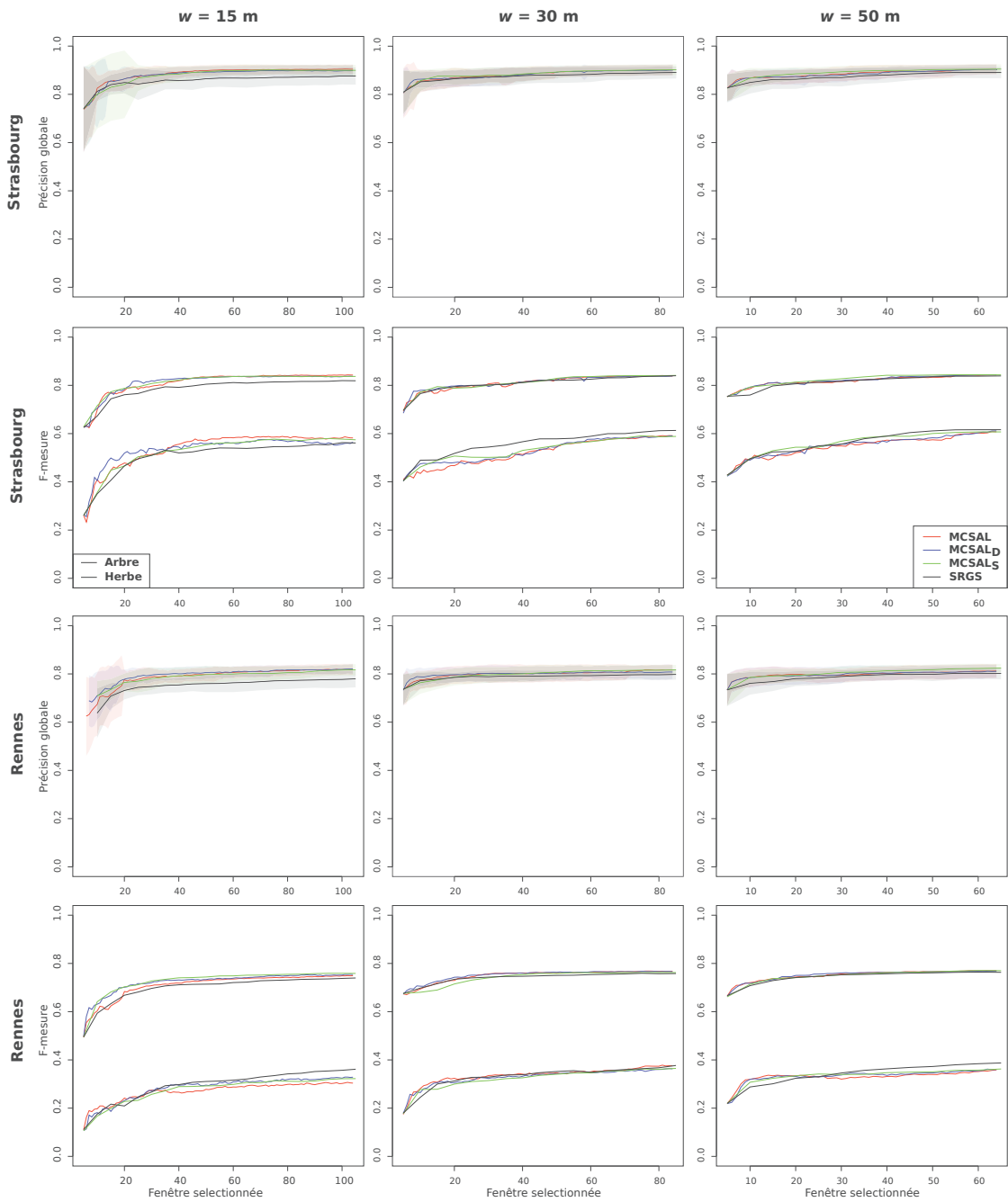


Figure 4.4 – Comparaison des performances globales et par classes des méthodes d'apprentissage actif avec SRGS pour Strasbourg et Rennes selon la taille des régions en nombre de fenêtres sélectionnées

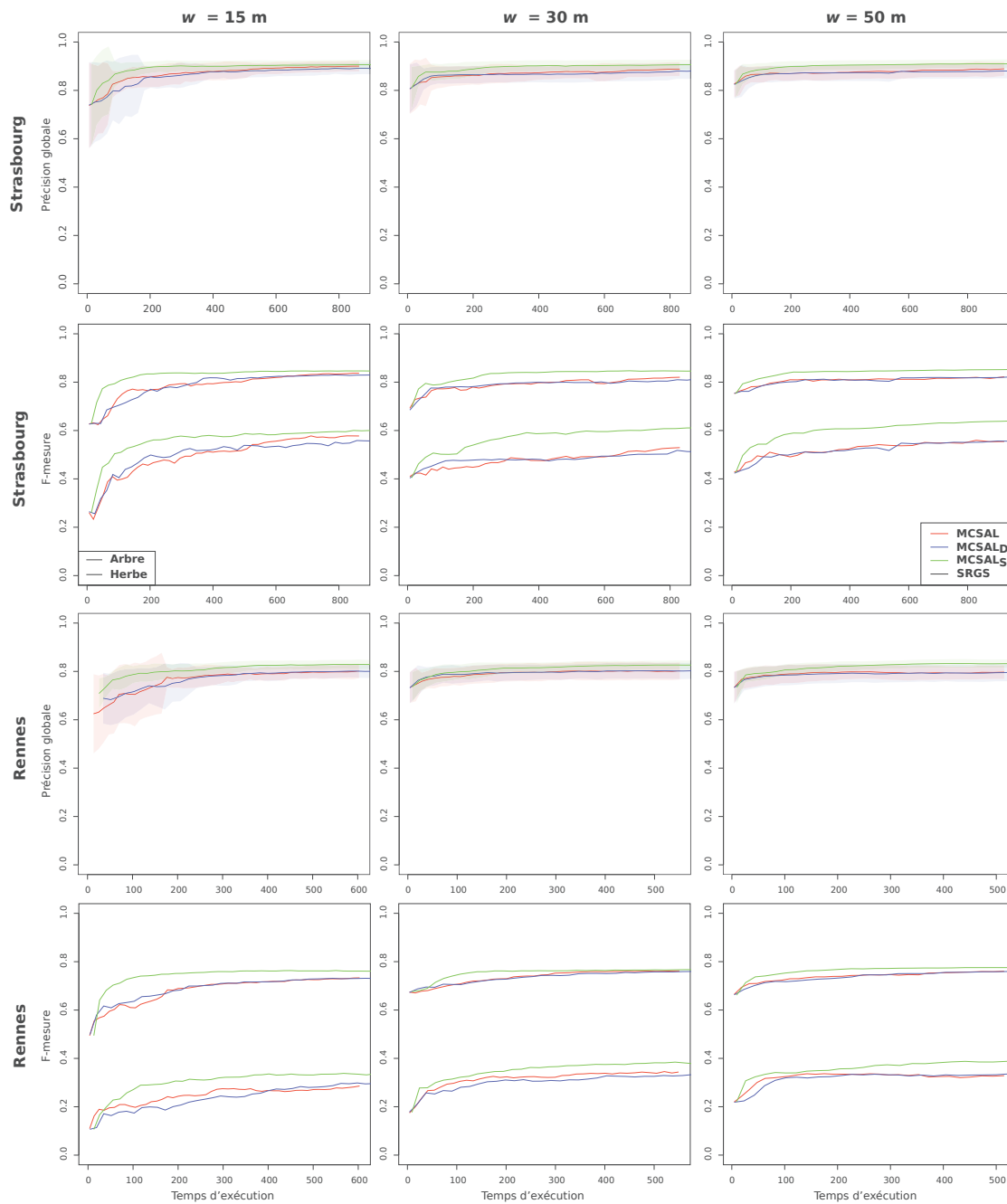


Figure 4.5 – Comparaison des performances globales et par classes des méthodes d'apprentissage actif avec SRGS pour Strasbourg et Rennes selon la taille des régions en temps d'exécution

En terme de précisions globales, les stratégies d'apprentissage actif proposées dans cette thèse montrent l'intérêt de les utiliser comparés à un traditionnel échantillonnage aléatoire stratifié. Toutefois, les résultats par classes sont plus contrastés, notamment pour la végétation herbacée, et montrent les limites de ces méthodes pour la classification de classes fortement minoritaires. Ainsi, l'utilisation d'un échantillonnage aléatoire stratifié s'avère compétitive comparée à des méthodes d'apprentissage actif. Pour ce type d'applications, il apparaît donc nécessaire d'évaluer un algorithme d'apprentissage actif par un échantillonnage stratifié plutôt que totalement aléatoire et de préférence par classe plutôt que globalement comme il est d'usage pour ce type de recherche.

#### 4.5.4 Apport des données auxiliaires pour l'amélioration des performances

Dans cette partie, plusieurs ensembles d'attributs sont créés afin de vérifier l'influence des différents types de données sur les performances des classifications par apprentissage actif pour le site d'étude de Strasbourg. Le tableau 4.2 résume les cinq ensembles d'attributs utilisés.

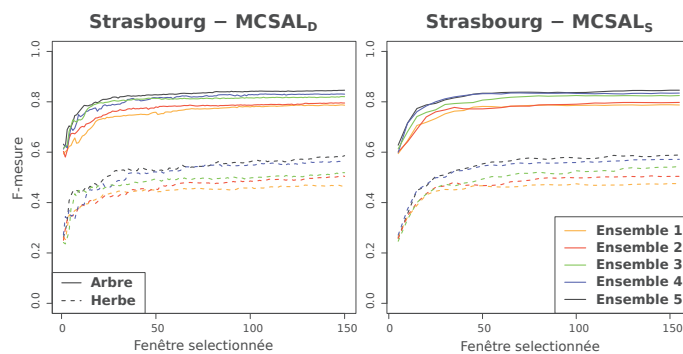
**Tableau 4.2** – Vue d'ensemble des différents jeux de données testés pour le site de Strasbourg

Jeu de données	Attributs	Nombre d'attributs
Ensemble 1	1 <sup>er</sup> niveau de segmentation	51
Ensemble 2	1 <sup>er</sup> + 2 <sup>e</sup> niveaux de segmentation	61
Ensemble 3	1 <sup>er</sup> + 2 <sup>e</sup> niveaux de segmentation + multi-temporels	145
Ensemble 4	1 <sup>er</sup> + 2 <sup>e</sup> niveaux de segmentation + hauteur	76
Ensemble 5	1 <sup>er</sup> + 2 <sup>e</sup> niveaux de segmentation + multi-temporels + hauteur	160

Les meilleurs résultats sont logiquement obtenus quand l'ensemble des attributs est utilisé, que ce soit avec  $MCSAL_D$  ou  $MCSAL_S$  (figure 4.6). Néanmoins lorsqu'une seule image est utilisée et que des informations sur la hauteur des objets ne sont pas disponibles, utiliser un second niveau de segmentation (ensemble 2) pour calculer des attributs permet d'améliorer les performances principalement pour la végétation herbacée comparé à l'utilisation de l'ensemble 1 (figure 4.6). L'ensemble 2 permet de mieux séparer les parcelles agricoles, non considéré comme de la végétation herbacée, et les surfaces en herbe. Les attributs multi-temporels (ensemble 3) vont aussi dans ce sens en permettant de mieux identifier les parcelles ayant changé d'occupation du sol. De plus, l'utilisation de l'ensemble 3 permet aussi d'améliorer de manière importante la classification des éléments arborés en les distinguant mieux de ceux herbacés. Prendre en compte les attributs de hauteur associée à ceux calculés à partir de l'image d'août (ensemble 4) améliore fortement les résultats autant pour la végétation arborée qu'herbacée. La différence entre l'utilisation de l'ensemble des variables et l'ensemble 4 est faible. En effet pour les deux stratégies d'apprentissage actif, l'amélioration apportée par l'ensemble 5 est d'environ 0,8 à 1,3 % pour la végétation arborée et de 2,5 à 3,8 % pour celle herbacée. La complexité de la classification est de plus fortement réduite avec

l'utilisation de l'ensemble 4. Cela permet donc de réduire fortement le temps de calcul ainsi que la consommation de mémoire, rendant l'application pour de plus grandes zones d'étude plus facilement réalisable.

La combinaison de l'image Pléiades principale et des attributs de hauteurs, sans l'utilisation des informations multi-temporelles, apparaît donc être un bon compromis. Cette troisième dimension peut être une donnée parfois difficile à obtenir, mais le fait que le système de satellites Pléiades permet des acquisitions stéréoscopiques apparaît être une solution intéressante pour obtenir informations spectrales et de hauteur par photogrammétrie avec une seule donnée.



**Figure 4.6** – Précisions des classifications pour MCSAL<sub>D</sub> et MCSAL<sub>S</sub> par classes de végétation selon le jeu de données utilisé

#### 4.5.5 Évaluation des erreurs selon l'aire et la pureté de la classe

Afin d'évaluer plus finement les causes d'erreurs des classifications, cette dernière section ne va pas considérer les plis de la validation croisée. Néanmoins pour vérifier la pertinence des mesures de performance sans ces plis une première évaluation est réalisée globalement en les considérant.

Pour les sites d'études de Rennes et de Strasbourg, une classification est réalisée avec MCSAL<sub>S</sub> et 40 itérations, soit 200 fenêtres avec une taille de 15 m, en considérant l'entropie normalisée. Les performances sont mesurées grâce à la F-mesure pour chaque classe tant en terme de segments qu'en terme d'aire (table 4.3). Pour chacune des classes et chacun des sites d'études, les résultats en segments apparaissent moins performants que ceux en aire. Les différences sont assez minimes pour la classe autre, mais peuvent être assez importantes pour les deux autres. En effet l'amélioration de la F-mesure pour la végétation herbacée de Strasbourg est de 0,07. Cette classe est celle qui apparaît la plus difficile à obtenir notamment sur le site de Rennes qui ne dispose que d'une seule source de données. De plus, la variabilité de la F-mesure selon le pli de la validation croisée est très importante.

La sélection d'exemples par la méthode proposée permet de corriger le déséquilibre initial entre les classes. Le nombre d'exemples final de chaque classe est similaire pour le site de Strasbourg et montre des différences plus importantes pour celui de Rennes, mais qui ne sont toutefois pas le reflet de la distribution initiale.

**Tableau 4.3** – Moyenne des F-mesures en aire et en segments et du nombre d'exemples sélectionnés par classes obtenues avec MCSAL<sub>S</sub> ( $w = 15$  m,  $n = 40$ ) par validation croisée en 16 plis (entre parenthèses les écarts-type de chaque valeur)

	Strasbourg			Rennes		
	Segments	Aire	Nombre d'exemples	Segments	Aire	Nombre d'exemples
Arbre	0,85 (0,05)	0,88 (0,04)	979 (52)	0,76 (0,06)	0,79 (0,06)	1209 (100)
Herbe	0,60 (0,09)	0,67 (0,11)	954 (34)	0,34 (0,08)	0,36 (0,11)	691 (25)
Autre	0,95 (0,01)	0,95 (0,02)	1025 (79)	0,90 (0,02)	0,91 (0,03)	986 (116)

L'analyse finale de la distribution des erreurs sur les deux sites d'études est réalisée par une classification réalisée à partir des données d'entraînement sélectionnées par MCSAL<sub>S</sub> avec 40 itérations une taille de fenêtre de 15 m sur l'ensemble des données labélisées. Les performances de ces deux classifications sont évaluées à partir de l'ensemble des exemples n'étant pas utilisés pour l'entraînement.

Les F-mesures pour l'ensemble de ces données sont meilleures pour les deux sites d'études que ceux obtenus dans les tests précédents réalisés avec la validation croisée (table 4.4). Les fenêtres sélectionnées par MCSAL<sub>S</sub> pour l'entraînement sont assez bien réparties sur l'ensemble des zones labélisées, et donc à proximité des exemples de validation, ce qui peut améliorer les résultats obtenus dus à l'autocorrélation spatiale existante dans une image. Les résultats obtenus ici sont donc positivement biaisés et montrent qu'évaluer une classification sans l'utilisation d'une validation croisée avec de larges zones indépendantes ne peut pas constituer une évaluation pertinente des performances. Ce test permet néanmoins de mieux comprendre quantitativement et spatialement les erreurs commises par la méthode proposée dans ce travail.

**Tableau 4.4** – F-mesures en aire et en segments et nombre d'exemples sélectionnés par classes obtenues avec MCSAL<sub>S</sub> ( $w = 15$  m,  $n = 40$ )

	Strasbourg			Rennes		
	Segments	Aire	Nombre d'exemples	Segments	Aire	Nombre d'exemples
Arbre	0,86	0,90	979	0,80	0,82	800
Herbe	0,65	0,73	936	0,41	0,47	588
Autre	0,96	0,96	1047	0,91	0,91	1575

Les performances des classifications pour les deux villes sont aussi évaluées selon les strates d'urbanisation définies dans la section 3.2.4 qui s'étendent de 1 pour les zones les plus végétalisées à 5 pour celles les plus artificialisées (figure 4.7). Elles montrent des résultats intéressants. Celles obtenues pour la végétation arborée ont tendance à diminuer lorsque le degré d'artificialisation augmente. Dans les espaces plus artificialisés, les surfaces arborées sont souvent de plus petites dimensions et de

nombreux arbres isolés ont été mal classés (figure 4.8.4 et 4.8.5) tandis que dans les zones les plus végétalisées de grandes surfaces d'arbre sont présentes (figure 4.8.1). Elles apparaissent donc plus faciles à classer pour la méthode développée dans cette recherche.

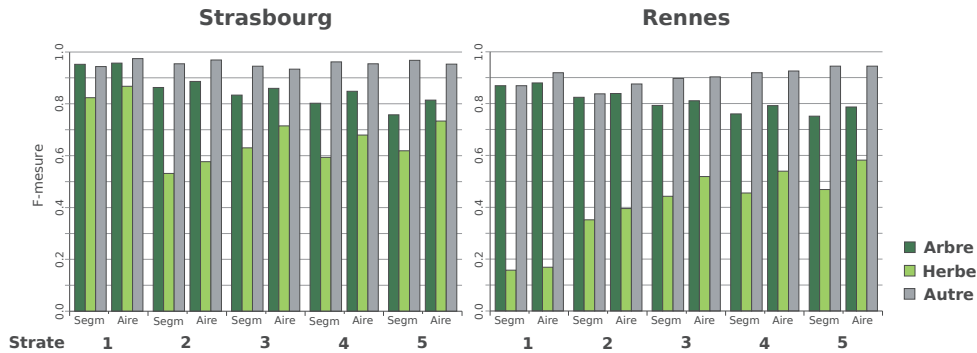


Figure 4.7 – F-mesures en aire et en segments par classes et par strate d'urbanisation obtenues avec MCSAL<sub>S</sub> ( $w = 15$  m,  $n = 40$ )

Pour la végétation herbacée les résultats sont assez stables quel que soit l'espace d'étude et ont tendance à augmenter légèrement avec le degré d'artificialisation exceptée pour la strate 1. Les performances obtenues pour cette strate d'urbanisation sont très différentes selon l'espace d'étude. En effet pour Strasbourg les résultats sont très bons tandis que le contraire apparaît pour Rennes. Pour le site de Rennes, ces zones sont principalement composées de parcelles agricoles ainsi que de surfaces arborées assez larges, telles que des forêts ou des haies. Très peu de surfaces herbacées sont présentes et, dès lors qu'une parcelle agricole est classée en herbe, les F-mesures obtenues sont très impactées. Sur le site de Strasbourg au contraire, de grandes surfaces herbacées sont présentes permettant ainsi d'améliorer fortement les résultats obtenus pour cette strate (figure 4.8.1). Comme pour les résultats globaux (tableau 4.4), les résultats pour les deux classes de végétation et pour tous les degrés d'urbanisation apparaissent meilleurs quand l'aire est considérée plutôt que les segments (figure 4.7).

La figure 4.9 décrit les résultats obtenus selon l'aire des segments. Quel que soit la classe de végétation ou le site d'étude, cette figure montre que les classifications ont tendance à être plus performantes quand les segments sont de plus grande taille. Les résultats calculés à partir de l'aire sont donc meilleurs que ceux obtenus à partir du nombre de segments.

Les résultats sont aussi évalués selon la « pureté » des segments pour la végétation arborée et la végétation herbacée (figure 4.9). La pureté correspond à la part que représente la classe qui a été attribuée au segment par rapport à la surface totale de ce segment. Dans cette étude, la vérité-terrain a été créée par photo-interprétation des images et non pas par attribution des classes directement aux segments. Elle a ensuite été projetée sur les segments et la classe qui a été assignée à chacun est celle qui était majoritaire. Comme dans cette étude trois classes sont considérées, la part de la classe attribuée est donc au

minimum supérieur à un tiers. La figure 4.9 montre que, de manière générale, plus le segment est pur, plus il a de chance d'être bien classé. Ces exemples sont évidemment placés à la limite des objets et ce sont ces limites qui sont les plus difficiles à placer par l'utilisateur lors de la numérisation.

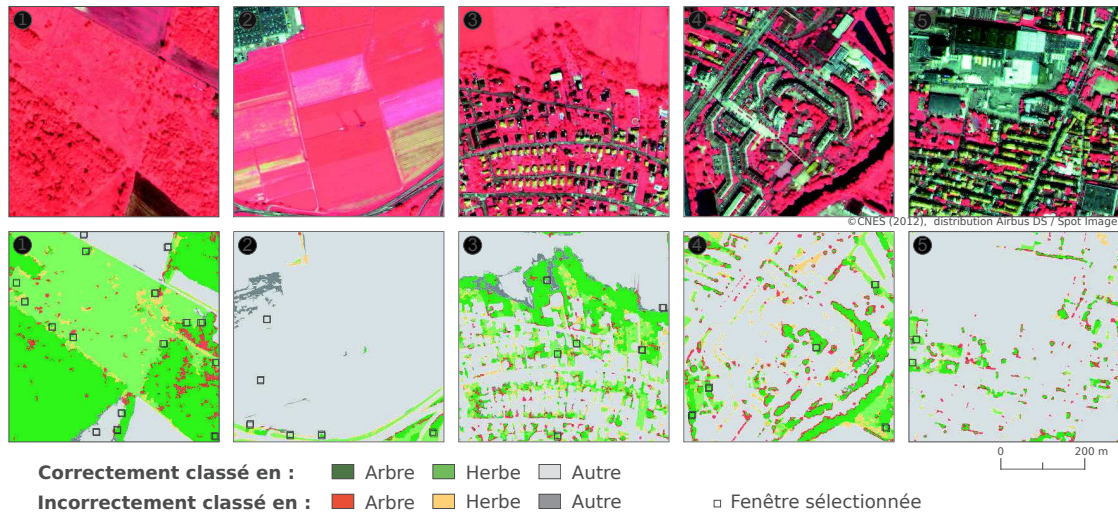


Figure 4.8 – Exemple de résultats de classification réalisée avec MCSAL<sub>S</sub> ( $w = 15$  m,  $n = 40$ ) par strate d'urbanisation pour le site de Strasbourg

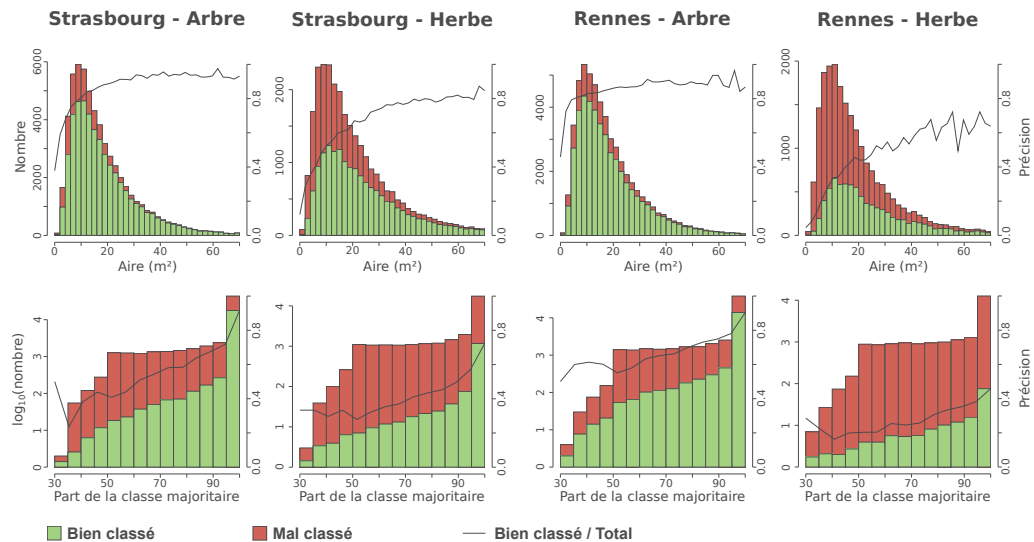


Figure 4.9 – Évaluation des résultats pour la végétation arborée et herbacée selon l'aire des segments (haut) et la pureté de la classe attribuée à un segment par la vérité-terrain (bas)

Tuia et Munoz-Mari (2013) proposent de prendre en compte l'incertitude de l'utilisateur pour ne pas intégrer ces exemples dans la classification par apprentissage actif. Ici, les classes sont directement attribuées aux segments à chaque nouvelle itération. Si une



numérisation de l'image est utilisée pour la labélisation, il serait donc possible de ne pas intégrer dans le modèle les exemples dont la pureté serait inférieure à un seuil. Une deuxième solution serait de leur donner un poids plus ou moins important dans le processus de classification selon une mesure permettant d'exprimer cette pureté, par exemple l'entropie.

## **Conclusion**

Ce chapitre a permis de proposer plusieurs stratégies d'apprentissage actif afin de diminuer le nombre d'échantillons nécessaire à l'obtention de classifications performantes et ainsi diminuer le coût nécessaire à la création de cette base d'apprentissage. Trois méthodes ont été développées et évaluées par un échantillonnage aléatoire stratifié qui est une méthode souvent utilisée en télédétection pour la sélection d'exemples diversifiés. Plusieurs points lors de cette évaluation se sont révélés intéressants.

Tout d'abord, l'utilisation d'une entropie normalisée telle qu'il est conseillé n'a pas permis de montrer son apport pour sélectionner des échantillons plus informatifs. Toutefois, seulement trois classes étaient considérées et des tests avec plus de classes devraient être menés afin de vérifier si la normalisation de l'entropie aurait un impact plus important.

Dans les publications portant sur l'apprentissage actif, les résultats sont souvent comparés à un simple échantillonnage aléatoire. Ce chapitre a toutefois révélé que s'il est stratifié, il permet d'être compétitif avec les stratégies développées dans ce travail et devrait donc être plus utilisé pour l'évaluation des méthodes d'apprentissage actif. De plus en termes de précisions globales, ces stratégies permettent toutes de surpasser l'échantillonnage aléatoire stratifié. La méthode MCSAL<sub>S</sub> s'est révélée la plus efficace, car sans augmenter le nombre de fenêtres à labéliser, elle permet de réduire fortement le temps d'exécution. Les performances par classes sont toutefois plus contrastées. En effet, celles obtenues pour la végétation herbacée sont parfois meilleures avec SRGS ce qui montre une limite importante aux stratégies d'apprentissage actif proposées pour la détection de classes fortement minoritaire.

Enfin, ce chapitre a aussi démontré l'apport que représente l'utilisation de données auxiliaires telles que les images multi-dates et les informations topographiques pour l'amélioration de l'identification de la végétation urbaine. Ces tests réalisés sur le site de Strasbourg ont montré que l'utilisation des informations de hauteurs sans les multi-temporelles est un bon compromis entre performances et complexité. Il serait donc intéressant de pouvoir profiter des capacités stéréoscopiques des satellites Pléiades pour obtenir simultanément les informations spectrales et topographiques.

Plusieurs points nécessitent néanmoins d'être améliorés dans les méthodes proposées. Le premier est de pouvoir utiliser non plus une représentation ponctuelle des segments

pour créer la carte d'entropie, car des pixels vides peuvent apparaître, mais une représentation continue sur la zone d'étude. Une autre solution serait de ne plus créer cette carte par un raster, mais utiliser les relations topologiques de chaque segment pour calculer l'entropie locale de chacun. La création du graphe d'adjacence permettrait de rendre compte de ces relations pour lequel il est possible de considérer des poids différents aux segments tels que leur surface pour calculer l'entropie locale afin de favoriser les segments de grandes dimensions.

Les deux chapitres précédents ont permis d'aborder deux points importants d'un processus d'analyse d'image par une approche orientée-objet et d'une classification supervisée, la segmentation et la sélection d'échantillons. Ces deux chapitres se sont concentrés sur deux extraits. L'objet du prochain chapitre est donc dans un premier temps de généraliser ces méthodes de segmentation et de sélection d'échantillons sur l'ensemble des sites d'étude de Strasbourg et de Rennes couvert par les images Pléiades. Dans un deuxième temps, la classification finale des images, les post-traitements et la validation des résultats obtenus sont réalisés.

*Valorisation scientifique associée à ce chapitre :*

**Rougier S.**, Puissant A., Stumpf A., Lachiche N., 2016, « Comparison of sampling strategies for object-based classification of urban vegetation from Very High Resolution satellite images », *International Journal of Applied Earth Observation and Geoinformation*, vol. 51, pp. 60–73

# Chapitre 5

## Généralisation de l'approche orientée-objet par apprentissage actif pour la classification de la végétation

### Sommaire

---

Introduction . . . . .	138
5.1 Problèmes liés à la taille des images et solutions utilisées . . . . .	138
5.1.1 Tuilage des zones d'étude . . . . .	139
5.1.2 Sélection de nouveaux exemples d'entraînement par apprentissage actif . . . . .	140
5.2 Sélection des méthodes de classification et de leurs paramètres . . . . .	141
5.2.1 Choix des algorithmes de classification . . . . .	141
5.2.2 Sélection des paramètres des méthodes de classification . . . . .	143
5.3 Classifications finales de la végétation arborée et herbacée . . . . .	144
5.3.1 Sélection des mailles pour la validation des classifications . . . . .	144
5.3.2 Évaluation des méthodes de classification par les mailles de validation . . . . .	145
5.3.3 Protocole de post-traitements et première évaluation des résultats . . . . .	148
5.4 Évaluation des bases de données végétation . . . . .	151
5.4.1 Évaluation des classifications selon les strates d'urbanisation . . . . .	152
5.4.2 Évaluation des classifications selon l'utilisation du sol . . . . .	152
5.4.3 Évaluation à l'échelle des objets . . . . .	156
Conclusion . . . . .	162

---

## Introduction

Ce chapitre vise à cartographier de manière exhaustive la végétation arborée et herbacée pour les sites de Strasbourg et de Rennes par la méthode orientée-objet par apprentissage actif présentée lors des deux chapitres précédents. Toutefois dans ces chapitres seuls des extraits d'images étaient utilisés pour tester les méthodes proposées. Le passage à une étendue plus grande pose le problème de la faisabilité de ces traitements du fait de la taille des images.

La section 5.1 présente tout d'abord les stratégies de traitements mises en place pour résoudre les problèmes liés à la taille des images afin de les segmenter, calculer les attributs et sélectionner les échantillons d'entraînement par apprentissage actif. Nous testons ensuite différents algorithmes de classification à partir des données d'entraînement sélectionnées afin de déterminer, pour chaque site, celui qui est le plus performant. La section suivante présente les méthodes de classification retenues et la sélection des paramètres optimaux pour chacune de ces méthodes (§ 5.2). La section 5.3 de ce chapitre présente les résultats de l'extraction de la végétation arborée et herbacée de Strasbourg et de Rennes. Cette section présente tout d'abord la sélection des données utilisée pour la validation puis les performances de chaque algorithme de classification retenu avec ces paramètres optimaux. Dans un deuxième temps, le protocole de post-traitements est décrit et réalisé et les résultats des classifications de la végétation et de l'application de ce protocole sont évalués. Enfin, la dernière section de ce chapitre évalue précisément ces classifications pour mieux cerner les causes et la localisation des erreurs (§ 5.4).

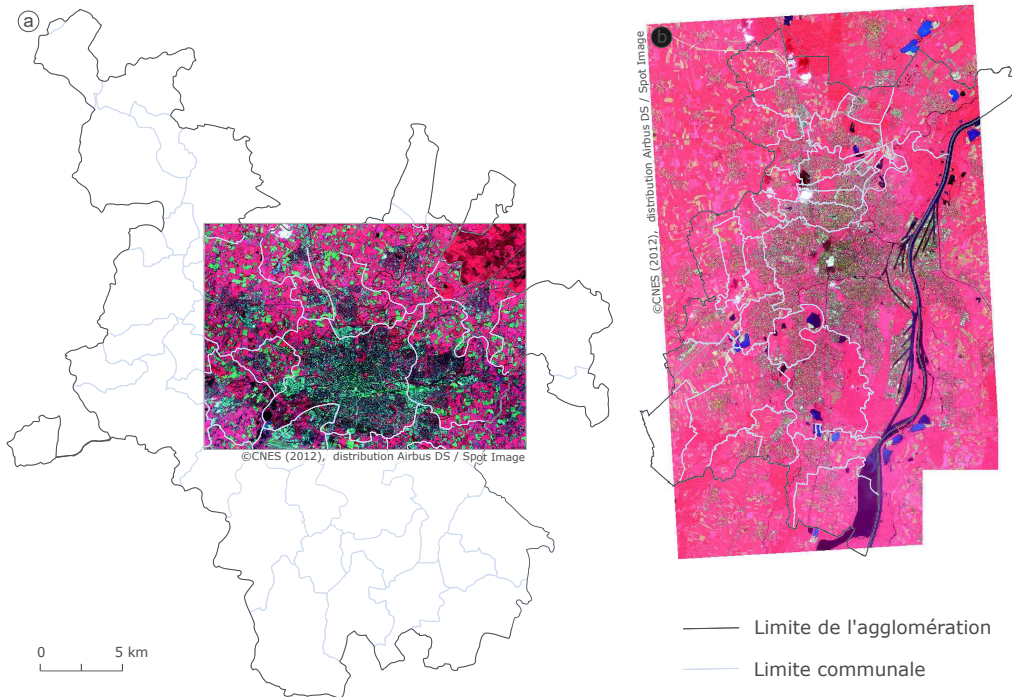
### 5.1 Problèmes liés à la taille des images et solutions utilisées

Pour le site de Strasbourg, la zone à traiter correspond à l'emprise en commun des images Pléiades du 14 août et du 21 septembre 2012 et du MNT fourni dans le cadre du partenariat CIGAL. Ainsi cette zone s'étend sur environ 540 km<sup>2</sup> et comprend la quasi-totalité de l'agglomération (figure 5.1.b). Cette zone intègre aussi une large partie située en Allemagne. Une partie de cette zone n'est pas comprise dans l'emprise de l'image du 23 avril 2013 (figure 3.2). Les attributs dérivés de cette image calculés pour les segments en dehors de cette emprise sont remplacés par la médiane comme il est proposé par Liaw et Wiener (2002) pour des classifications avec Random Forest.

Pour Rennes qui dispose d'une seule image, l'ensemble de l'image n'est pas utilisé. En effet, la partie sud de l'image est très voilée ne permettant pas d'observer la végétation de manière claire et donc d'obtenir une classification satisfaisante. Seule la partie nord a donc été conservée (figure 5.1.a) ce qui représente une zone d'environ 275 km<sup>2</sup>.

La classification de ces deux zones n'est pas sans difficulté. La taille est en effet un facteur limitant puisque la consommation de mémoire vive est ici très importante que ce soit pour l'étape de segmentation ou pour le calcul d'attributs puisque le logiciel eCognition

ne permet pas d'appliquer la méthode Multi-Resolution sur de larges images, mais aussi lors de la classification avec R qui conserve l'ensemble des données en mémoire vive. Ainsi il a été nécessaire de mettre en place une stratégie adaptée pour ces deux étapes.



**Figure 5.1** – Localisation des extraits des images Pléiades considérés pour les sites de Rennes (a) et de Strasbourg (b)

### 5.1.1 Tuilage des zones d'étude

Afin de réaliser la segmentation et le calcul de l'ensemble des attributs sur nos deux zones d'études, un tuilage est adopté. La taille approximative des tuiles a été trouvée de manière empirique en maximisant leur taille afin de réduire leur nombre tout en rendant possible l'ensemble des calculs. Ainsi le site de Strasbourg est composé de 12 zones et celui de Rennes de 6. L'ensemble de ces tuiles est traité de manière indépendante en utilisant les mêmes paramètres de segmentation que dans la partie précédente et en en considérant aussi les mêmes attributs (chapitre 4).

Au final, la segmentation a produit environ 11 millions de segments pour le site de Rennes et environ 22 millions pour celui de Strasbourg. L'ensemble de ces données a ensuite été chargé dans R en conservant toujours ce découpage en tuile.

### 5.1.2 Sélection de nouveaux exemples d'entraînement par apprentissage actif

Afin de réaliser la classification de l'ensemble de l'image, de nouveaux exemples d'entraînement sont sélectionnés pour que toute la diversité de nos images soit considérée et non seulement une partie. La méthode pour sélectionner ces nouveaux échantillons est celle d'apprentissage actif MCSAL<sub>S</sub> puisque c'est celle qui a permis d'obtenir de meilleurs résultats d'après le chapitre précédent. Toutefois, certains des paramètres ont été modifiés afin de pouvoir réaliser cette étape puisque le même problème que pour la segmentation est apparu. La taille du raster pour la projection des valeurs d'entropie a été fixée à 10 m. En effet, il n'était pas possible de conserver la même valeur de 5 m que précédemment et cette taille permet de plus de n'avoir aucun pixel sans segments qui pouvait poser problème lors du chapitre précédent. De plus, pour des questions de temps d'exécution une fenêtre  $w$  de 100 m est utilisée.

Chaque tuile est traitée indépendamment pour la sélection de ces nouvelles fenêtres dans le but de réduire le nombre d'itérations qui est ici fixé à 4, mais aussi d'avoir des exemples bien distribués sur l'ensemble de l'image. C'est-à-dire qu'à chaque itération de MCSAL<sub>S</sub>, cinq nouvelles fenêtres par tuile de l'image vont être sélectionnées. Tout comme pour le chapitre 4 l'indice rendant compte de l'artificialisation va être discrétisé en cinq strates, mais ici pour chaque découpage réalisé pour la segmentation et le calcul des attributs. Ainsi 60 nouvelles fenêtres de 100 m sont ajoutées au jeu d'entraînement pour le site de Strasbourg à chaque itération et 30 pour celui de Rennes. Pour les deux sites, la carte de stratification est basée sur le MSAVI moyen. Il n'était en effet pas possible de calculer l'indice d'urbanisation pour le site de Strasbourg puisqu'une partie de l'image comprend l'Allemagne et qu'aucune donnée semblable à la base de données d'occupation des sols de l'Alsace n'était disponible.

Finalement se sont donc 240 nouvelles fenêtres qui ont été photo-interprétées pour Strasbourg et 120 pour Rennes. La numérisation de ces nouveaux exemples d'entraînement a suivi les mêmes règles que la numérisation des mailles précédentes (§ 3.2.4) avec tout de même une différence pour Strasbourg. En effet précédemment, la photo-interprétation n'était réalisée que sur l'image d'août ce qui était possible puisque les mailles avec des nuages n'étaient pas considérées lors de la sélection des mailles d'entraînement et de validation. Lors de l'application de la méthode d'apprentissage actif, certaines zones comprenaient des parties nuageuses et lorsque le cas se présentait, alors la numérisation de la végétation a été réalisée sur l'image de septembre.

Pour Strasbourg, 373 870 segments constituent le jeu d'entraînement et pour Rennes 273 503. Pour ces deux sites, les classes sont toujours fortement déséquilibrées avec une faible présence d'arbre, mais surtout de végétation herbacée (figure 5.2).

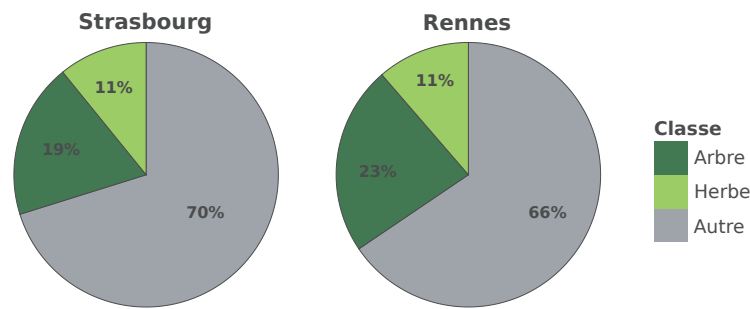


Figure 5.2 – Répartition par classe des segments des jeux de données d'entraînement de Rennes et de Strasbourg

## 5.2 Sélection des méthodes de classification et de leurs paramètres

L'objectif de cette section est de produire les bases de données décrivant les a végétation arborée et herbacée pour l'ensemble des sites d'études de Strasbourg et de Rennes. La première étape est de réaliser les classifications définitives pour ces deux sites en utilisant les données d'entraînement sélectionnées par apprentissage actif et en comparant différents algorithmes souvent utilisés dans le domaine de la classification d'image. Les différents algorithmes de classification sont précisés dans la section 5.2.1 et le choix des paramètres optimaux pour chacune de ces méthodes constitue l'objet de la section 5.2.2.

### 5.2.1 Choix des algorithmes de classification

Quatre méthodes de classification sont testées afin de vérifier laquelle est le plus performante pour extraire la végétation urbaine à partir des échantillons qui ont été sélectionnés grâce à la méthode d'apprentissage actif. Ces quatre algorithmes sont souvent utilisés en apprentissage automatique ainsi qu'en télédétection et leur comparaison a montré que les performances des classifications obtenues sont assez proches (Laliberte *et al.*, 2006 ; Duro *et al.*, 2012a ; La Rosa et Wiesmann, 2013). Il s'agit de celui des plus proches voisins ( $k$ -NN –  $k$ -Nearest Neighbor), des machines à vecteur de support (SVM), de Random Forest, et d'un arbre de décision, C 5.0. L'ensemble de ces méthodes de classification présente aussi l'avantage d'être intégrées dans R. Pour ces différentes méthodes, plusieurs paramètres sont optimisés grâce au package caret (Kuhn et Johnson, 2013). Ces paramètres sont, pour chaque méthode, déterminés automatiquement en utilisant une validation croisée stratifiée en 10 plis. La stratification utilisée ici ne repose pas sur des mailles indépendantes comme c'était le cas dans le chapitre précédent, mais uniquement sur la classe des exemples. Ces exemples n'étant pas réellement indépendants spatialement entre ceux d'entraînement et ceux de tests dans la validation croisée, les résultats obtenus ne peuvent être considérés directement comme pertinents

puisqu'ils seraient positivement biaisés (§ 4.5.5). Toutefois, elle permet ici de déterminer les paramètres les plus performants pour chacune des méthodes. Une deuxième phase de validation est réalisée à partir des mailles de validation qui sont elles totalement indépendantes, et ce pour chacune des quatre méthodes testées ici.

La méthode RF a déjà été expliquée en détail lors des chapitres précédents et ne sera pas expliquée ici. Pour cette méthode néanmoins un seul paramètre est optimisé ici, il s'agit de  $mtry$ , le nombre de variables aléatoirement sélectionnées à chaque nœud. Les trois autres algorithmes sont très succinctement exposés ici.

K-NN est un algorithme qui va rechercher dans l'espace des données les voisins les plus proches de chaque exemple sans label afin de déterminer sa classe. La classe attribuée est celle la plus représentée dans les  $k$  voisins les plus proches qui sont étiquetés,  $k$  étant un paramètre défini par l'utilisateur.

Les SVM sont des méthodes de classification supervisées très utilisées actuellement dans de nombreux domaines, car elles sont très performantes. Il s'agit ici de méthodes de classification non linéaire qui utilisent une fonction de noyaux et qui peuvent être utilisées dans des espaces de données de grandes dimensions. Le principe est de trouver la fonction qui permettra de séparer au mieux l'hyperplan des exemples utilisés pour l'entraînement. Les exemples de chaque classe les plus près de cet hyperplan sont appelés les vecteurs support. La méthode que nous utilisons dans cette recherche est une fonction à base radiale où deux paramètres sont définis par l'utilisateur :  $\sigma$  et coût.  $\sigma$  correspond à la largeur du noyau de la fonction radiale et est déterminé automatiquement grâce au package `caret`. Le coût est donc le seul paramètre à réellement déterminer de manière empirique et représente une pénalité donnée par l'utilisateur aux résidus.

Enfin, le dernier algorithme comparé dans cette recherche est le C 5.0. Cette méthode est une amélioration du C 4.5 qui permet d'apprendre avec deux types de modèles : l'un basé sur un arbre de décision et l'autre basé sur un ensemble de règles. Afin de définir la meilleure coupe à chaque nœud de l'arbre lorsque ce modèle est utilisé, cette méthode se base sur l'entropie qui mesure la diversité des classes dans chaque feuille à la différence des CART qui eux utilisent l'indice de Gini. Si cette feuille n'est pas pure alors un nouveau nœud est créé à partir des exemples qui la composent. Cet arbre va ensuite être élagué afin d'éviter l'effet de sur apprentissage et pour réduire la taille de l'arbre. Dans le cadre d'un modèle à base de règles, un arbre initial est construit puis simplifié plus fortement par élagage afin que les règles soient suffisamment simples. À la différence du modèle à base d'arbre de décision, un exemple peut appartenir à plusieurs classes et un degré de confiance est donc attribué pour chacune. La classe attribuée est celle avec la plus grande confiance. Une autre évolution de C 5.0 est d'intégrer une procédure de boosting. Le principe du boosting est de combiner plusieurs classifieurs faibles, c'est-à-dire des classifieurs permettant d'obtenir des résultats simplement meilleurs que l'aléatoire, afin de renforcer un classifieur fort. Chaque arbre est ici évalué séparément pour permettre le calcul d'un indice de confiance. Ainsi la classe avec l'indice de



confiance le plus élevé pour chacun des exemples est attribuée. Deux paramètres sont donc testés pour cet algorithme : le type de modèle, arbre ou règle, ainsi que le nombre d'essais, le nombre d'itérations de boosting.

### 5.2.2 Sélection des paramètres des méthodes de classification

La figure 5.3 présente les résultats obtenus pour les dix plis de validation croisée avec les meilleurs paramètres pour chacune des méthodes ainsi que les paramètres retenus. Pour ces deux sites et pour RF, le nombre de variables sélectionnées aléatoirement pour chaque nœud est plus important que le paramètre par défaut utilisé traditionnellement avec cette méthode. Insérer un nombre important d'itérations de boosting dans la méthode C 5.0, ici 90, permet d'améliorer les résultats comme utiliser un nombre relativement important de voisins dans la méthode k-NN. Cette figure montre que pour les deux sites d'études les résultats de k-NN sont beaucoup moins performants qu'avec les trois autres méthodes. Si SVM apparaît légèrement plus performant sur le site de Rennes, pour celui de Strasbourg c'est plutôt RF qui permet les meilleurs résultats.

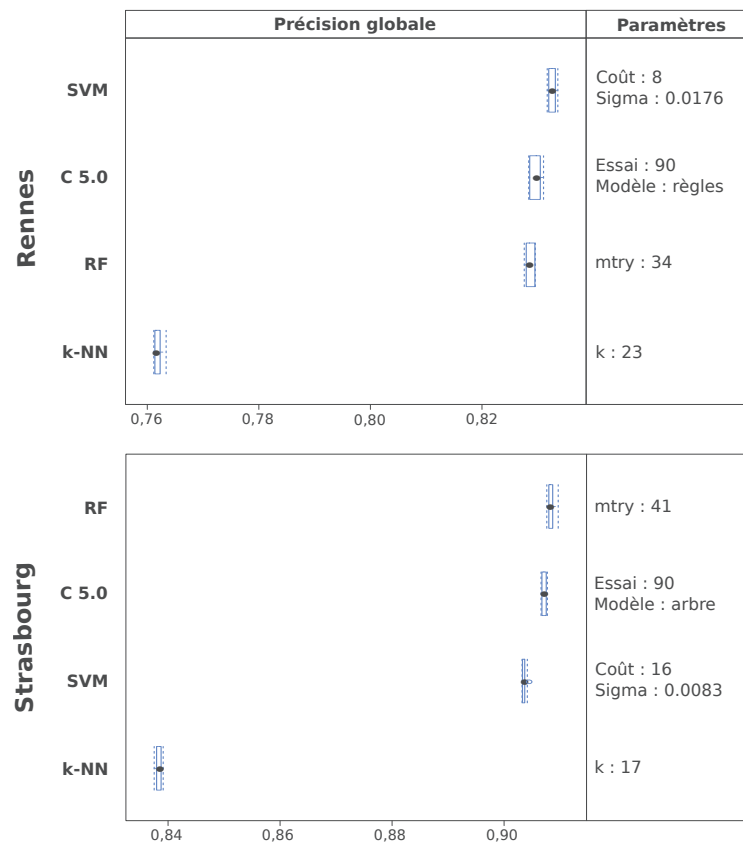


Figure 5.3 – Résultats de la sélection des paramètres par validation croisée des quatre algorithmes de classification testés pour Rennes et Strasbourg

### 5.3 Classifications finales de la végétation arborée et herbacée

L'objectif de cette section est de réaliser les classifications finales de la végétation arborée et herbacée de Strasbourg et de Rennes. Dans un premier temps, la section 5.3.1 présente les données qui ont été construites par photo-interprétation des images pour constituer le jeu de données de validation. La section suivante permet de sélectionner à partir de ce jeu de validation la méthode de classification permettant d'obtenir les meilleures performances pour chacun de ces sites en considérant pour chacune de ces méthodes les paramètres sélectionnés dans la section 5.2.2. Enfin, la section 5.3.3 présente et évalue le protocole de post-traitements appliqué à ces deux classifications afin de produire les bases de données finales.

#### 5.3.1 Sélection des mailles pour la validation des classifications

Pour évaluer nos deux classifications, de nouvelles mailles destinées à la validation ont été sélectionnées afin de couvrir toute la diversité de l'image. Une stratification basée sur le MSAVI a été utilisée pour les deux sites d'études et les deux sont validés en considérant environ 1 % de la surface des images. Un maillage a donc été réalisé avec des mailles carrées de 100 m de côté et les mailles intersectant celles utilisées pour l'entraînement ne sont pas considérées. Le MSAVI est ensuite calculé pour l'ensemble en utilisant l'image d'août pour le site de Strasbourg et est ensuite discrétisé en un nombre de strates égal au nombre de mailles à photo-interpréter afin d'en sélectionner une par strate de manière aléatoire. Pour Strasbourg 536 mailles constituent le jeu de test et pour Rennes 275.

La répartition de ces mailles et de leur occupation du sol pour le site de Strasbourg et uniquement sur la partie française se rapproche fortement de celle de l'ensemble de l'agglomération (figure 5.4). Les différences en terme de répartition sont au maximum de 2 % ce qui montre que l'utilisation d'une stratification basée sur le MSAVI peut être utilisée afin d'obtenir une bonne représentation de la diversité de la zone d'étude.

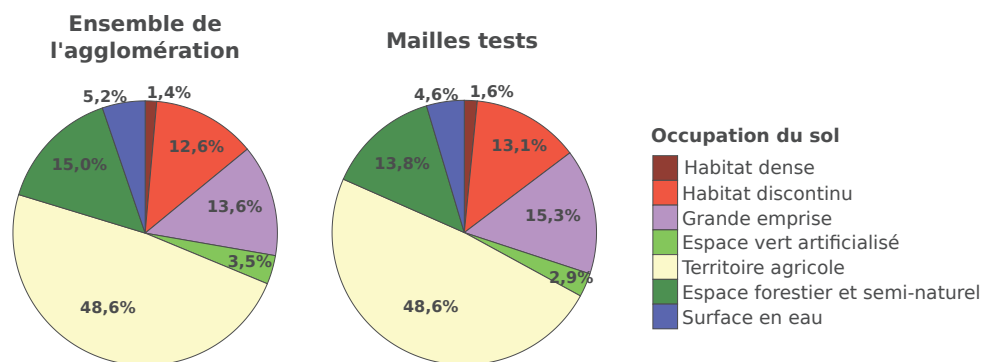


Figure 5.4 – Occupation du sol de l'Eurométropole de Strasbourg et des mailles de validation

### 5.3.2 Évaluation des méthodes de classification par les mailles de validation

Chaque méthode de classification considérée a ensuite pu être testée avec ces paramètres optimaux sur les 100 mailles de validation. Les performances obtenues avec l'algorithme k-NN, que ce soit en termes de segments ou d'aire, sont très en deçà de celles réalisées avec les trois autres méthodes pour nos deux sites d'études comme c'était déjà le cas lors de la détermination des meilleurs paramètres (tables 5.1 et 5.2). Les méthodes RF, SVM et C 5.0 permettent d'obtenir des résultats assez comparables.

Que ce soit pour Rennes ou pour Strasbourg, les classifications réalisées avec SVM apparaissent légèrement plus performantes quand le nombre de segments est utilisé pour le calcul des précisions. En effet, l'amélioration de F-mesure est d'environ 2 % pour le site de Rennes quelle que soit la classe de végétation considérée.

Si la différence est minime pour Strasbourg et la végétation arborée, la différence de F-mesure est beaucoup plus importante pour la végétation herbacée atteignant environ 5 %. En considérant maintenant l'aire des segments, les différences entre ces trois méthodes de classification sont moins importantes.

**Tableau 5.1** – F-mesures par classes en segments et en aire pour chaque algorithme testé avec ses paramètres optimaux pour le site de Rennes

	Segments			Aire		
	Arbre	Herbe	Autre	Arbre	Herbe	Autre
k-NN	0,746	0,171	0,894	0,754	0,165	0,901
RF	0,817	0,323	0,917	0,843	0,355	0,924
SVM	0,836	0,353	0,924	0,852	0,346	0,911
C 5.0	0,812	0,332	0,915	0,829	0,347	0,925

Pour Rennes notamment, les F-mesures des classes herbe et autre apparaissent moins bonnes avec SVM en segments qu'en aire. Pour cet algorithme, et au contraire de ce qui été remarqué dans la section 4.5.5, les segments de grandes tailles n'apparaissent pas plus faciles à classer que les petits. À l'inverse, les performances sont meilleures lorsqu'elles sont mesurées avec l'aire que ce soit pour RF et C 5.0. Pour le site de Strasbourg par contre, les performances, quelle que soit la méthode d'apprentissage utilisée, sont améliorées lorsque l'aire est considérée.

**Tableau 5.2** – F-mesures par classes en segments et en aire pour chaque algorithme testé avec ses paramètres optimaux pour le site de Strasbourg

	Segments			Aire		
	Arbre	Herbe	Autre	Arbre	Herbe	Autre
k-NN	0,801	0,327	0,918	0,818	0,350	0,918
RF	0,899	0,457	0,945	0,923	0,485	0,954
SVM	0,902	0,511	0,944	0,923	0,524	0,953
C 5.0	0,897	0,460	0,954	0,921	0,483	0,954

La classification réalisée avec les séparateurs à vaste marge est assez nettement celle qui permet d'obtenir de meilleurs résultats. En effet si les F-mesures de la végétation arborée et de la classe autre sont très semblables, celle obtenue pour la végétation herbacée est bien meilleure. C'est donc cette méthode qui est appliquée à l'ensemble de l'image pour le site de Strasbourg. Pour le site de Rennes, les classifications sont quant à elles beaucoup plus difficiles à départager. Les meilleures F-mesures sont différentes selon l'algorithme utilisé : RF pour la végétation herbacée, SVM pour celle arborée et C 5.0 pour la classe autre. Notre choix s'est porté sur RF afin d'être appliqué sur l'ensemble de la zone d'étude de Rennes pour deux raisons. La première est que la moyenne des F-mesures des trois classes est ici légèrement plus élevée que pour les deux autres et la deuxième est que la prédiction sur l'ensemble des données est beaucoup plus rapide avec RF (tableau 5.3).

**Tableau 5.3** – Temps d'exécution en minutes pour l'entraînement avec l'ensemble des données labélisées et pour la prédiction des données de test (1 % des images) de chaque algorithme testé

	Strasbourg		Rennes	
	Entraînement	Prédiction	Entraînement	Herbe
k-NN	/	460,7	/	116,1
RF	160,3	0,4	54,3	0,2
SVM	493,8	41,9	227,1	22,3
C 5.0	354,3	2,30	1429,4	13,3

Les figures 5.5 et 5.6 présentent la distribution des segments selon différents attributs qui sont apparus *a priori* déterminants pour comprendre mieux les erreurs commises par les deux méthodes d'apprentissage. La brillance moyenne, le MSAVI moyen ainsi que l'aire des segments ont été sélectionnés. Pour ces deux sites, l'aire apparaît importante pour expliquer ces erreurs puisque plus les segments sont grands, plus il est probable qu'ils soient bien classés, les faux positifs et négatifs étant principalement de petite dimension. Ce constat confirme celui de la section 4.5.5 et montre de plus que cet effet ne concerne pas seulement la méthode RF, mais aussi SVM pour Strasbourg et que les omissions ou les commissions sont tout autant concernées.

Les deux autres attributs sélectionnés pour comprendre la distribution des erreurs sont la moyenne de la brillance et celle du MSAVI par segments. C'est pour les plus sombres, avec une brillance faible, que les confusions de la végétation herbacée sont les plus importantes, aussi bien pour les faux négatifs que les positifs, et ce pour les deux sites d'étude. Les faux positifs sont très souvent situés en bordure des groupes d'arbres, dans leur ombre. Pour la végétation arborée par contre omissions et commissions sont différenciées selon leur brillance. En effet mis à part pour Strasbourg où une fréquence importante apparaît dans les valeurs de brillance faibles, les commissions ont une plus grande probabilité d'être commises lorsque la brillance est élevée. Au contraire, les faux négatifs sont concentrés sur les objets ayant une brillance faible. Ces segments correspondent notamment à des segments de petites dimensions situés au sein de groupement d'arbres et liés à leur ombre portée que la segmentation de niveau 2 et les attributs de voisinage n'ont pas permis de corriger.

La moyenne du MSAVI montre pour cette classe des résultats différents. C'est en effet pour les segments de végétation arborée avec une moyenne faible que les erreurs de commissions et d'omissions sont les plus fréquentes. Pour la végétation herbacée par contre, les résultats sont beaucoup moins clairs. En effet pour Rennes, seuls les faux positifs sont plus probables pour cette classe lorsque le MSAVI est faible tandis que pour Strasbourg, faux positifs et faux négatifs sont plus fréquents dans les valeurs hautes et basses du MSAVI comparé aux segments qui sont bien classés.

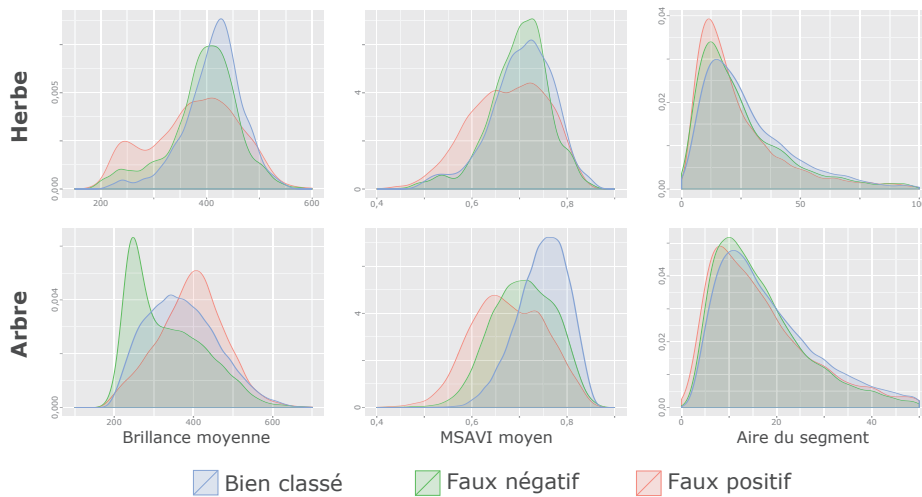


Figure 5.5 – Répartition par classe des segments correctement classés, des faux positifs et négatifs selon la moyenne de leur brillance, du MSAVI et de l'aire pour le site de Rennes

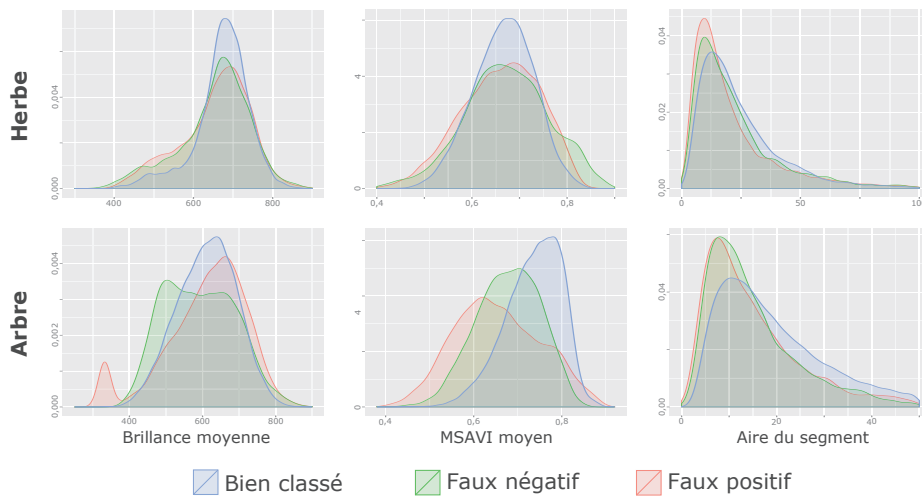


Figure 5.6 – Répartition par classe des segments correctement classés, des faux positifs et négatifs selon la moyenne de la brillance, du MSAVI et de l'aire pour le site de Strasbourg

### 5.3.3 Protocole de post-traitements et première évaluation des résultats

Afin d'améliorer la précision des bases de données décrivant les formes végétales de Strasbourg et de Rennes, un ensemble de post-traitements a été appliqué aux deux classifications sélectionnées dans la section précédente. Le premier est d'agréger l'ensemble des segments qui partagent un bord en commun et qui sont de la même classe au sein d'un même objet. Cette opération a permis d'obtenir environ 470 000 objets pour l'agglomération de Strasbourg et 420 000 pour Rennes.

Le deuxième post-traitement consiste à corriger les confusions entre la végétation herbacée et les parcelles agricoles. De nombreuses parcelles ont en effet été classées en herbe sans être considérées comme telles dans la vérité-terrain du fait que nous nous intéressons uniquement à la végétation herbacée urbaine. Afin de supprimer une partie de ces confusions, le RPG de 2012 est utilisé ici. Dans cette thèse, cette base a ainsi permis de supprimer les objets de végétation herbacée qui se superposent aux parcelles agricoles identifiées par le RPG. Ceux de végétation arborée ont par contre été conservés qu'ils soient ou non au sein de l'espace urbain.

Enfin, le dernier post-traitement effectué est la suppression des objets de trop petite dimension. Un seuil de  $6,5 \text{ m}^2$  est utilisé et correspond à celui que nous avons fixé lors de la création de la vérité-terrain (§ 3.2.4). Ainsi les objets inférieurs à ce seuil sont fusionnés à l'objet voisin, celui avec lequel la lisière commune est la plus grande. Ces classifications ainsi que leurs post-traitements ont permis d'obtenir une cartographie détaillée de la végétation arborée et herbacée pour les villes de Rennes (figure 5.7) et de Strasbourg (figure 5.8).

Les résultats avant post-traitements sont ici légèrement différents de ceux du chapitre 5.3.2 du fait que les résultats ne sont pas évalués à partir des segments ou de leur aire, mais qu'ils sont le résultat de l'intersection des classifications, post-traitées ou non, avec la vérité-terrain de validation.

L'utilisation de cette série de post-traitements permet d'améliorer de manière importante les précisions obtenues (figures 5.9 et 5.10). L'amélioration la plus importante concerne les précisions utilisateurs de la végétation herbacée, et ce pour les deux sites d'étude puisqu'elle est d'environ 14 % pour Rennes et de 7 % pour Strasbourg. Cet effet se répercute par la suite sur la F-mesure et est ici la conséquence de la suppression des éléments herbacés lorsqu'ils sont présents dans une parcelle agricole du RPG. Avec ces post-traitements, les F-mesures obtenues pour les deux sites pour cette classe sont équivalentes et atteignent 54 %. L'amélioration illustre la difficulté pour les deux algorithmes d'apprentissage utilisés à séparer la classe herbe du reste de la végétation basse tel que les cultures et montre donc la nécessité d'utiliser ou de trouver d'autres attributs, d'autres sources de données ou des séries temporelles plus longues ou plus adaptées pour les séparer.

L'augmentation des autres mesures de précision est due au fait de la suppression des objets de petite dimension, celle d'une surface inférieure à  $6,5 \text{ m}^2$ . Si cet effet n'a permis

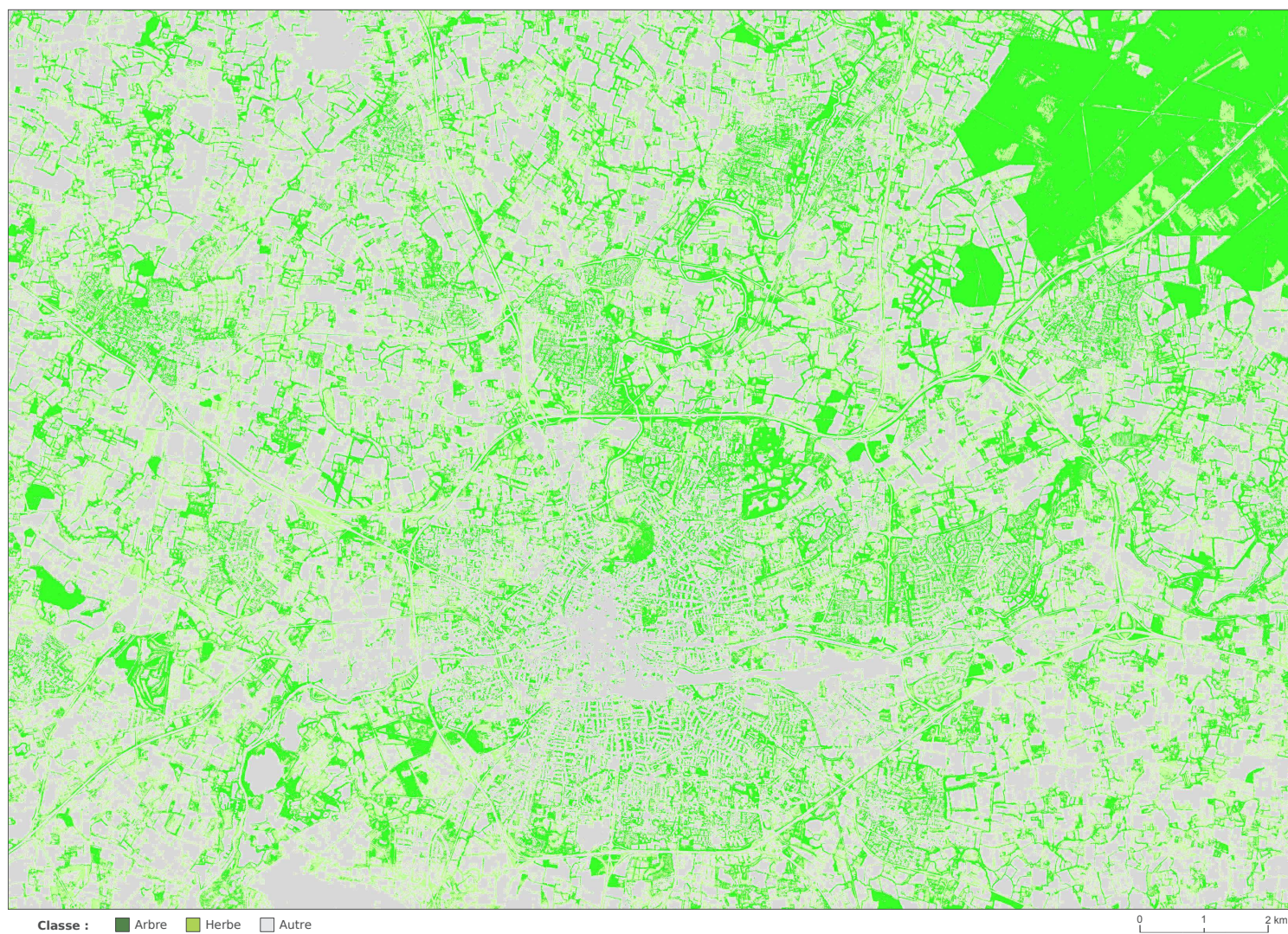


Figure 5.7 – Classification post-traitée de la végétation arborée et herbacée de Rennes

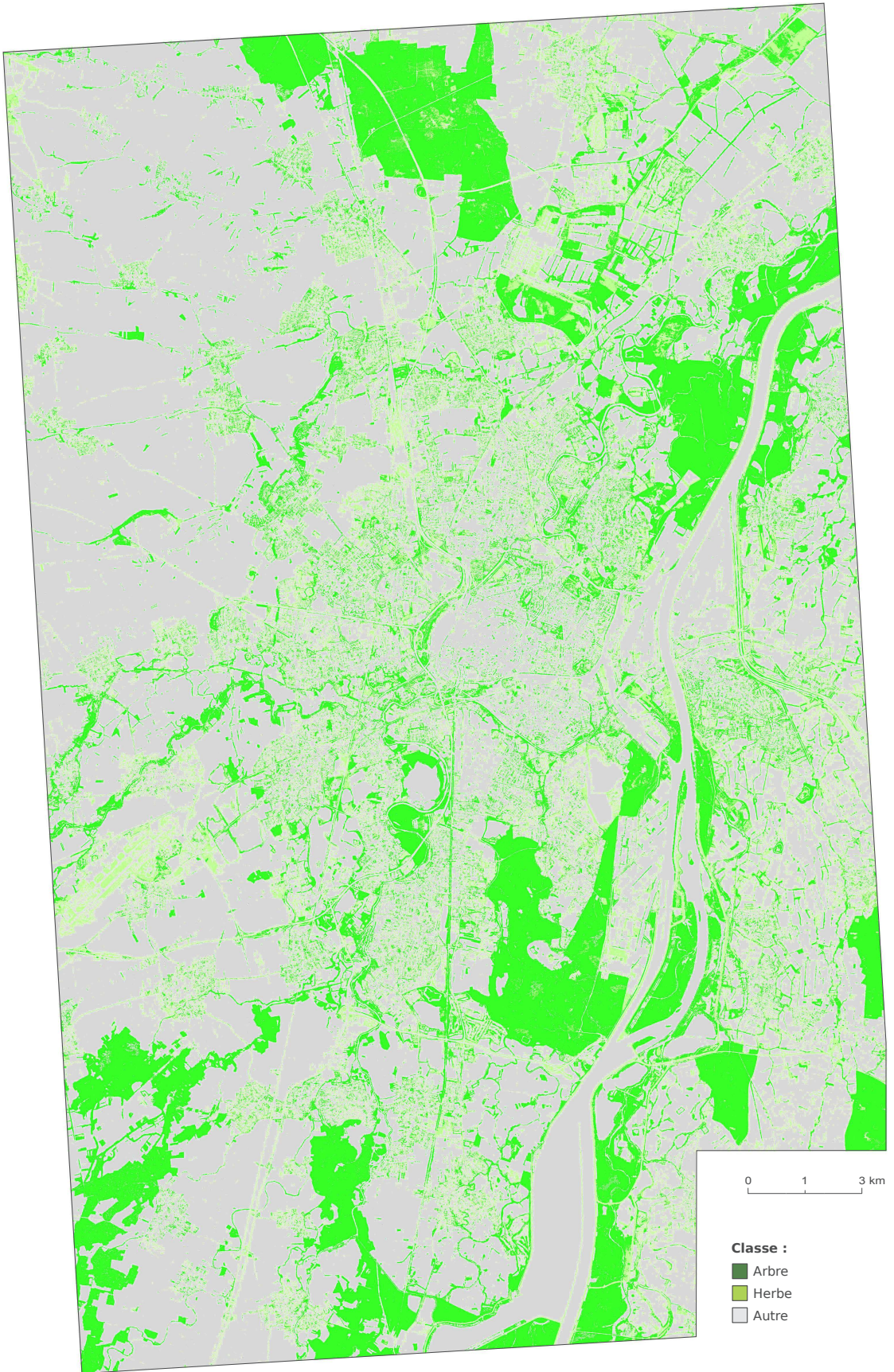


Figure 5.8 – Classification post-traitée de la végétation arborée et herbacée de Strasbourg



qu'une amélioration maximum des précisions de 2 %, visuellement elle apparaît plus fidèle à la vérité-terrain puisque cette étape a permis de supprimer de nombreux trous notamment dans les massifs boisés où les ombres créaient des confusions. La végétation arborée obtient ainsi une F-mesure de 91 % pour Strasbourg et de 81 % pour Rennes. Ce dernier post-traitement a de plus permis de supprimer un nombre important d'objets des classifications, environ 77 000 pour Strasbourg, soit 14 % du nombre total et 70 000 représentant 18 % pour Rennes. Ces deux bases sont ainsi fortement allégées ce qui facilite leur usage pour la suite de ce travail.

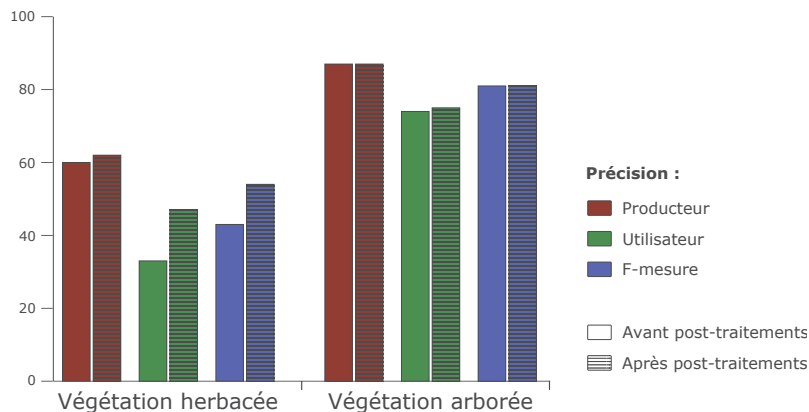


Figure 5.9 – Précisions de la classification de Rennes réalisée avec RF avant et après post-traitements

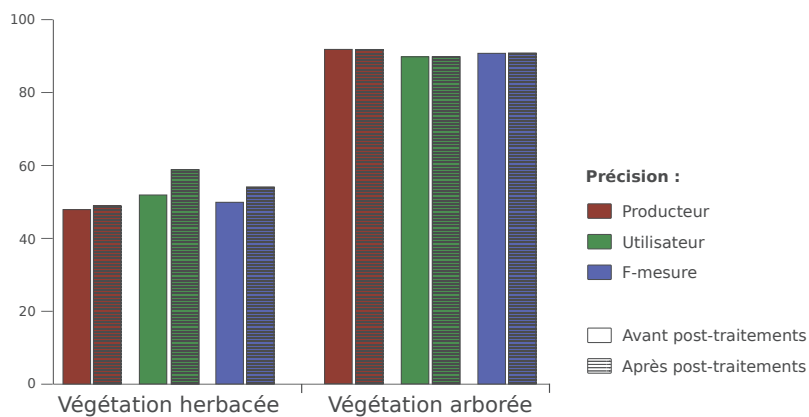


Figure 5.10 – Précisions de la classification de Strasbourg réalisée avec SVM avant et après post-traitements

## 5.4 Évaluation des bases de données végétation

Les performances des deux classifications post-traitées sont évaluées plus finement dans cette section afin de mieux saisir les causes et la localisation des erreurs. Trois échelles

d'analyse sont utilisées. La première est à l'échelle des grilles basées sur le MSAVI utilisé pour la stratification et une deuxième est réalisée à partir de l'occupation du sol. Enfin, une dernière analyse est réalisée à l'échelle de l'objet, ceux issus de la classification pour vérifier leur adéquation avec la vérité-terrain.

#### 5.4.1 Évaluation des classifications selon les strates d'urbanisation

De manière assez générale, l'augmentation du degré d'artificialisation voit diminuer les précisions obtenues pour la classe arbre et augmenter celles de la végétation herbacée (table 5.4). Pour les strates les plus végétalisées, la végétation arborée occupe une place très importante sous la forme de grande surface telle que de larges patches forestiers. Ces objets apparaissent assez faciles à classer pour les deux algorithmes et concernent la strate de MSAVI 1 pour Rennes et 1 et 2 pour Strasbourg. Pour la strate plus artificialisée, ces valeurs diminuent de manière assez importante. La végétation arborée ne représente dans ces strates qu'une plus faible fraction et est majoritairement composée d'arbres isolés qui sont plus difficilement à classer correctement.

La végétation herbacée montre de réelles différences comparées à l'analyse réalisée dans le chapitre 4. En effet, les résultats de la strate d'artificialisation 1 sont inversés entre Rennes et Strasbourg. C'est ici lié à la nature de l'occupation des sols des mailles qui ont été sélectionnées. Si lors du chapitre précédent, la strate 1 pour le site de Strasbourg comprenait une part assez importante d'herbe, elle est ici très faible ce qui impacte fortement les résultats à chaque erreur commise. L'effet inverse s'est produit pour le site de Rennes. De plus, de nombreuses parcelles agricoles ne sont pas représentées à Strasbourg, ce qui n'a pas permis d'éliminer de nombreuses confusions avec la classe autre.

**Tableau 5.4** – F-mesures par strate d'artificialisation pour Strasbourg et Rennes à partir des classifications post-traitée

	Strasbourg			Rennes		
	Arbre	Herbe	Autre	Arbre	Herbe	Autre
1 – Très végétalisée	0,95	0,25	0,98	0,90	0,52	0,93
2 – Végétalisée	0,95	0,47	0,97	0,74	0,45	0,94
3 – Médian	0,90	0,36	0,94	0,77	0,48	0,90
4 – Artificialisée	0,85	0,59	0,94	0,76	0,58	0,92
5 – Très artificialisée	0,85	0,60	0,98	0,74	0,59	0,94

#### 5.4.2 Évaluation des classifications selon l'utilisation du sol

La BDOCS de 2012 est utilisée dans cette section pour le site de Strasbourg et a nécessité une simplification afin que suffisamment de mailles de validation soient contenues dans chacun des types et qu'elle puisse correspondre en partie à la base utilisée pour Rennes. Une typologie en huit classes a été retenue en combinant plusieurs niveaux de

cette base de données (figure 5.12.a). Cette base n'étant disponible que sur la partie française, seules les mailles de validation situées à l'ouest du Rhin sont considérées dans cette section, soit 450 sur les 536 initialement utilisées. Tous les types ne seront pas explicités ici. L'habitat discontinu comprend aussi ben des espaces d'habitat collectif que pavillonnaire. Le type grandes emprises est composé de l'ensemble des espaces urbains spécialisés tels que les emprises hospitalières et universitaires ainsi que des espaces commerciaux industriels ou le réseau. Enfin, les espaces verts artificialisés comprennent les parcs urbains ainsi que l'ensemble des équipements sportifs.

Pour le site de Rennes, les données d'occupation du sol utilisées ont été réalisées par photo-interprétation à l'échelle de l'îlot, défini à partir des entités linéaires de la BD TOPO® de l'IGN : les réseaux routiers, ferrés et hydrographiques. La zone d'étude n'a pas été interprétée dans son ensemble et seule la tache urbaine centrale est considérée (figure 5.12.b). Ainsi ce sont seulement 58 sur les 275 mailles de validation qui sont contenues dans cette zone urbaine et qui constituent donc l'évaluation à cette échelle pour ce site d'étude. La typologie pour cet espace comprend moins de types que pour celles de Strasbourg. En effet, seuls les espaces urbains sont considérés et la typologie est donc de cinq types. L'habitat continu, discontinu et les grandes emprises sont définis de la même manière que pour Strasbourg. Les surfaces en eau ont été obtenues automatiquement en considérant que chaque îlot composé en majorité de surfaces en eau issues de la BD TOPO® appartient à ce type. Le type autre est donc ici composé principalement par les équipements sportifs, les parcs urbains, les réseaux routiers ou ferrés et dans une moindre mesure par des îlots majoritairement composés par des parcelles agricoles.

Selon ces différents types d'occupation du sol, les performances sont assez hétérogènes (figures 5.11 et 5.13). Généralement, la faible part d'une classe d'intérêt va de pair avec une précision relativement médiocre et à l'inverse une part importante permet d'avoir des précisions meilleures. Par exemple pour Strasbourg, la végétation herbacée obtient une F-mesure très faible pour les territoires agricoles ainsi que les surfaces en eau où elle est fortement minoritaire. Pour Rennes, les précisions de cette classe sont aussi assez bien corrélées à la part qu'elle représente pour chaque type d'îlot. La classe autre qui apparaît généralement assez facile à classer, mais peut obtenir des précisions très moyennes lorsqu'elle ne représente qu'une part faible comme dans les espaces forestiers. À l'inverse, les espaces forestiers par exemple permettent d'obtenir une F-mesure élevée pour la végétation arborée. Quelques différences sont néanmoins présentes.

Pour les espaces verts artificialisés, la végétation herbacée, qui représente une part très importante de ces espaces, obtient une F-mesure assez basse. Ici, de grands patchs ont été classés en autre. C'est notamment le cas de nombreux terrains de sport enherbés dont la ressemblance avec des parcelles agricoles peut entraîner ce type d'erreur. C'est pour la classe autre que ce type d'erreur se répercute de manière importante pour le site de Rennes. Dans les espaces d'habitat continu de Strasbourg par contre, les précisions, que ce soit pour la végétation arborée ou la végétation herbacée, sont relativement bonnes, car le contraste entre ces deux classes et les autres objets est assez fort. Pour Rennes, la part que représente la végétation arborée est, pour ce type d'occupation du sol, très faible.

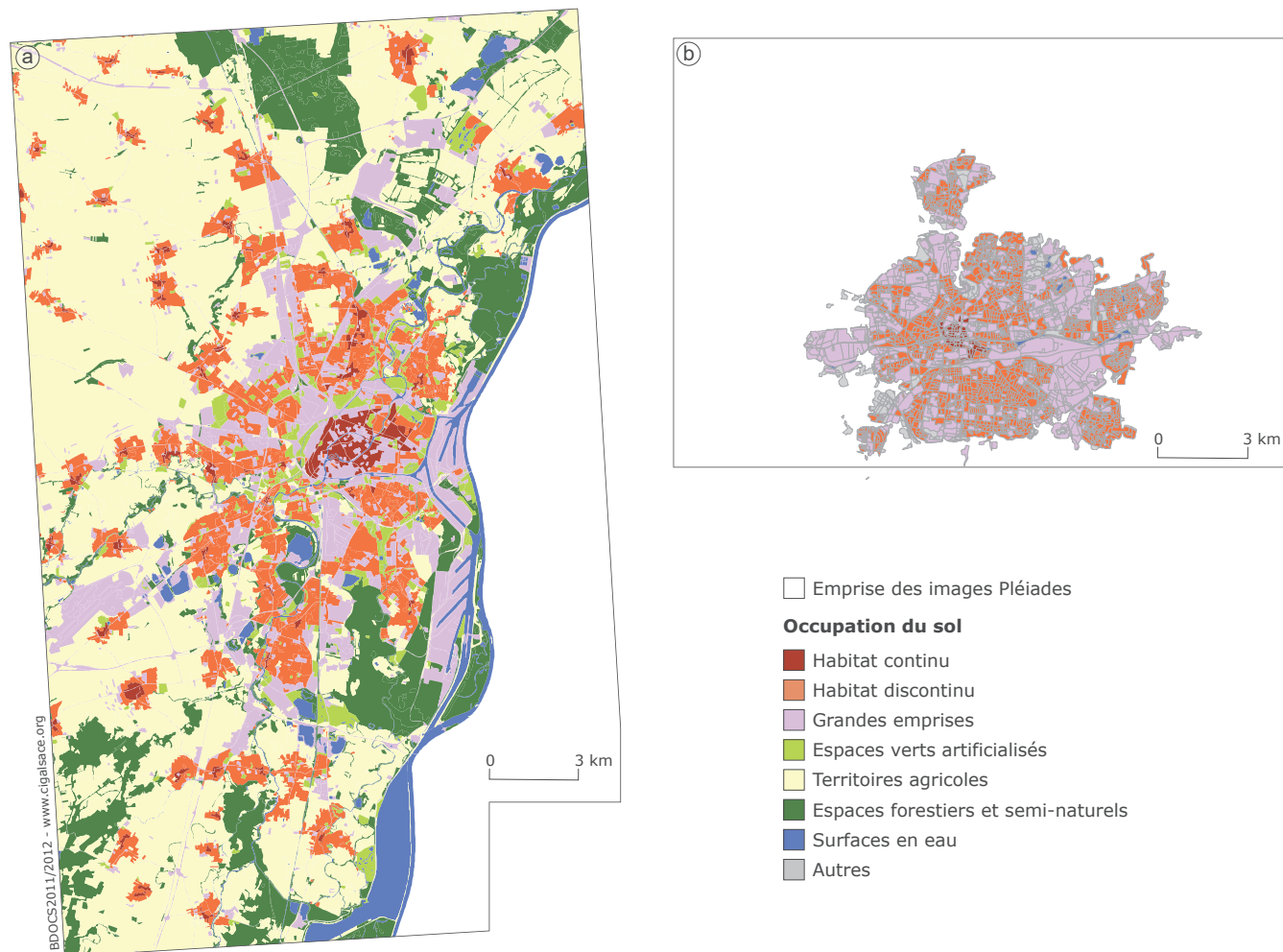
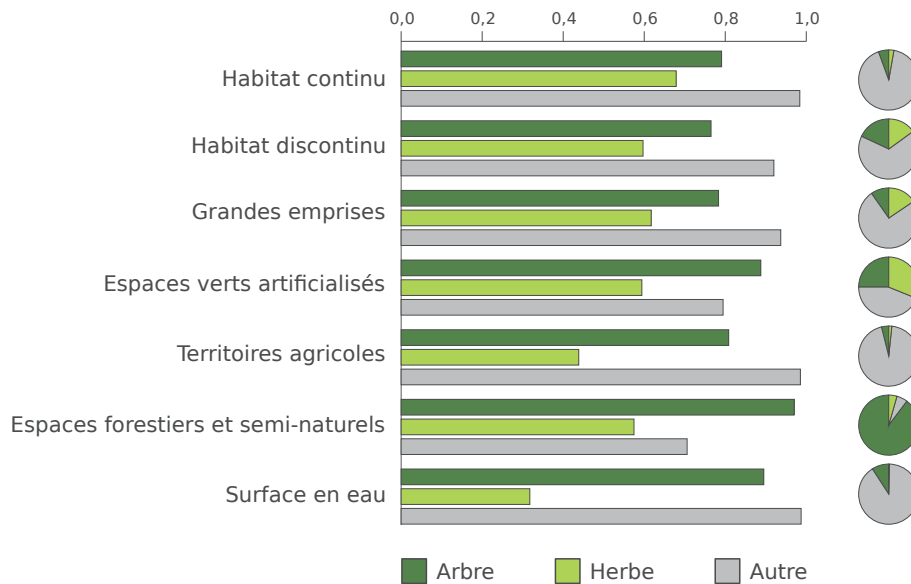
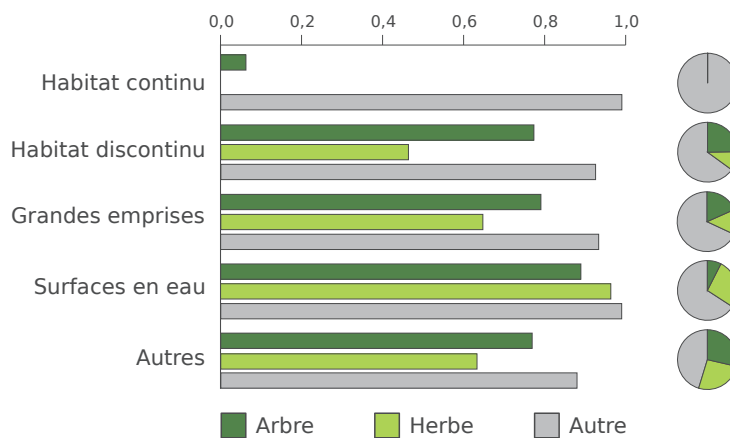


Figure 5.12 – Occupation du sol des sites de Strasbourg issue de (a) la base de données de la région Alsace (BDOCS2012 – CIGAL v2) et de (b) Rennes issue de la photo-interprétation des îlots (L. Jacquemin – Nature Par Cœur)

En effet, il n'existe pour cette classe qu'un seul élément arboré de très petite dimension et en quasi-totalité mal identifié ce qui explique la précision très faible. De plus, aucun patch de végétation herbacée n'était présent dans les îlots d'habitats continus dans nos exemples de validation ne permettant donc pas le calcul de la F-mesure.



**Figure 5.11** – Évaluation par type d'occupation du sol issue de la base de données de la région Alsace de la classification de Strasbourg par la F-mesure. Les diagrammes circulaires représentent la part de chaque classe obtenue à partir de la vérité-terrain



**Figure 5.13** – Évaluation par type d'occupation du sol issue de la photo-interprétation des îlots de la classification de Rennes par la F-mesure. Les diagrammes circulaires représentent la part de chaque classe obtenue à partir de la vérité-terrain.

Les F-mesures obtenues pour la végétation arborée sont majoritairement comprises entre 0,76 et 0,80 ce qui paraît assez faible au regard de celle obtenue précédemment de

manière globale pour le site de Strasbourg (figure 5.10) et dans une moindre mesure pour celui de Rennes (figure 5.9). En effet elles étaient proches de 0,91 et de 0,81. Les espaces forestiers couvrant une large part de nos sites d'étude, les performances obtenues sont donc positivement biaisées. À l'inverse, la végétation herbacée pour Strasbourg obtient des F-mesures proches de 0,6 excepté pour les surfaces en eau et les territoires agricoles. Les espaces agricoles représentent eux aussi une part très importante de l'agglomération et ont donc une forte influence sur les résultats obtenus pour cette classe. Pour Rennes, la végétation herbacée obtient des F-mesures meilleures que celle calculée globalement notamment en raison de la non-représentation des espaces agricoles dans cette typologie.

### 5.4.3 Évaluation à l'échelle des objets

La dernière évaluation se fait ici à l'échelle des objets résultants de la classification. Seuls ceux de végétation arborée et herbacée sont évalués grâce à la méthode proposée par Ardila *et al.* (2012). Celle-ci compare l'adéquation entre les objets de la classification et ceux de la vérité terrain en utilisant des méthodes issues de l'évaluation supervisée des segmentations (chapitre 3).

Tout d'abord, faux positifs et faux négatifs sont extraits en considérant chaque classe indépendamment. Par exemple, un arbre appartenant à la vérité-terrain qui n'intersecte pas un arbre de la classification est considéré comme un faux négatif, peut importe s' il se superpose à un objet de classe herbe ou autre.

Tous les autres objets, ceux qui ne sont ni des faux négatifs ni des faux positifs, sont alors évalués selon une sur ( $OI$ ) et une sous-identification ( $UI$ ) qui peuvent être calculées grâce aux formules :

$$OI = 1 - \frac{Aire(O \cap R)}{Aire(O)} \quad (5.1)$$

$$UI = 1 - \frac{Aire(O \cap R)}{Aire(R)} \quad (5.2)$$

où  $O$  correspond à l'objet de la classification et  $R$  à celui de référence. Dans le cas où un objet de la classification est décrit par plusieurs de la vérité-terrain, alors les moyennes de la sous et de la sur-identification sont calculées. Ces deux mesures peuvent ensuite être combinées afin d'obtenir l'erreur totale  $ET$  de délimitation de chaque objet par :

$$ET = \sqrt{\frac{OI^2 + UI^2}{2}} \quad (5.3)$$

Ces trois mesures sont comprises entre 0 et 1 où une adéquation parfaite entre un objet issu de la classification et son correspondant dans la vérité-terrain serait égale à 0.

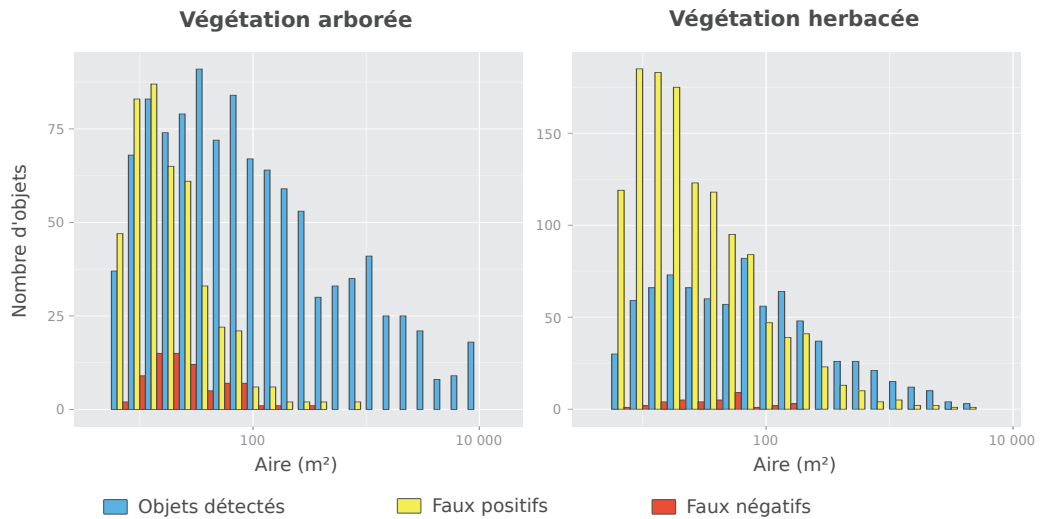
Les résultats obtenus avec cette méthode montrent que le nombre de faux positifs est très important particulièrement pour la végétation herbacée puisqu'ils représentent 61 % du nombre total d'objets herbe issus de la classification pour Rennes (figure 5.14) et 44 % pour Strasbourg (figure 5.15). Aussi bien pour la végétation arborée qu'herbacée, ces objets sont majoritairement de petites tailles. Les faux positifs de la classe herbe sont très souvent situés dans l'ombre des arbres (figure 5.17.3) ou dans leur partie la plus claire (figure 5.17.2) pour ceux de plus petite dimension, mais aussi sur des parcelles agricoles pour les plus grands. Le post-traitement consistant à supprimer les parcelles grâce au RPG n'est pas parfait puisque certaines bordures de parcelle n'ont pas été éliminées ainsi que d'autres qui n'étaient pas représentées dans cette base (figure 5.17.4). Les faux négatifs sont par contre beaucoup moins nombreux pour cette classe et sont souvent liés à des confusions avec la classe arbre sur des objets de petites dimensions (figure 5.18.2), mais aussi comme il a été mentionné lors de la section précédente, sur de plus grands patches tels que des terrains de sport.

Le nombre de faux positifs pour la végétation arborée est plus faible même si de nombreuses erreurs subsistent pour nos deux sites d'étude. En effet, en plus des confusions avec l'herbe, des objets situés dans l'ombre des bâtiments où le MSAVI peut être élevé du fait de la présence de végétation non visible lors de la photo-interprétation (figure 5.18.4). Cette erreur souvent rencontrée est liée à la ressemblance de ces segments avec ceux d'ombre présents au sein de massifs boisés qui ont eux été intégrés à la végétation arborée lors de l'apprentissage. Les faux négatifs sont beaucoup plus présents pour cette classe. En majorité de petites dimensions, ils sont concentrés dans des espaces où les arbres isolés sont très fréquents tels que les jardins des lotissements (figure 5.18.2) ou le long des voies de circulation (figure 5.17.3). De plus, il peut apparaître localement des confusions entre végétation arborée et herbacée, traduites par des faux négatif et positif, qui ne sont pas des erreurs de classification (figure 5.18.2). C'est ici principalement des erreurs liées à la photo-interprétation de l'image qui montre la difficulté de réalisation de cette tâche.

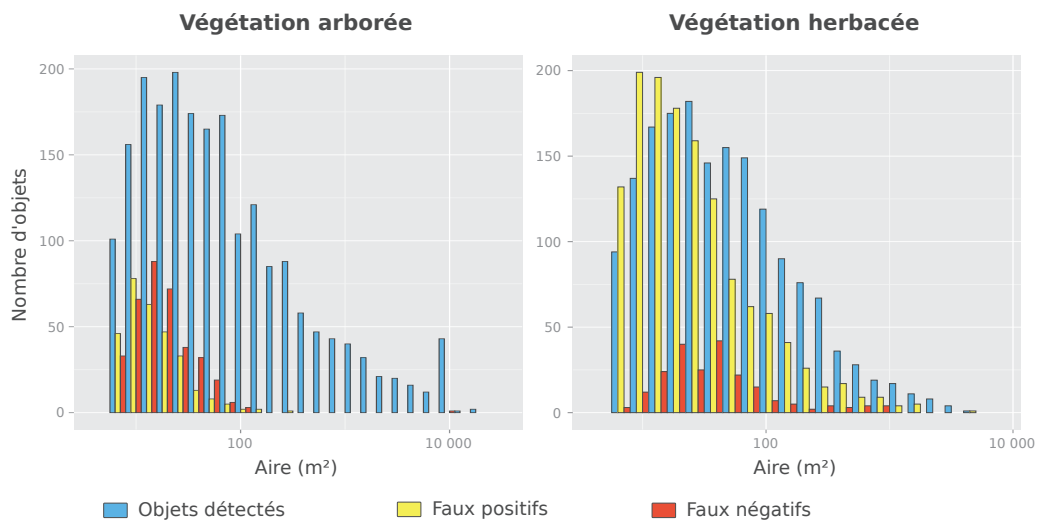
Les erreurs de délimitations mesurées entre les objets détectés et ceux de la vérité terrain sont assez élevées (figure 5.16.a). En effet les médianes pour l'herbe sont de 0,59. Pour la végétation arborée, ces erreurs sont plus faibles autant pour Rennes où la médiane de l'erreur totale est de 0,48 que pour Strasbourg où elle est de 0,37. Ces erreurs importantes peuvent expliquer en partie les F-mesures assez faibles obtenues pour la végétation herbacée. Ces F-mesures ont en effet été obtenues à partir de la matrice de confusions réalisée par l'intersection puis le calcul de l'aire des objets ainsi créés.

L'erreur totale est en grande partie due à une sous-identification des objets (figure 5.16.b). Cet effet est particulièrement important pour la végétation herbacée de Strasbourg. Les fonctions de densité de ces erreurs de sur et de sous-identification montrent bien que presque tous les objets herbe pour ce site connaissent très peu de problèmes de sur-identification tandis qu'une majorité est concernée par ceux de sous-identification. Pour Rennes, cet effet est moins présent puisqu'une part conséquente d'objets connaît aussi des défauts de sur-identification. Cette erreur de délimitation est assez marquée

visuellement sur les résultats des classifications des deux sites et peut s'expliquer par un morcellement de patch d'herbe de la vérité terrain en plusieurs objets résultants de la classification qui maximise les erreurs de sous-identification pour ces multiples objets (figures 5.17.4 et 5.18.4).

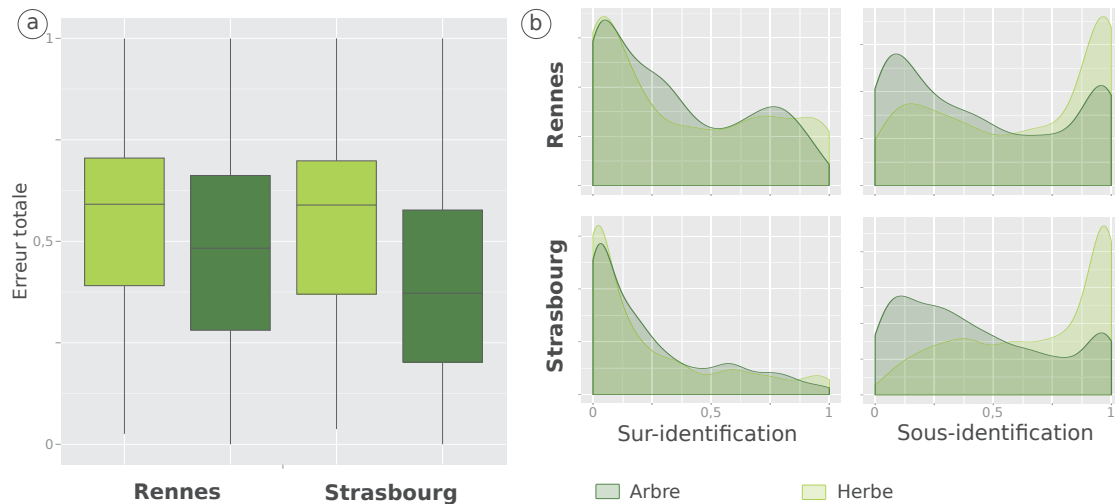


**Figure 5.14** – Distribution pour Rennes des objets détectés, des faux négatifs et positifs selon l'aire des objets (selon une échelle logarithmique). L'aire considérée est celle des objets de la vérité terrain pour les faux négatifs et celle des objets de la classification pour les objets détectés et les faux positifs



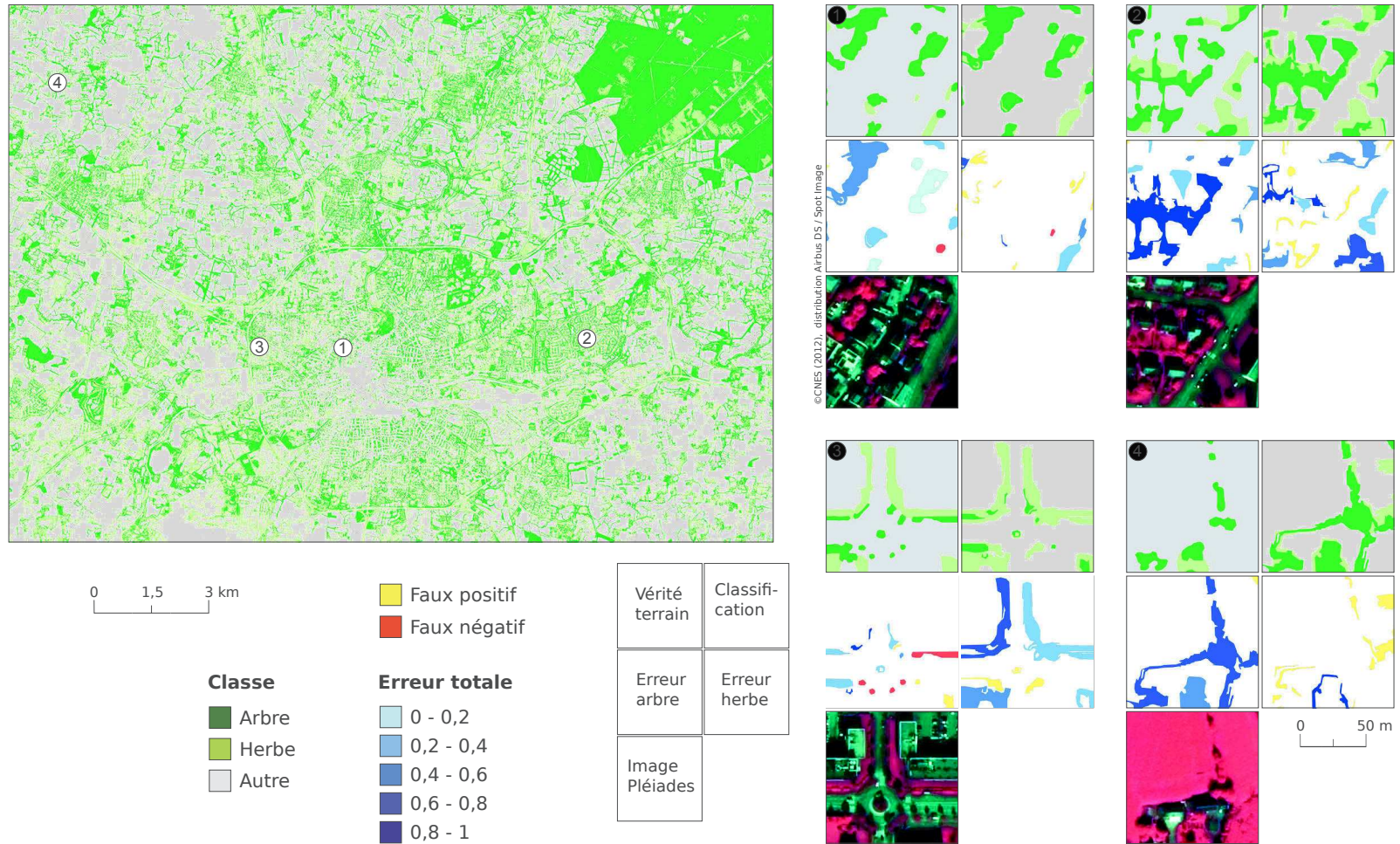
**Figure 5.15** – Distribution pour Strasbourg des objets détectés, des faux négatifs et positifs selon l'aire des objets (selon une échelle logarithmique). L'aire considérée est celle des objets de la vérité terrain pour les faux négatifs et celle des objets de la classification pour les objets détectés et les faux positifs



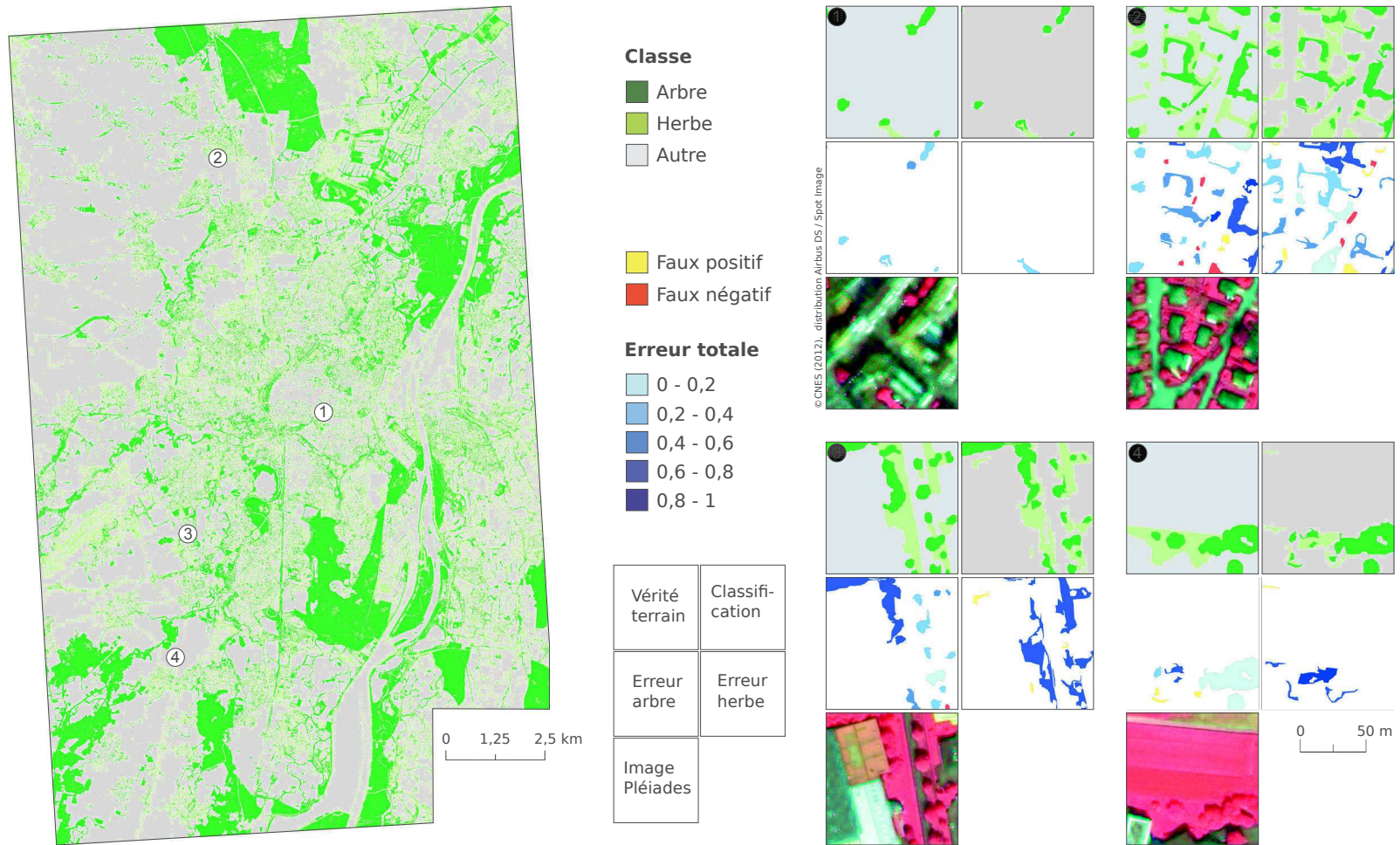


**Figure 5.16** – Distribution de l'erreur totale  $ET$  pour les objets détectés de Rennes et de Strasbourg (a) et distribution des fréquences de sur et de sous-identification par classe de végétation (b)

Les erreurs concernant la végétation arborée sont beaucoup moins nombreuses que pour la végétation herbacée. Une majorité des objets identifiés sont en effet caractérisés par une sur et une sous-identification inférieure à 0,5 avec une tendance à être plutôt sous-identifiés. Comme précédemment, cette classe peut aussi connaître un morcellement des objets définis par la vérité terrain, mais dans une moindre mesure (figures 5.17.2 et 5.18.2). Pour Rennes un phénomène plus important de sur-identification peut-être observé. On assiste ici à l'effet inverse, c'est-à-dire que certains objets issus de la classification vont être composé de plusieurs issus de la vérité terrain (figures 5.17.1-4 et 5.18.3).



**Figure 5.17** – Classification post-traitée du site d'étude de Rennes et exemples de quatre mailles de validation. Ces exemples sont décrits par la vérité terrain, le résultat de la classification, l'image ainsi que les erreurs pour les deux classes de végétation. Chaque classe est évaluée selon un schéma un contre tous



**Figure 5.18** – Classification post-traitée du site d'étude de Strasbourg et exemples de quatre mailles de validation. Ces exemples sont décrits par la vérité terrain, le résultat de la classification, l'image ainsi que les erreurs pour les deux classes de végétation. Chaque classe est évaluée selon un schéma un contre tous

## Conclusion

Les méthodes proposées ont permis d'extraire la végétation arborée et herbacée pour les zones d'études de Strasbourg et de Rennes à partir d'images satellites THRS Pléiades. Dans un premier temps, les méthodes de segmentation d'image et de sélection d'échantillons d'apprentissage évaluées lors des deux chapitres précédents et réalisées sur des zones tests ont été généralisées à l'ensemble des zones d'études. Pour rendre ces opérations possibles sur de larges espaces, ces méthodes ont donc été adaptées. La réalisation d'un tuilage a permis d'effectuer l'ensemble des étapes réalisées lors de ces deux chapitres.

La seconde partie de ce chapitre s'est concentrée sur la réalisation des classifications finales pour les deux sites, suivi de leur post-traitement et de leur validation. Quatre algorithmes d'apprentissage automatique ont ainsi été paramétrés et évalués : k-NN, SVM, RF et C 5.0. SVM s'est révélé être le plus performant pour le site de Strasbourg tandis que pour celui de Rennes, les meilleures performances ont été obtenues avec RF. Le protocole de post-traitements mis en place se déclinait en trois étapes importantes à savoir l'agrégation des segments adjacents appartenant à la même classe, la suppression des objets herbe situés sur des parcelles agricoles et la suppression des objets d'une dimension inférieurs à  $6,5 \text{ m}^2$ . Il a permis une amélioration significative des résultats, notamment pour le site de Rennes. En effet, de nombreuses confusions entre végétation herbacée et parcelles agricoles étaient présentes alors que pour le site de Strasbourg ces confusions étaient moins nombreuses grâce à l'utilisation d'attributs multi-temporels.

Les performances obtenues pour les deux classes de végétation peuvent être très différentes selon le degré d'artificialisation ou le type d'utilisation du sol. La végétation arborée est généralement bien identifiée dans les espaces les moins artificialisés tandis que l'inverse apparaît pour la végétation herbacée. L'analyse de ces confusions a montré que la part que représentent ces classes dans ces différentes strates ou utilisations du sol en est certainement une des causes. À l'échelle des objets, l'évaluation a révélé que les objets de petites dimensions, qu'ils soient arborés ou herbacés, sont les plus mal détectés pour les deux sites d'études et que la sous-identification des objets est un problème plus important que la sur-identification. Au final, les bases de données créées se révèlent être très performantes pour la végétation arborée où les F-mesures obtenues sont de 91 % pour Strasbourg et de 81 % pour Rennes tandis que pour la végétation herbacée les résultats sont de relativement mauvaise qualité puisque les F-mesures obtenues sont d'environ 54 %.

## Conclusion de la partie II

À partir d'images Pléiades, cette partie a permis la création d'une base de données décrivant la végétation qui faisait défaut pour les agglomérations de Strasbourg et de Rennes en deux strates : la végétation arborée et la végétation herbacée. Une méthodologie qui se base sur le paradigme de l'analyse d'image satellite orientée-objet associée à une classification supervisée a été mise en place tout au long de cette partie afin de proposer plusieurs solutions aux défis liés à ce type d'approche : le choix d'une segmentation adaptée, la sélection d'échantillons pour l'entraînement et la classification sur de grandes images.

Le chapitre 3 avait pour objectif de répondre à ce premier défi. Deux algorithmes de segmentation, Multi-Resolution et Mean Shift, ont été évalués par des méthodes supervisées, non supervisées et par classification. L'évaluation par classification était utilisée comme référence afin de vérifier si l'une des deux autres pouvait être utilisée pour sélectionner la segmentation optimale du fait qu'elles sont moins coûteuses en temps de traitements. Les résultats ont montré que les segmentations sélectionnées par ces trois méthodes étaient très différentes rendant ainsi difficile le choix d'une segmentation par une méthode supervisée ou non si l'objectif est d'obtenir de bonnes performances pour l'extraction de la végétation urbaine. Ce chapitre a permis de confirmer qu'une sur-segmentation permettait d'obtenir les meilleurs résultats et a montré que l'algorithme Multi-Resolution était le plus simple à utiliser du fait de performances relativement stables, quelle que soit l'échelle de segmentation sélectionnée.

Le chapitre 4 avait pour objectif de proposer des méthodes d'échantillonnage par apprentissage actif afin de constituer un jeu d'entraînement divers et informatif. Trois méthodes différentes ont été développées et comparées à une méthode classique d'échantillonnage stratifié. Ce chapitre a permis de montrer que la combinaison de stratégies d'échantillonnage active et stratifiée permettait d'obtenir les meilleures performances en sélectionnant un nombre limité d'exemples d'entraînement. Toutefois, certains problèmes inhérents à la méthode sont apparus comme le passage d'une représentation vectorielle des segments à une représentation matricielle par leur centre de gravité qui entraînait la création de pixels vides et qui perturbait le choix de nouvelles régions. Une des améliorations nécessaires à cette méthode serait donc de conserver la forme vectorielle de la segmentation. Ce chapitre a aussi permis d'évaluer l'apport de données

auxiliaires pour le site de Strasbourg permettant de montrer l'importance des informations de hauteurs pour la classification de la végétation urbaine. L'acquisition de ces informations peut être difficile à obtenir, mais peut être surmontée grâce aux capacités des satellites Pléiades à prendre des images stéréoscopiques.

Enfin le chapitre 5 avait pour objectifs de généraliser à l'ensemble des images Pléiades de Strasbourg et de Rennes la segmentation et le calcul des attributs ainsi que la sélection des échantillons par apprentissage actif. Toutefois, cette généralisation s'est heurtée au problème de la taille des images et une stratégie par tuilage a été mise en place afin de rendre possible l'application de la méthodologie développée lors des deux chapitres précédents. De plus, ce chapitre a permis de tester différents algorithmes de classification afin d'améliorer les performances de l'extraction de la végétation arborée et herbacée. Enfin un protocole de post-traitements a été appliqué qui a permis d'améliorer fortement les précisions des classifications principalement pour la végétation herbacée. Les classifications finales des deux sites d'études ont pu être finement évaluées et ont montré que l'extraction de la végétation arborée est performante grâce à cette méthodologie. Celles pour l'extraction de la végétation herbacée sont par contre beaucoup moins bonnes et il serait donc nécessaire de considérer d'autres attributs ou d'autres données auxiliaires pour permettre de les améliorer.

La constitution de ces bases de données végétation était une étape préalable nécessaire pour l'analyse de la structure urbaine par la combinaison de la trame grise et la trame verte du fait qu'aucune base de données existante ne permettait d'obtenir une cartographie exhaustive des éléments verts. La mise en place de cette méthodologie pour l'extraction de la végétation à partir d'image satellite THRS a permis d'obtenir cette cartographie et est intéressante, car elle permet d'avoir une représentation multi-strates de celle-ci. Plus spécifiquement pour la végétation arborée, elle permet d'obtenir plus de précision que la BD TOPO® avec une surface minimum de cartographie beaucoup plus petite. De plus, grâce à la capacité de revisite des satellites Pléiades, cette base de données végétation pourrait être mise à jour très régulièrement. Toutefois, une implémentation plus efficiente devrait être mise en place en automatisant le processus de tuilage qui était ici réalisé manuellement et en créant une interface pour la labélisation des exemples d'entraînement sélectionnés par apprentissage actif qui éviterait les aller et retour entre le logiciel SIG et celui de traitement. Grâce à la construction de cette base de données géographique à grande échelle, la prochaine partie de cette thèse va pouvoir se consacrer à cartographier la structure urbaine des agglomérations de Strasbourg et de Rennes à l'échelle des tissus.







## Partie III

### Cartographie et analyse quantitative de la structure urbaine à partir de données multi-sources

Introduction de la partie III .....	169
<b>CHAPITRE 6</b> : Création d'une base de données d'éléments urbains multi-sources ..	171
<b>CHAPITRE 7</b> : Application de méthodes d'extraction de connaissances .....	203
Conclusion de la partie III .....	253



## Introduction de la partie III

Cette partie vise à proposer une méthodologie pour la cartographie la plus automatique possible et l'analyse quantitative des tissus urbains à une échelle du 1/5000<sup>e</sup> au 1/10 000<sup>e</sup> à partir d'une base de données d'occupation du sol à grande échelle fondée sur des données existantes pour la trame grise et fondée sur la trame verte extraite dans la partie II.

Pour rappel, la notion de tissu fait référence à une portion d'espace qui nécessite d'être délimité afin de rendre compte de l'organisation des éléments gris et verts qui le composent. De très nombreux découpages existent dans la littérature (§ 2.3.3.1) et il est donc indispensable d'identifier l'unité d'analyse la plus adaptée pour automatiser leur cartographie. Par ailleurs, la définition des tissus urbains ne peut être uniquement basée sur l'occupation du sol. Afin d'élargir leur analyse quantitative au-delà de la morphologie urbaine, d'autres indicateurs et des données auxiliaires par exemple statistiques sont ainsi proposés et utilisés. Dans ce contexte, il s'agit également d'identifier les plus pertinents.

Le premier chapitre de cette partie a ainsi pour premier objectif de constituer cette base de données d'occupation des sols adaptée afin d'analyser les éléments constitutifs des trames urbaines grise et verte. Cette base de données est construite à partir de sources différentes adaptées pour la cartographie des espaces urbains. À partir de celle-ci, une série d'indicateurs selon leur discipline d'origine tels que l'urbanisme ou l'écologie du paysage sont calculés. D'autres indicateurs originaux permettant de tenir compte du contexte spatial sont aussi proposés. Enfin l'unité d'analyse la plus adaptée pour la cartographie des tissus urbains est évaluée et un maillage adapté est proposé.

Le deuxième chapitre de cette partie propose une méthodologie d'extraction des connaissances afin d'automatiser le processus de cartographie des tissus urbains, fondée sur une trame grise et une trame verte, adaptée et constituée dans la base d'occupation du sol construite dans le chapitre précédent. Suite à l'analyse de l'ensemble des indicateurs calculés pour le maillage retenu précédemment, la cartographie des tissus est réalisée à l'aide d'une classification non supervisée (clustering) qui est ensuite interprétée et évaluée.

L'ensemble de la démarche mise en place dans cette partie est testé et développé sur l'agglomération de Strasbourg avant d'être validé sur celle de Rennes afin de vérifier la

reproductibilité de la méthode. La figure E présente les étapes de la démarche méthodologique de cette partie.

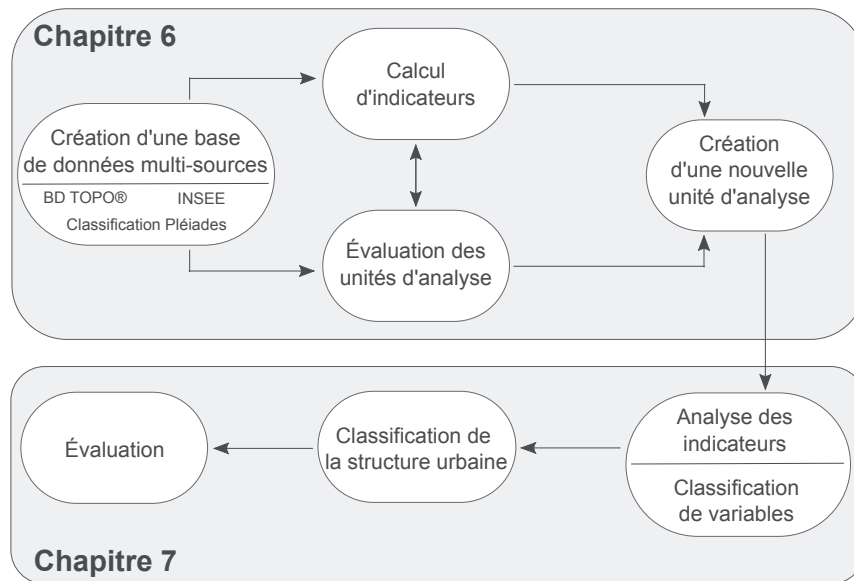


Figure E – Structuration de la démarche méthodologique suivie dans la partie I

# Chapitre 6

## Création d'une base de données d'éléments urbains multi-sources

### Sommaire

---

Introduction . . . . .	172
6.1 Vers une base de données d'occupation du sol à grande échelle . . . . .	172
6.1.1 Choix des données pour cartographier la trame grise . . . . .	172
6.1.2 Choix des données pour cartographier les trames verte et bleue . . . . .	174
6.1.3 Les données INSEE sur la population . . . . .	175
6.2 La caractérisation de la structure urbaine . . . . .	176
6.2.1 Indicateurs issus de l'urbanisme . . . . .	176
6.2.2 Indicateurs issus de l'écologie du paysage . . . . .	180
6.2.3 Autres indicateurs proposés dans cette recherche . . . . .	182
6.3 Unités testées pour caractériser la structure urbaine . . . . .	188
6.3.1 Découpage par carroyage . . . . .	188
6.3.2 Découpage à la parcelle morphologique . . . . .	189
6.3.3 Découpage à l'îlot . . . . .	190
6.4 Sélection de l'unité d'analyse pour la cartographie de la structure urbaine . . . . .	193
6.4.1 Sélection des critères d'évaluation . . . . .	193
6.4.2 Évaluation des mailles pour l'analyse de la structure urbaine . . . . .	195
6.5 Proposition d'un nouveau maillage pour la cartographie de la structure urbaine . . . . .	198
Conclusion . . . . .	202

---

## Introduction

Ce chapitre présente dans un premier temps les différentes bases de données qui sont utilisées pour calculer toute une série d'indicateurs afin de cartographier les structures urbaines de Strasbourg et de Rennes (§ 6.1). Afin de décrire de manière la plus exhaustive possible cette structure, un grand nombre d'indicateurs provenant de plusieurs disciplines sont considérés et sont présentés dans la section 6.2 de ce chapitre. Dans un troisième temps, nous présentons les différentes unités d'analyse proposées et testées dans cette thèse afin de caractériser les espaces urbains et périurbains (§ 6.3). Enfin la section 6.4 a pour objectif d'évaluer ces unités et de déterminer laquelle est la plus adaptée pour la caractérisation de la structure urbaine.

### 6.1 Vers une base de données d'occupation du sol à grande échelle

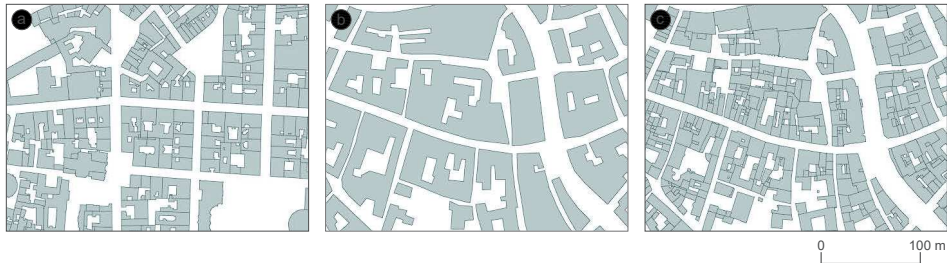
Pour cartographier la structure urbaine, il est indispensable de disposer d'une cartographie à grande échelle des éléments constituant l'occupation du sol qui constituent de la trame grise et de la trame verte. Les données permettant de décrire les éléments gris sont tout d'abord exposées (§ 6.1.1) puis celles utilisées pour la description des éléments verts et bleus (§ 6.1.2). La section 6.1.3 présente les données auxiliaires (données INSEE) utilisées pour compléter l'analyse morphologique de la structure urbaine et la modélisation choisie afin de pouvoir intégrer ces données dans notre analyse.

#### 6.1.1 Choix des données pour cartographier la trame grise

De nombreuses bases de données permettent de décrire les éléments gris des milieux urbains à une échelle fine du 1/2000<sup>e</sup> au 1/5000<sup>e</sup> (§ 2.2.2). Afin d'assurer la reproductibilité de l'approche permettant de décrire les tissus urbains, le choix des bases de données pour les éléments gris s'est porté vers des bases de données nationales qui sont plus exhaustives et de qualité plus homogène pour l'ensemble de la France. Ainsi nous nous appuyons sur deux types de bases, la base de données topographique de l'IGN (BD TOPO®) et une base de données cadastrale pour Strasbourg celle des bâtiments du cadastre numérisé fourni par l'Eurométropole de Strasbourg.

La BD TOPO® est la base de référence de l'IGN qui décrit avec une précision métrique différents éléments du paysage en intégrant la dimension verticale obtenue par photogrammétrie. Cette base est décomposée en dix thèmes (IGN, 2014b) ; trois sont utilisés dans cette thèse afin de décrire les éléments gris. Les deux premiers sont les réseaux routiers et ferrés, le troisième est le bâti. Plusieurs couches décrivant les bâtiments sont disponibles et trois sont utilisés : le bâti indifférencié, le bâti remarquable et le bâti industriel. La source des bâtiments peut être différente et provenir soit de la BD TOPO®

et être obtenus par photogrammétrie ou être issus du cadastre. Ainsi pour la zone d'étude de Rennes les deux types de représentation sont présents dans la BD TOPO®, les bâtiments issus du cadastre représentent 99 % de l'ensemble tandis que pour Strasbourg l'ensemble des bâtiments provient de la BD TOPO®. La différence entre les sources des données offre une représentation différente des espaces urbains (figure 6.1).



**Figure 6.1** – Exemples de bâtiments du centre-ville de Rennes issus (a) de la BD TOPO® (2015) et de Strasbourg issus (b) de la BD TOPO® (2014) et (c) du cadastre (2013)

Pour Strasbourg les deux modes de représentation ont été conservés, car elles permettent une représentation en deux niveaux qui peut être un marqueur des formes urbaines tel que dans les centres urbains où les bâtiments de la BD TOPO® apparaissent de grandes dimensions, mais fortement découpés par ceux du cadastre (figure 6.1). Pour le site de Rennes, les bâtiments contigus ont donc été fusionnés afin de permettre cette double représentation. Nous parlons dans la suite de ce travail de bâtiments agrégés et de bâtiments non agrégés ou du cadastre. D'autres prétraitements ont été réalisés sur ces éléments. Tout d'abord, les trois couches de la BD TOPO® ont été fusionnées sans que soit conservée leur classe (indifférencié, remarquable, industriel). Une évaluation a été réalisée par croisement de ces couches et de la BDOCS de 2012 sur le site de Strasbourg et a révélé que 24 % des bâtiments sont mal classés.

De plus les informations de hauteur ont été conservées ou attribuées différemment selon les sites d'étude. Dans la BD TOPO® la hauteur des bâtiments est renseignée, mais peut être manquante. De plus dans la base de données cadastrale de l'Eurométropole, cette information n'est pas disponible. Pour ce site nous disposons d'un MNE (§ 4.3) et la hauteur a pu être renseignée grâce à cette donnée pour les deux niveaux de bâtiments. Pour Rennes, seules les hauteurs de la BD TOPO® sont disponibles et pour le niveau agrégé des bâtiments, celle-ci a été attribuée par la moyenne au prorata de la surface. Que ce soit pour Rennes ou pour Strasbourg certaines hauteurs de bâtiment peuvent être erronées ou manquantes. Ainsi pour Strasbourg lorsque les hauteurs étaient inférieures à 1 m (1,6 % des bâtiments) et pour Rennes lorsqu'elles étaient manquantes (17,6 % des bâtiments), une hauteur de 1 m a été attribuée afin de rendre possible le calcul de certains attributs dans la suite de ce chapitre.

Pour les deux sites d'étude, le réseau routier de la BD TOPO® est utilisé. Cette base distingue plusieurs niveaux qui permettent de le hiérarchiser selon l'importance des

tronçons pour le trafic routier (IGN, 2014b). Cinq niveaux d'importance sont présents dans cette base :

1. Liaisons les plus importantes composées principalement d'autoroutes et de quasi-autoroutes qui permettent la liaison entre les métropoles
2. Liaisons entre départements avec un fort trafic
3. Liaisons interurbaines à l'intérieur d'un département
4. Voies permettant les déplacements rapides intra-urbains ainsi que les liaisons entre bourg et hameau dans les espaces ruraux
5. Les voies permettant de desservir l'intérieur d'une commune

Enfin, un dernier niveau existe (NC) qui répertorie tous les autres tronçons ne faisant pas partie de ces catégories et qui est principalement composé de chemins.

La même base de données est utilisée pour la représentation du réseau ferré qui est lui aussi discrétisé en plusieurs catégories :

- Les lignes à grande vitesse
- Les voies ferrées principales
- Les voies de service
- Les voies non exploitées
- Le transport urbain (tramway, métro aérien)
- Funiculaire ou crémaillère

Le tableau 6.1 présente pour chaque élément gris et pour chaque site la base de donnée utilisée ainsi que sa date de production.

**Tableau 6.1** – Bases de données utilisées pour la cartographie des éléments gris

Élément gris	Strasbourg	Rennes
Bâtiments non agrégés	Cadastre – Eurométropole de Strasbourg – 2013	BD TOPO® – IGN – 2015
Bâtiments agrégés	BD TOPO® – IGN – 2014	BD TOPO® – IGN – 2015
Réseau routier	BD TOPO® – IGN – 2014	BD TOPO® – IGN – 2015
Réseau ferré	BD TOPO® – IGN – 2014	BD TOPO® – IGN – 2015

### 6.1.2 Choix des données pour cartographier les trames verte et bleue

Pour la cartographie des trames vertes et bleues, plusieurs éléments sont pris en compte : la végétation arborée et herbacée, les parcelles agricoles et les surfaces en eau. La description de ces éléments dans nos deux agglomérations s'appuie sur différentes bases de données. La plupart de ces bases ont déjà été présentées précédemment (§ 2.2.2 et 5.3.3) et sont donc ici seulement mentionnées. La première est la couche végétation issue de la classification d'images Pléiades en deux classes : la végétation arborée et la végétation herbacée. Nous intégrons aussi aux éléments verts les différentes parcelles agricoles issues du RPG de 2012 qui ont été utilisées dans la chaîne de post-traitements de la classification des images (§ 5.3.3).



La base de données permettant d'obtenir les éléments décrivant le réseau hydrographique et les surfaces en eau est la BD TOPO®. La couche surface en eau décrit toutes les surfaces dont la largeur est supérieure à 20 m et les cours d'eau faisant plus de 7,5 m de large. Le tableau 6.2 résume les différentes bases de données utilisées pour la cartographie des trames verte et bleue.

**Tableau 6.2** – Bases de données utilisées pour la cartographie des éléments verts et bleus

Éléments verts et bleus	Strasbourg	Rennes
Arbre	Classification Pléiades 2012/2013	Classification Pléiades 2012
Herbe	Classification Pléiades 2012/2013	Classification Pléiades 2012
Parcelle agricole	RPG 2012 – ASP	RPG 2012 – ASP
Surface en eau	BD TOPO® – IGN – 2014	BD TOPO® – IGN – 2015

La figure 6.2 présente le résultat de la base de données géographique qui permet de décrire les éléments des trames grise, verte et bleu de l'agglomération de Strasbourg et la figure 6.3 présente celle réalisée pour le site de Rennes.

### 6.1.3 Les données INSEE sur la population

Les IRIS sont le découpage le plus fin où est diffusé le recensement de la population par l'INSEE et sont souvent utilisés pour la cartographie de la structure urbaine (§ 2.3.3). Ces données, continues sur l'ensemble du territoire français, peuvent être désagrégées selon l'objet utilisé pour reconstruire le tissu urbain, par exemple à la maille (Bonhomme, 2013) ou au bâtiment (Sahraoui *et al.*, 2014), afin d'obtenir une répartition plus précise de la démographie. Toutefois, un problème méthodologique se pose avec ces données. En effet, les données démographiques sont dans des zones géographiques prédéfinies, dont certaines sont parfois non peuplées (Weeks, 2001). Les IRIS par exemple sont continus sur l'ensemble du territoire et doivent avoir une taille suffisamment large afin de faire disparaître les problèmes de confidentialité. Ainsi lors de la désagrégation de ce type de données, des problèmes peuvent survenir, par exemple à l'échelle de la maille où certaines auront une population sans qu'aucun bâtiment n'y soit présent, ou du bâtiment, où tous seront peuplés sans qu'il s'agisse forcément de bâtiments résidentiels.

L'INSEE a récemment mis en ligne des données carroyées sur la population avec des mailles de 200 × 200 m. Cette base présente l'avantage de ne pas être continue sur l'ensemble du territoire permettant de seulement représenter les carreaux peuplés à la différence des IRIS. Cette base décrit les revenus fiscaux localisés qui sont établis à partir de la déclaration de revenus des personnes physiques et de la taxe d'habitation. La localisation des habitants se fait donc à l'adresse puis est agrégée au sein des mailles. Elle est constituée de 18 variables permettant de décrire la structure par âge des individus, les caractéristiques des ménages et leurs revenus de 2010. Pour respecter la confidentialité des ménages, seul le nombre d'habitants par carreaux est donné et les carreaux dont le nombre de ménages est inférieur à 11 sont agrégés en rectangles afin de répondre à cette règle. Il est donc nécessaire, pour plus de précision, de désagréger ces données à

l'échelle du carreau. Nous avons suivi ici la méthode fournie par l'INSEE (INSEE, 2013). Ce carroyage a récemment été utilisé par Baro *et al.* (2016) afin de caractériser les tissus urbains dans un découpage radioconcentrique du centre vers la périphérie. Le nombre d'habitants, associé à la surface totale et moyenne des bâtiments de la BD TOPO®, a permis d'obtenir une typologie des tissus urbains des villes de Besançon, Tours et Nancy par seuillages successifs afin de construire un arbre de décision.

En conclusion, maintenant que nous disposons d'une base de données d'occupation des sols complétée par des données sociodémographiques, l'étape suivante consiste à dériver de ces bases des indicateurs permettant de caractériser la structure urbaine c'est-à-dire de qualifier quantitativement ces éléments, leur morphologie ainsi que leur organisation spatiale.

## 6.2 La caractérisation de la structure urbaine

Afin de décrire la structure urbaine, de très nombreux indicateurs ont été développés par différentes disciplines de recherche. Clifton *et al.* (2008) classent ces indicateurs en cinq catégories d'approches permettant de quantifier ces formes à différentes échelles et avec différents objectifs : l'écologie du paysage, la structure économique, la planification des transports, le design communautaire et le design urbain.

À partir de la base de données constituée dans la section 6.1, nous formulons l'hypothèse que les indicateurs souvent utilisés en urbanisme associés à ceux de l'écologie du paysage permettent de caractériser les formes de la trame grise et des trames vertes et bleues. Toutefois afin d'obtenir une cartographie des tissus urbains la plus complète et opérationnelle possible (pas uniquement morphologique), nous proposons de les compléter par d'autres indicateurs permettant de rendre compte de leurs caractéristiques sociodémographiques, mais aussi des indicateurs d'accessibilité, de visibilité et de centralité.

### 6.2.1 Indicateurs issus de l'urbanisme

En urbanisme, de nombreux indicateurs ont été proposés afin de caractériser les formes urbaines et leur choix est guidé par la typologie qui est recherchée (§ 1.2.3). La densité est un critère très souvent utilisé pour comparer et cartographier les tissus urbains (AUCAME, 2010 ; Berghauser Pont et Haupt, 2005 ; Salat, 2011). Par exemple, Schneider et Woodcock (2008), dont l'objectif est de comparer les formes de développement urbain de plusieurs grandes agglomérations dans le monde, vont utiliser des métriques permettant de quantifier les changements entre deux dates en terme de densité de bâti et de population, mais aussi de surface construite et de fragmentation du paysage, pour ces dernières, à partir de métriques issues de l'écologie du paysage. Pour des analyses de la forme urbaine avec une finalité bioclimatique telle que l'analyse des besoins en

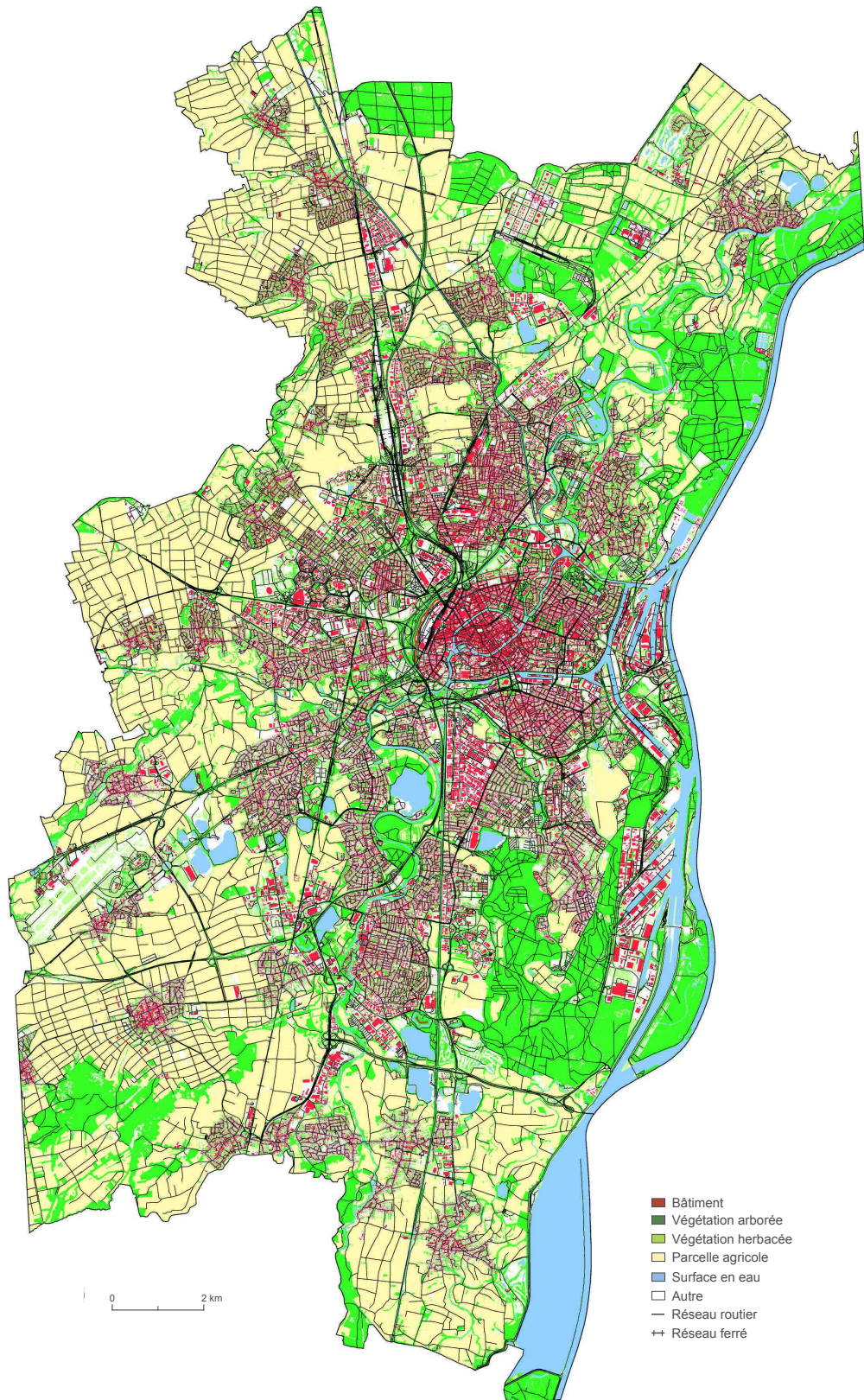


Figure 6.2 – Cartographie des éléments constitutifs des trames grise, verte et bleue de l'agglomération de Strasbourg

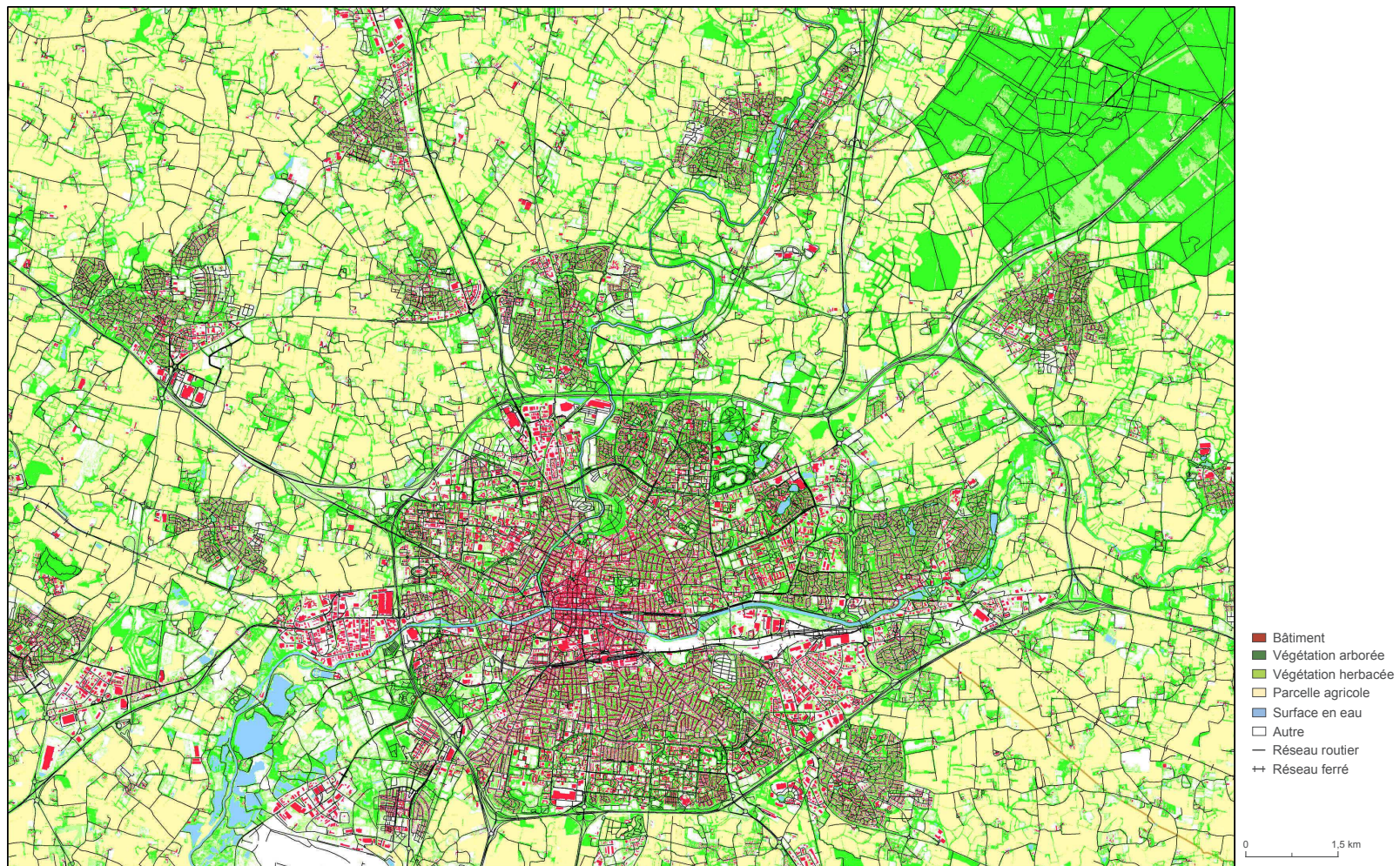


Figure 6.3 – Cartographie des éléments constitutifs des trames grise, verte et bleue de l'agglomération de Rennes

énergie, la densité des surfaces bâties est essentielle (§ 1.2.2) et est souvent associée à des critères de forme du bâti tel que leur compacité en deux ou trois dimensions (APUR, 2007 ; Bonhomme, 2013).

Pour l'étude de l'îlot de chaleur urbain, Stewart et Oke (2012) proposent de construire les LCZ avec une quantité limitée d'indicateurs. L'objectif est que leur cartographie soit facilement reproductible et repose donc sur dix indicateurs basés principalement sur la morphologie des tissus, leur composition, mais aussi leur albédo et l'utilisation humaine d'énergie. Pour ce type d'étude en climatologie urbaine et en analyse thermique des bâtiments, la hauteur est une donnée essentielle puisqu'elle détermine la surface d'interface entre ces deux milieux, mais modifie aussi l'ensoleillement au niveau du sol (Salat, 2011). Cette information a aussi montré qu'elle permettait, dans une logique de classification fonctionnelle des tissus, d'améliorer fortement les résultats et paraît donc importante à prendre en compte (Aubrecht *et al.*, 2009 ; Hecht *et al.*, 2015).

Pour Bonhomme (2013) qui a fait une revue de littérature détaillée sur les indicateurs utilisés en urbanisme, ils peuvent être décomposés en cinq axes importants qui correspondent chacun à un niveau d'échelle, du bâtiment au tissu :

- L'axe bâtiment : usage, âge, forme, taille, positionnement dans la ville, matériaux de construction
- L'axe bloc (ensemble de bâtiments contigus) : mixité des bâtiments, forme, taille
- L'axe découpage foncier : taille, forme, usage du sol
- L'axe espace public : place, voie, stationnement, espace vert
- L'axe insertion dans le tissu : histoire, forme (continue / discontinue), structure (radiale / linéaire / orthogonale), fonction (multifonctionnalité / zonage), centralité (monocentrique / polycentrique / décentralisé)

Salat (2011) avec une approche bioclimatique de la forme urbaine propose une série de 62 indicateurs regroupés en sept thèmes permettant de décrire la forme à l'échelle des tissus :

- Indicateurs de forme : géométrie des éléments, leur volume, leur empreinte dans l'espace
- D'intensité : mesurer et qualifier la densité ou la concentration d'un type d'objet
- De distribution spatiale : concentration relative ou dispersion d'objet à une échelle donnée par rapport à la totalité des objets recensés dans échelle supérieure
- De diversité : mixité, variété de types d'objet similaire à une échelle donnée
- De complexité : mesurer par l'entropie
- De proximité : distance entre deux choses
- De connectivité : accessibilité relative ou interconnexion spatiale d'un système ou d'un réseau

L'ensemble des indicateurs utilisés par Salat pour décrire les tissus urbains est aussi classiquement utilisé en écologie du paysage. La section suivante s'attache à montrer les

différents indicateurs qu'a proposés cette discipline, appelés métriques paysagères, et leur utilisation dans la littérature pour l'étude des formes urbaine.

### 6.2.2 Indicateurs issus de l'écologie du paysage

De nombreuses métriques ont été développées avec l'essor de l'écologie du paysage afin de décrire, de manière objective, la structure du paysage, son évolution et de modéliser ses effets sur la biodiversité (Herrault *et al.*, 2016). Ils permettent de caractériser une zone d'étude selon la forme des éléments d'occupation du sol et de leur organisation spatiale. Ces indicateurs sont aujourd'hui très largement utilisés que ce soit par des collectivités pour décrire leur territoire ou dans la recherche, pas seulement en écologie, mais aussi dans de nombreux autres domaines notamment en morphologie urbaine (Uemaa *et al.*, 2013). Tran *et al.* (2014) par exemple les utilisent pour décrire la morphologie des IRIS de l'agglomération de Strasbourg et de vérifier son influence sur les pratiques de mobilité de ses habitants. Afin de mieux comprendre la répartition de l'exposition aux pollutions sonores et aux particules fines, Weber *et al.* (2014) ont caractérisé les formes urbaines à l'échelle du tissu urbain et étudié les corrélations entre métriques paysagères et niveau d'exposition et ont montré que ces indicateurs pouvaient être utilisés afin de prédire la concentration des polluants selon le type de structure urbaine. Ils ont aussi très largement été utilisées pour décrire l'évolution de grandes villes et ont ainsi permis de caractériser le type de processus à l'œuvre tel que le mitage, l'étalement urbain ou la densification, sur des périodes de temps long, d'environ 30 ans, grâce la disponibilité d'images Landsat (Aguejdad, 2009 ; Taubenböck *et al.*, 2014), de prédire ses évolutions (Herold *et al.*, 2005), mais aussi de comparer la forme urbaine des villes à l'échelle de l'agglomération (Huang *et al.*, 2007 ; Schneider et Woodcock, 2008). Ces métriques paysagères ont aussi été employées afin de décrire la structure des blocs urbains par le bâti et la végétation (Herold *et al.*, 2003b ; Vanderhaegen et Canters, 2010) et ont montré qu'elles permettaient d'améliorer de manière importante les performances d'une classification fonctionnelle de ces blocs à partir de cartes d'occupation du sol extraites d'images satellites.

L'utilisation importante de ces indicateurs peut s'expliquer notamment par la mise à disposition de nombreux logiciels ou compléments disponibles librement. Fragstat est certainement le plus utilisé actuellement et regroupe plus de 100 métriques (McGarigal *et al.*, 2012), mais d'autres, tels que le logiciel Chloé2012 de l'Institut National de la Recherche Agronomique (INRA), ont aussi été développés permettant de calculer d'autres métriques paysagères (Boussard et Baudry, 2014). Dans Fragstat, l'analyse de la structure du paysage ou des éléments qui le composent se fait ici en trois niveaux : à celui de l'objet, de la classe d'occupation du sol et enfin du paysage.

Les métriques au niveau du patch sont calculées pour l'ensemble des objets d'un paysage. Elles permettent de caractériser spatialement un objet, par sa taille, sa compacité ou sa dimension fractale par exemple, mais aussi son contexte. De plus, ce niveau peut aussi

être utilisé afin de mesurer la déviation d'un objet par rapport à l'ensemble de sa classe ou de l'ensemble des objets du paysage.

Au niveau de la classe, les métriques calculées dans un premier temps au niveau de l'objet sont agrégées et permettent ainsi d'obtenir différentes statistiques afin de rendre compte de sa configuration spatiale. Ce niveau permet aussi de caractériser le paysage par le calcul d'indicateur permettant de mesurer par exemple l'abondance ou la fragmentation d'une classe particulière. Enfin, le niveau du paysage permet de caractériser l'ensemble de la mosaïque des objets. Les mesures de ce niveau peuvent être de simples mesures agrégées, calculées à l'échelle des objets sans prendre en considération leur classe d'appartenance comme au niveau précédent, mais elles permettent aussi de caractériser le paysage en terme de composition ou de diversité par exemple.

Le niveau du paysage dans le logiciel Fragstat ne considère pas nécessairement l'ensemble de la zone d'étude, mais peut aussi être un découpage de cet espace, que ce soit par un carroyage ou un découpage existant, où toutes les métriques seront calculées indépendamment pour chacune de ces entités. Ainsi les métriques au niveau du paysage peuvent être utilisées afin de caractériser et comparer des unités d'analyse et ainsi, dans cette recherche, qualifier îlots et parcelles morphologiques.

Le nombre de métriques qu'il est possible de calculer est très important et peut entraîner de nombreux problèmes afin d'être utilisé dans des modèles statistiques liés à des colinéarités ou des multicolinéarités entre de nombreuses variables (Riitters *et al.*, 1995 ; Cushman *et al.*, 2008). Il est ainsi nécessaire, soit de procéder à un tri important des métriques calculées, soit de détecter ces problèmes dans un premier temps.

De plus, la plupart des logiciels, tels que Fragstat, utilisent un mode de représentation matricielle des données, qui peuvent fortement impacter les indicateurs selon la résolution spatiale des pixels considérée et ainsi conduire à une caractérisation erronée du paysage (Wu, 2004). Les métriques paysagères calculées au niveau de la classe sont très sensibles au choix de la résolution spatiale et à l'unité cartographique minimale, plus que celles du niveau du paysage (Kelly *et al.*, 2011). La sélection d'une échelle peut donc avoir des conséquences importantes sur la recherche de facteurs explicatifs de biodiversité par ces métriques (Brennan et Schnell, 2005). De plus, avec les nouvelles méthodes de classification d'image telle que les méthodes orientées-objet (chapitre 3), la demande d'indicateurs paysagers calculés à partir d'une représentation vectorielle a augmenté et de nombreux outils directement intégrés dans des SIG sont apparus tels que Patch Analyst (Rempel *et al.*, 2012) ou PolyFrag (MacLean et Congalton, 2013).

Dans cette recherche nous avons donc fait le choix de calculer ces métriques en utilisant une représentation vectorielle des données et qui sont calculées grâce aux logiciels R et Postgis. Nous utilisons principalement des métriques au niveau de la classe et dans une moindre mesure à celle du paysage qu'il est possible de calculer à partir de données vectorielles.

### 6.2.3 Autres indicateurs proposés dans cette recherche

Les indicateurs proposés dans cette section sont mentionnés par Salat (2011) et Bonhomme (2013) pour des recherches en bioclimatique urbaine. Le choix a été fait de les séparer, car ils nécessitent d'être explicités plus en détail du fait de leur méthode de calcul ou des bases de données sur lesquelles ils s'appuient qui sont originales. Les indicateurs d'accessibilité et de visibilité présentés ici (§ 6.2.3.1) peuvent être rapprochés de ce que Salat nomme comme des indicateurs de proximité et de connectivité, ceux sur la population des indicateurs d'intensité (§ 6.2.3.2) quant à ceux basés sur les graphes, ils permettent d'évaluer la centralité des découpages testés (§ 6.2.3.3).

#### 6.2.3.1 Indicateurs d'accessibilité et de visibilité

L'accessibilité aux aménités urbaines peut être un critère permettant de différencier certains types de tissus de par une répartition de leur densité du centre vers la périphérie. Sahraoui *et al.* (2014) vont ainsi considérer l'accessibilité à plusieurs services urbains selon différents modes de transport afin de créer une typologie des espaces résidentiels. Ainsi l'accessibilité au centre urbain défini par l'hôtel de ville en terme de temps de transport en voiture est calculée pour chaque bâtiment. Les accessibilités à différents services, l'école, la boulangerie et le bureau de poste sont aussi calculées, mais en considérant cette fois le temps de transport en voiture et par la marche. Ces différentes mesures d'accessibilité leur permettent, avec d'autres indicateurs, de créer une typologie de ces espaces en six types.

Mais l'accessibilité aux infrastructures de transport est aussi une dimension importante permettant l'explication du choix de la localisation de certains types de tissus ou de bâtiments dans la ville. En effet pour Mangin (2004), la localisation des tissus urbains récents situés en périphérie des villes telles que les grands pôles économiques ou commerciaux qui se sont développées à partir des années 70 repose sur deux principes. Dans son livre *la ville franchisée*, il explique que les grandes enseignes commerciales par exemple font le choix de s'implanter à proximité de routes rapides et donc dans des zones très accessibles afin de pouvoir attirer un grand nombre de clients. De plus, une deuxième explication avancée par l'auteur est que les grandes enseignes sont présentes à proximité des grands réseaux de transport très fréquentés afin d'être visibles par le plus grand nombre leur permettant ainsi de maximiser leur visibilité.

Afin de rendre compte de ces logiques d'implantation, deux distances des bâtiments agrégés aux réseaux routiers sont calculées. La première est une distance entre les bâtiments et le réseau par le plus court chemin exprimée en mètres pour mettre en évidence leur accessibilité, la deuxième est une distance à vol d'oiseau afin de pouvoir apprécier la visibilité. Nous utilisons ici le réseau routier de la BD TOPO® avec la distinction entre les différents niveaux d'importance du réseau routier pour rendre compte de l'accessibilité et de la visibilité des bâtiments.



### 6.2.3.2 Indicateurs sociodémographiques

La population est une donnée très souvent intégrée pour cartographier les espaces urbains, qu'elle soit combinée à des données de type vecteur (Bonhomme, 2013 ; Sahraoui *et al.*, 2014) ou des données raster (Lu et Weng, 2006 ; Schneider et Woodcock, 2008). À partir de données d'occupation du sol obtenue par classification d'images par exemple, ces données démographiques peuvent ainsi permettre de retrouver l'utilisation du sol.

Dans cette recherche, le carroyage de l'INSEE de 2010 sur la population est utilisé (§ 6.1.3). Les données ont été désagrégées à l'échelle du bâtiment, car les unités d'analyse qui sont utilisées et présentées dans la section 6.3 ne correspondent pas au carroyage de l'INSEE. Cette désagrégation des données a été réalisée selon le volume des bâtiments plutôt que l'aire afin d'obtenir une meilleure estimation de la population (Aubrecht *et al.*, 2009). Ces données ont ensuite été agrégées au sein de la maille utilisée. Cette estimation de la population n'est pas non plus totalement juste à l'image de ce que l'on pourrait obtenir avec les IRIS. En effet, la taille des carreaux, même si elle est fine, inclut parfois des bâtiments non résidentiels qui sont donc, avec cette méthode, peuplée. Pour obtenir une meilleure estimation de la population, il serait donc nécessaire de reproduire dans un deuxième temps cette estimation après qu'une classification fonctionnelle des tissus ait été réalisée. Le carroyage nous permet aussi d'obtenir des informations sur la part de ménages propriétaire, mais aussi de ceux vivant en logements collectifs. Ces indicateurs sont des informations qui peuvent être importantes dans la définition des tissus (Bonhomme, 2013 ; Sahraoui *et al.*, 2014) et sont donc utilisés dans cette recherche.

### 6.2.3.3 Indicateurs de centralité

La localisation des tissus urbains dans la ville a une influence importante sur la morphologie des îlots, mais aussi du bâti (§ 1.2.1.2). Nous proposons donc dans cette recherche de prendre en considération des informations contextuelles pour chaque unité d'analyse et ainsi de pouvoir apprécier leur distribution relative au sein de la ville grâce à des mesures de centralité calculées à partir de graphes.

Les graphes permettent de modéliser des phénomènes spatiaux représentés par des sommets et leur relation telle que le voisinage à partir d'arêtes. De nombreuses métriques peuvent être dérivées de graphes. Barnsley et Barr (1997) vont décrire par les graphes les relations d'adjacences, mais aussi d'inclusion entre les objets extraits d'une image afin d'inférer l'utilisation du sol. À partir de l'extraction des bâtiments d'une image Ikonos, Doğrusöz et Aksoy (2007) vont construire un graphe planaire en déterminant les relations de voisinage entre les bâtiments grâce à une tessellation de Voronoï. Les tissus urbains sont ici décrits en deux types, organisés et désorganisés, déterminés à partir des angles que forment les arêtes des graphes puis réaliser un clustering par seuillage sur l'arbre couvrant de poids minimal. Hermosilla *et al.* (2012), pour réaliser une classification fonctionnelle des parcelles cadastrales, vont construire des graphes,

indépendant pour chaque îlot urbain, permettant de décrire le nombre d'adjacences, le degré, ainsi que la distance entre les centres de gravité de deux parcelles. De nombreuses mesures calculées à partir de graphes telles que des indicateurs de centralité, des relations d'adjacences entre différents objets extraits d'une image et de connectivité associés à des indicateurs de morphologie du bâti, ont été évaluées par Walde *et al.* (2014) pour la classification fonctionnelle d'îlots urbains grâce aux mesures d'importance des variables de Random Forest. Les mesures d'adjacences, suivies de celles de centralité puis de connectivité, sont celles ayant le plus d'impacts sur les performances de la classification.

Dans cette thèse, nous proposons d'utiliser différentes mesures de centralité calculées à partir de graphes construits non pas à partir de l'occupation des sols, mais des relations entre les mailles des découpages utilisés. Pour la construction des relations de voisinage, la règle de Queen a été appliquée, c'est-à-dire qu'une maille est considérée comme voisine à une autre si au moins un sommet est commun. L'objectif d'utiliser ce type de mesure au niveau des mailles est d'ajouter des informations contextuelles pour chacune et de pouvoir apprécier leur distribution relative au sein de la ville. Ainsi quatre mesures de centralité sont utilisées.

Les deux premières sont la centralité de degré (le nombre de voisins) et la centralité de vecteur propre qui considère aussi le nombre de voisins des sommets connectés. Ces deux mesures permettent d'évaluer la connectivité de chacune des mailles. Dans le cadre de maille régulière, comme pour le carroyage, la centralité de degré ne permet pas une réelle différenciation des mailles, sauf pour celles qui sont situées aux limites de la zone d'étude et ne sont donc pas considérées pour ces unités dans la suite de ce travail.

Les deux autres mesures de centralité permettent d'évaluer la position relative des mailles dans l'agglomération. La première est l'excentricité qui correspond au nombre de liens nécessaire pour relier le sommet le plus distant. Ainsi une forte valeur de cet indicateur signifie que le nœud n'est pas central. À l'inverse, la centralité de proximité va être plus importante si la maille est centrale dans la zone d'étude. Elle correspond à l'inverse de la somme normalisée des plus courts chemins pour rejoindre l'ensemble des nœuds du graphe.

Le tableau 6.3 résume l'ensemble des attributs calculés pour chaque unité d'analyse présentée dans la section suivante.

**Tableau 6.3** – Vue d'ensemble des attributs calculés pour décrire les différentes mailles

Nom	Attributs / Indicateurs	Classe considérée
Type d'indicateur		
Indicateurs de forme de la maille (urbanisme et écologie du paysage)		
<b>maille.aire</b>	Aire (en m <sup>2</sup> ) de la maille	
Indicateur morphologique		
<b>maille.perim</b>	Périmètre (en m) de la maille	
Indicateur morphologique		
<b>maille.SHAPE</b>	Indice de forme de la maille	
Indicateur morphologique		
	$SHAPE = \frac{perim}{2\pi\sqrt{\frac{aire}{\pi}}}$	
	Supérieur à 1, où 1 correspond à la forme la plus compacte (cercle)	
<b>maille.FRAC</b>	Indice de forme fractale de la maille	
Indicateur morphologique		
	$FRAC = \frac{2\ln(0.25perim)}{\ln(aire)}$	
	Varie entre 1 et 2, où 1 correspond à une forme simple et 2 à des formes très fractales	
Indicateurs de forme des éléments gris, verts et bleus (urbanisme et écologie du paysage)		
<b>aire.mean</b>	Aire moyenne (en m <sup>2</sup> )	Toutes les classes
Indicateur morphologique		
<b>aire.sd</b>	Écart-type des aires (en m <sup>2</sup> )	Toutes les classes
Indicateur morphologique		
<b>SHAPE.mean</b>	Moyenne des indices de forme des éléments	Toutes les classes
Indicateur morphologique		
<b>SHAPE.sd</b>	Écart-type des indices de forme des éléments	Toutes les classes
Indicateur morphologique		
<b>FRAC.mean</b>	Indice de forme fractale moyen des éléments	Toutes les classes
Indicateur morphologique		
<b>FRAC.sd</b>	Écart-type des indices de forme fractale des éléments	Toutes les classes
Indicateur morphologique		
<b>H.mean</b>	Hauteur moyenne (en m)	Bâtiment agrégé et non agrégé
Indicateur morphologique		
<b>H.sd</b>	Écart-type des hauteurs (en m)	Bâtiment agrégé et non agrégé
Indicateur morphologique		
<b>volume.mean</b>	Volume moyen (en m <sup>3</sup> )	Bâtiment agrégé et non agrégé
Indicateur morphologique		
<b>volume.sd</b>	Écart-type des volumes (en m <sup>3</sup> )	Bâtiment agrégé et non agrégé
Indicateur morphologique		

Nom Type d'indicateur	Attributs / Indicateur	Classe considérée
<b>compacite3D.mean</b> Indicateur morphologique	Compacité en trois dimensions moyenne (Bonhomme, 2013) $compacite3D = \frac{perim \times H + 2 \times aire}{volume^{2/3}}$	Bâtiment agrégé et non agrégé
<b>compacite3D.sd</b> Indicateur morphologique	Écart-type des compacités en trois dimensions	Bâtiment agrégé et non agrégé
<b>orientation.mean</b> Indicateur morphologique	Orientation moyenne (en °) Varie de 0 à 180°, 0-180 correspondant à une orientation nord-sud et 90 à une orientation est-ouest	Bâtiment agrégé et non agrégé
<b>orientation.sd</b> Indicateur morphologique	Écart-type des orientations (en °)	Bâtiment agrégé et non agrégé
<b>Indicateurs issus de l'écologie du paysage</b>		
<b>PLAND</b> Indicateur d'abondance	Part d'occupation du sol en % (correspond à la densité nette en urbanisme)	Toutes les classes
<b>PD</b> Indicateur de densité et de fragmentation	Densité d'éléments (en nombre pour 100 ha) $PD = \frac{Nombre\ d'\ elements}{maille.aire} \times 10\ 000 \times 100$	Toutes les classes
<b>Meff</b> Indicateur de fragmentation	Largeur effective de maille modifiée (en m <sup>2</sup> ) (Moser <i>et al.</i> , 2006) $Meff = \frac{\sum_1^n A_i \times A_i^{cmpl}}{maille.aire}$ où $n$ correspond au nombre de patch, $A_i$ à l'aire du patch à l'intérieur de la maille et $A_i^{cmpl}$ à l'aire complète du patch	Toutes les classes
<b>LPI</b> Indicateur de dominance	Part en % du plus grand patch	Toutes les classes
<b>ED</b> Indicateur de densité et morphologique	Densité de lisière (en m/ha) $ED = \frac{\sum perim}{maille.aire} \times 10\ 000$	Toutes les classes
<b>COHESION</b> Indicateur de connectivité	Indice de cohésion des patches en % $COHESION = \left[ 1 - \frac{\sum_1^n perim}{\sum_1^n perim \sqrt{aire}} \right] \left[ \frac{1}{\sqrt{maille.aire}} \right]^{-1} \times 100$ Entre 0 et 100, approche de 0 quand la classe d'intérêt à une part faible dans la maille, qu'elle est fortement subdivisée et moins connectée	Toutes les classes
<b>SHDI</b> Indicateur de diversité	Indice de Shannon $SHDI = -\sum_1^m (PLAND \times \ln PLAND)$ où $m$ correspond au nombre de classes Supérieur à 0, 0 correspond à une maille composée d'une seule classe	

Nom	Attributs / Indicateur	Classe considérée
Type d'indicateur		
<b>route.ED</b> Indicateur de densité	Densité de route (en m/ha) par niveau d'importance dans la maille	Routes
<b>voie.ferrée.ED</b> Indicateur de densité	Densité de voies ferrées (en m/ha) par niveau d'importance dans la maille	Voies ferrées
Indicateurs d'accessibilité et de visibilité		
<b>dist.route.NN</b> Indicateur de visibilité	Distance moyenne (en m) à la route la plus proche à vol d'oiseau par niveau d'importance	Bâtiments agrégés
<b>dist.route.NNroute</b> Indicateur d'accessibilité	Distance moyenne (en m) à la route la plus proche par la réseau et par niveau d'importance	Bâtiments agrégés
Indicateurs sociodémographiques		
<b>insee.densHab</b> Indicateur de densité	Densité d'habitant en hab/km <sup>2</sup> dans la maille	
<b>insee.densMen</b> Indicateur de densité	Densité de ménage en men/km <sup>2</sup> dans la maille	
<b>insee.parColl</b>	Part de ménage habitant en logement collectif dans la maille	
<b>insee.parProp</b>	Part de ménage propriétaire de leur logement dans la maille	
Indicateurs de centralité		
<b>graph.degree</b> Indicateur de centralité	Degré des sommets du graphe Nombre de voisins d'une maille Maximal lorsque le sommet est fortement connecté	
<b>graph.eigCentrality</b> Indicateur de centralité	Centralité de vecteur propre d'un sommet du graphe Mesure qui rend compte de la connexion d'un sommet à des sommets très connectés Maximale lorsque le sommet est connecté à des sommets très connectés	
<b>graph.eccentricity</b> Indicateur de centralité	Excentricité d'un sommet du graphe Nombre de liens nécessaires pour relier le lien le plus distant Maximale lorsque le nœud est moins central	
<b>graph.closeness</b> Indicateur de centralité	Centralité de proximité d'un sommet du graphe Normalisation de somme des longueurs pour relier l'ensemble des autres sommets Varie de 0 à 1, 1 correspond à une centralité importante	

## 6.3 Unités testées pour caractériser la structure urbaine

Trois types d'unités d'analyse régulièrement utilisés dans la littérature sont testées : le carroyage (§ 6.3.1), la parcelle (§ 6.3.2) et l'îlot (§ 6.3.3). Nous cherchons à identifier quel est le découpage le plus adapté pour cartographier les formes urbaines à l'échelle des tissus urbains.

### 6.3.1 Découpage par carroyage

Le carroyage est un découpage arbitraire de l'espace souvent utilisé pour cartographier les tissus urbains. Le choix d'une taille de maille est un paramètre important qui peut influencer la qualité des indicateurs calculés pour chaque maille et de nombreuses tailles différentes ont été utilisées dans la littérature pour décrire les espaces urbains pouvant varier de 50 (Antoni, 2001) à 250 m (Bonhomme, 2013) selon les objectifs de l'étude (§ 2.3.3.1). En écologie du paysage, la taille des carreaux est en général plus grande que celles mentionnées précédemment. En effet, le terrain d'étude est souvent plus étendu et il est nécessaire d'avoir une maille suffisamment grande pour permettre de discerner des structures paysagères influençant la biodiversité. Ainsi les tailles de mailles souvent retenues dans ce domaine sont très diverses et leur surface peut aller de quelques hectares à quelques milliers d'hectares (Barreto *et al.*, 2010 ; Mimet *et al.*, 2014).

Dans ces recherches, la maille est de forme carrée, mais des formes différentes peuvent aussi être utilisées afin de créer un carroyage. La maille carrée est la plus classiquement utilisée, mais il est aussi possible de procéder à un maillage hexagonal, en nid d'abeille, de l'espace (Birch *et al.*, 2007). Les deux formes considérées, carrée et hexagonale, possèdent chacune des avantages, mais aussi des limites. La maille carrée présente l'avantage d'être symétrique et orthogonal et est aussi le plus simple ce qui permet son utilisation aisée dans de nombreux modèles tant en écologie du paysage (Birch *et al.*, 2007) qu'en climatologie urbaine (Bonhomme, 2013). Il permet aussi une représentation multi-échelles facile de l'espace en fusionnant ou découpant chacune des mailles d'un niveau sans qu'aucune superposition ne soit créée et une intégration facile avec des données matricielles, à la différence de l'utilisation d'une maille hexagonale. Par contre, l'utilisation d'un maillage carré peut être ambiguë lorsque l'on considère le voisinage direct de chacune des mailles, notamment pour des recherches portant sur la connectivité ou le déplacement d'espèces. Il est en effet nécessaire de déterminer la règle pour le voisinage qui peut considérer uniquement les mailles dont au moins un segment est partagé comme voisin, et ainsi obtenir quatre voisins pour chacune, ou au moins un sommet est partagé et ainsi en considérer huit pour chaque maille. Avec un maillage hexagonal, cette ambiguïté sur ce nombre de voisins est levée, rendant son utilisation préférable pour des travaux où le voisinage est important (Birch *et al.*, 2007).

En conclusion vu les avantages et les inconvénients de chacune, les deux types de formes de maille sont comparés afin de vérifier lequel est le plus adapté pour caractériser la structure urbaine. Trois tailles de maille pour chacune des formes sont considérées (tableau 6.4). Ainsi pour la maille de forme carrée, les longueurs de côté considérées sont de 125, 250 et 500 m. Les trois mailles de forme hexagonale ont la même surface que celle de forme carrée afin de permettre leur comparaison et ont ainsi des côtés d'une longueur de 77,5, 155 et 310 m. Dans la suite de ce travail afin de faciliter la lecture, nous parlons de tailles de maille de 125, 250 ou de 500 m quelle que soit la forme considérée.

### 6.3.2 Découpage à la parcelle morphologique

La parcelle cadastrale est un découpage de l'espace très souvent utilisé dans les études de morphologie urbaine du fait qu'il s'agit de la plus petite unité permettant l'analyse et la compréhension des tissus (§ 1.2.1.1). Pour la cartographie des tissus urbains, cette unité d'analyse a régulièrement été considérée (§ 2.3.3.1). Les parcelles cadastrales sont intégrées à la BD Parcellaire® pour le site de Rennes, mais pour Strasbourg, les parcelles disponibles sur le site internet de l'Eurométropole ne l'étaient pas encore au moment de la réalisation de ce chapitre (§ 2.2.2). Nous utilisons dans ce travail la définition de la parcelle morphologique donnée par Hamaina *et al.* (2012) et utilisons sa méthode afin de les construire. Elles sont construites avec un diagramme de Voronoï réalisé à partir de l'empreinte au sol des bâtiments.

La tessellation de Voronoï permet de créer une partition d'un plan en régions, basée sur le calcul de distance, généralement entre des points, et représente en quelque sorte l'aire d'influence morphologique de chacun des bâtiments. Chaque cellule permet ainsi de caractériser les espaces ouverts souvent utilisés pour décrire la morphologie des tissus et de s'affranchir d'un découpage existant de l'espace. Pour Hamaina, ces parcelles qui sont basées sur des caractéristiques structurelles des bâtiments, sont plus pertinentes que des découpages existants qu'ils soient administratifs ou fonciers pour l'étude des formes urbaines. En effet les parcelles cadastrales ne sont pas nécessairement en adéquation avec la réalité architecturale, certaines parcelles pouvant être regroupées au sein d'un même îlot de propriétés (Allain, 2004).

Afin de réaliser cette tessellation, la méthode proposée par Dong (2008) est utilisée dans ce travail. Elle possède plusieurs avantages. Cette méthode permet de ne pas se limiter seulement à une représentation ponctuelle des bâtiments comme c'est souvent le cas, mais de considérer aussi des lignes ou des polygones pour la construction de la tessellation. Un poids peut aussi être attribué à chacun des bâtiments dans le but de pondérer le calcul des distances, normalement équidistantes, pour fixer les limites des cellules entre deux bâtiments. De plus, elle est disponible librement et est intégrée dans le logiciel Arcgis. Toutefois, cette méthode n'est pas sans limites. En effet, elle nécessite de rasteriser dans un premier temps les polygones pour créer le diagramme. Une taille de pixel trop large peut entraîner la fusion de deux polygones voisins si leur

distance est inférieure à la résolution du raster, ou même les voir disparaître si leur surface est trop petite. À l'inverse, une résolution trop fine nécessite un raster de grande dimension entraînant de ce fait un temps de calcul très long et des limites en termes de consommation de mémoire vive qui peuvent être rapidement atteintes.

Afin de construire ces diagrammes de Voronoï, l'empreinte au sol des bâtiments agrégés a été utilisée comme graines. L'utilisation des bâtiments non agrégés était en effet impossible du fait des limites de cette méthode. Comme nous l'avons mentionné, elle passe par une rastérisation des éléments qui aurait fusionné les bâtiments contigus. De plus la couche non agrégée des bâtiments contient plus de bâtiments ce qui empêche l'exécution de la tessellation à cause de la limite de mémoire vive que peut utiliser cette méthode qui est vite atteinte. Enfin les bâtiments isolés ou ceux situés en bordure de zone bâtie, une distance de 50 m a été fixée comme limite de leur aire d'influence morphologique qui correspond au seuil que Loriot (2008) utilise pour la construction de la tache urbaine. Les routes ainsi que les réseaux ferroviaires les plus importants ont aussi été considérés comme des ruptures dans le tissu urbain et donc comme des limites aux parcelles morphologiques des bâtiments. Cette méthode permet d'attribuer un poids différent pour le calcul de distance. Hamaina *et al.* (2012) utilisaient le volume pour pondérer la création du diagramme. Toutefois, comme nous considérons déjà la surface des bâtiments et non des points pour les représenter, nous utilisons donc uniquement la hauteur des bâtiments afin de pondérer ce calcul avec l'hypothèse que plus un bâtiment est haut, plus son aire d'influence est grande comparée à ces voisins de plus petite hauteur.

Dans les espaces urbains denses, où le réseau de rues est dense et où les bâtiments sont tous mitoyens au sein d'un îlot, les cellules du diagramme de Voronoï (parcelles morphologiques) ont tendance à être très similaires aux îlots urbains délimités par les routes (figure 6.4.a). Dans les espaces plus discontinus, notamment pour les zones d'habitat pavillonnaire, les cellules obtenues sont plus proches visuellement des parcelles cadastrales et découpent ainsi les îlots en autant de parcelles qu'il existe de bâtiments (figure 6.4.b – tableau 6.4).

### 6.3.3 Découpage à l'îlot

La définition d'un tissu urbain renvoie classiquement à un découpage de l'espace urbain en îlot (§ 1.2.1.1). C'est une unité très largement utilisée en urbanisme pour la construction de typologie de tissu (§ 2.3.3.1). Elle est, de plus, celle de base utilisée par l'IGN, l'ossature, afin de construire la nouvelle couche d'occupation du sol grande échelle nationale (OCS GE) qui permettra d'obtenir à une échelle fine des informations sur l'utilisation du sol, uniformisée pour l'ensemble du territoire français (IGN, 2013). En urbanisme, l'îlot correspond à un ensemble de parcelles cadastrales contiguës délimité par les voies de circulation (Allain, 2004).



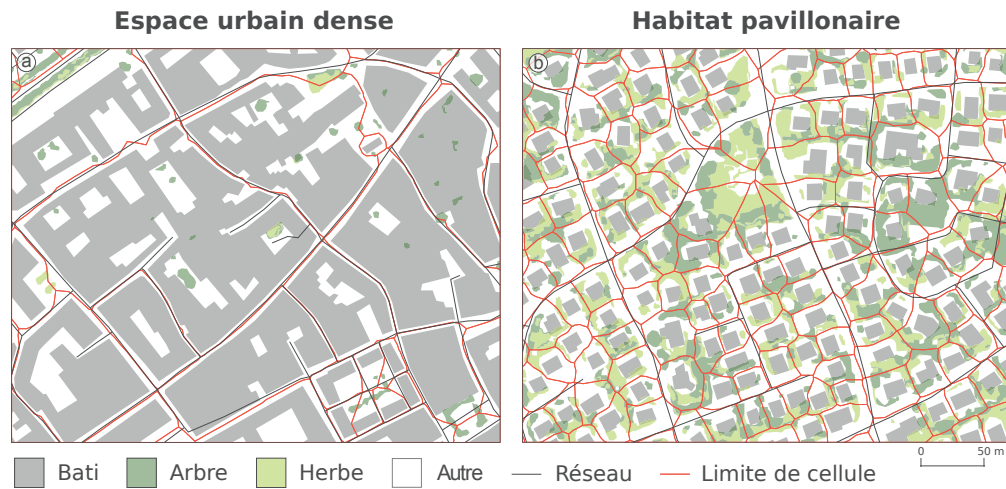


Figure 6.4 – Exemple du diagramme de Voronoï obtenu en espace urbain dense (a) ou pavillonnaire (b) pour Strasbourg

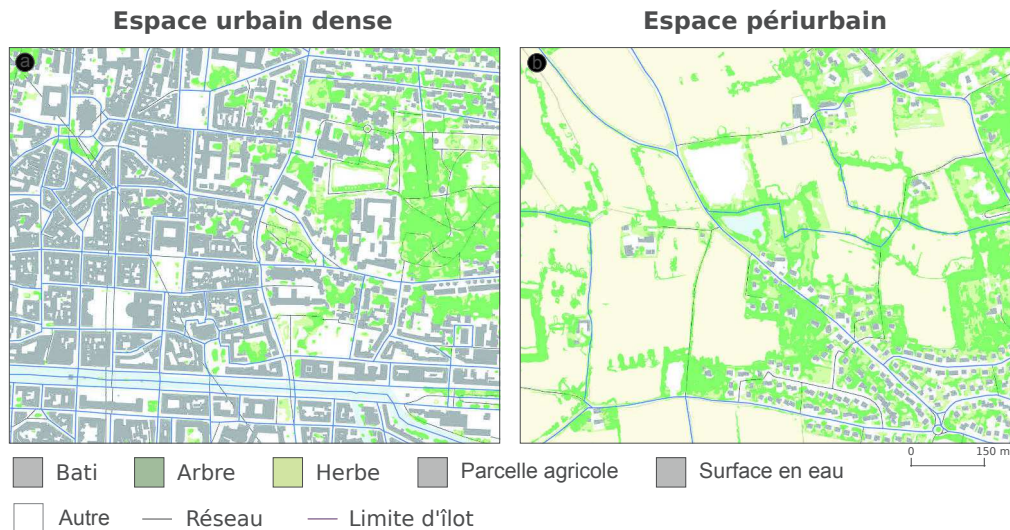
Quand les parcelles cadastrales ne sont pas disponibles, une solution classique est d'utiliser le linéaire de communication afin de construire les îlots (Novack et Kux, 2010 ; Hussain *et al.*, 2012). Les éléments linéaires considérés par Puissant *et al.* (2011) pour la création des îlots sont les routes et les réseaux ferrés les plus importants ainsi que l'ensemble du réseau hydrographique. Lorsque ces différents réseaux forment un cycle fermé alors un îlot est créé. La définition des îlots est donc ici différente de celle qui a été mentionnée précédemment puisque les espaces publics tels que la rue ne délimitent pas l'îlot, mais sont intégrés dans chacun des îlots (figure 6.5).



Figure 6.5 – Exemple d'îlots construits (a) à partir des parcelles cadastrales et (b) à partir des réseaux de communication (les limites d'îlots sont figurées en rouge)

Nous utilisons dans cette recherche les différents éléments linéaires de la BD TOPO®. Toutefois avec cette méthode de construction de nombreux îlots de petites dimensions peuvent être créés notamment dans les zones où la densité des réseaux, quels qu'ils soient, est importante ou lorsque des petits objets sont entourés par un réseau, par

exemple un rond-point. Ainsi tous ceux dont la surface est inférieure à 1500 m<sup>2</sup> ont été fusionnés au voisin avec lequel il partage la plus longue bordure. Ce seuil a été déterminé de manière empirique et visuelle et en vérifiant la taille minimale des îlots où un bâtiment est inclus. Les tailles des îlots créés peuvent être très différentes. Elles sont petites quand la densité de réseau est très importante comme dans les centres-ville et de grandes dimensions dans les espaces périurbains (figure 6.6).



**Figure 6.6** – Exemple de découpage en îlot obtenu en espace urbain dense (a) ou périurbain (b) pour Rennes

Les unités d'analyse proposées sont comparées et évaluées dans la section suivante afin de déterminer laquelle est la plus adaptée pour la cartographie de la structure urbaine des agglomérations de Strasbourg et de Rennes sur la base des indicateurs présentés dans le tableau 6.3 qui sont calculés pour l'ensemble de ces unités. Le tableau 6.4 résume les différentes unités d'analyse considérées et le nombre de mailles de chacune pour chaque site.

**Tableau 6.4** – Unité d'analyse testée pour la cartographie de la structure urbaine et nombre de mailles pour Strasbourg et pour Rennes

Unité d'analyse	Strasbourg	Rennes
Parcelle morphologique	56775	44640
Îlot	5535	3517
Carrée (125 m)	20108	17967
Carrée (250 m)	5147	4560
Carrée (500 m)	1345	1160
Hexagonale (125 m)	20080	17972
Hexagonale (250 m)	5154	4610
Hexagonale (500 m)	1347	1188

## 6.4 Sélection de l'unité d'analyse pour la cartographie de la structure urbaine

Le choix d'une unité adaptée à l'analyse de la structure urbaine est primordial puisqu'il va fortement influencer la qualité des différents indicateurs (Long *et al.*, 2008) qui elle-même aura une influence importante sur la classification de ces unités par la suite. En effet si sa taille est trop petite elle ne permet d'analyser la structure de chacun des tissus tandis que si elle trop grande plusieurs tissus différents peuvent être englobé dans une même unité.

L'objectif de cette section est de vérifier s'il existe un découpage adapté pour la cartographie de la structure urbaine. Cette évaluation est effectuée sur l'agglomération de Strasbourg puis validée sur le site d'études de Rennes. La section 6.4.1 présente les différents critères choisis pour évaluer les découpages puis la section 6.4.2 détaille les résultats de cette évaluation.

### 6.4.1 Sélection des critères d'évaluation

#### 6.4.1.1 Critères d'évaluation existants pour le choix d'une unité d'analyse

Dans sa thèse, Bonhomme (2013) teste quatre tailles de maille de forme carrée : 50 m, 100 m, 250 m et 500 m. Pour elle, une maille de taille adaptée doit répondre à quatre objectifs : permettre d'obtenir des unités homogènes en réduisant l'écart-type d'indicateurs au sein de la maille, différencier les mailles les unes des autres, éviter la fragmentation des bâtiments au sein de la maille et enfin limiter le temps de calcul à des fins de modélisation. Plusieurs mailles pour chaque type de tissus sont sélectionnées aléatoirement et photo-interprétées afin d'évaluer la pertinence de chaque taille de maille pour chacun des types. L'analyse statistique menée dans cette thèse s'appuie sur l'analyse de la variance au sein des mailles qu'il faut minimiser et entre les mailles qu'il faut maximiser. De plus, un critère est ajouté à cette analyse pour mesurer la fragmentation des bâtiments. L'objectif est que la fragmentation soit faible afin que les indicateurs dérivés de ces éléments ne soient donc pas calculés sur des parties de bâtiments. Toutefois, cette analyse ne permet pas réellement de voir un découpage se détacher des autres, car certains critères sont optimisés avec certaines unités d'analyse et pas d'autres et une maille est de 250 m est finalement choisie, car elle représente un bon compromis entre la qualité des indicateurs et le temps de calcul pour la suite de sa démarche.

Long (2003) utilise une approche similaire à la précédente et va à partir de quartier représentatif évaluer trois mailles carrées de 50 m, 100 m et 200 m. Après avoir écarté celle 50 m car elle entraîne une trop grande fragmentation des tissus, les indicateurs utilisés pour la cartographie des tissus sont aussi utilisés pour l'évaluation de la taille des mailles afin de choisir celle la plus adaptée : la densité bâtie, la hauteur moyenne des bâtiments, l'écart-type de la hauteur au sein de la maille, l'espacement moyen entre

les bâtiments et le nombre de bâtiments par hectare. Différentes statistiques descriptives sont calculées pour chaque indicateur et type de tissus permettant de retenir une maille de 200 m.

#### 6.4.1.2 Critères d'évaluation retenus

Plusieurs critères ont été retenus dans ce travail pour déterminer l'unité d'analyse la plus adaptée à l'étude de la trame grise et de la trame verte. Pour cette analyse il est nécessaire d'avoir une vérité terrain permettant de décrire l'occupation du sol. Pour l'agglomération de Strasbourg, nous disposons de la BDOCS datant de 2012 de la région Alsace qui nous permet de réaliser cette analyse pour l'ensemble de ce site tant dans les espaces urbains que dans les non urbains. La typologie retenue afin d'évaluer ces unités d'analyse est celle en 27 classes présentées par la figure 2.7. Cette typologie est celle la plus proche de celle qui a été retenue dans le cadre du projet VALI-URB (figures A et B).

Les trois premiers critères d'évaluation utilisés sont issus de la littérature : la fragmentation des éléments, la variance intra-classe et la variance inter-classes. Nous considérons la fragmentation des bâtiments comme un critère de choix, mais nous l'étendons ici à l'ensemble des autres classes d'occupation du sol : la végétation arborée et herbacée, les surfaces en eau et les parcelles agricoles. Pour une classification, qu'elle soit supervisée ou non, l'idéal est que les indicateurs pour une classe soient très semblables et en même temps très différents des autres classes. Ainsi pour chaque découpage et chacun de leurs indicateurs, les variances sont calculées avec pour objectif de minimiser celles intra-classe et maximiser celles inter-classes.

Ces dernières analyses correspondent en statistiques à des analyses de variances univariées. Pour approfondir l'évaluation des unités d'analyse, nous proposons d'utiliser l'analyse de variance généralisée aux données multidimensionnelles, la MANOVA (*Multivariate Analysis of Variance*). L'objectif de ce test statistique est de vérifier si des groupes permettent d'expliquer des différences dans les données, si leurs moyennes d'un point de vue multidimensionnelles sont différentes. Ce test est toutefois très sensible aux problèmes de multicollinéarité dans les données pouvant empêcher sa réalisation. Une solution pour s'en détacher est de réaliser préalablement une ACP (Huberty et Olejnik, 2006) qui permet par combinaison linéaire des variables de créer des axes orthogonaux non corrélés. Plusieurs statistiques permettent de tester l'hypothèse d'homogénéité des moyennes de la MANOVA. Le plus souvent utilisé est le lambda de Wilks, mais la trace de Pillai-Bartlett est celui qui apparaît le plus robuste aux violations des différents postulats (Olson, 1974). Nous ne nous intéressons pas ici à savoir si les groupes permettent d'expliquer des différences dans les données. En effet, le nombre de mailles pour chacun des découpages est relativement important et l'hypothèse nulle d'homogénéité des moyennes entre les groupes serait toujours rejetée. La valeur de la trace de Pillai-Bartlett permet, à partir des matrices de variance covariance, de mettre en évidence la part de variance expliquée par les groupes par rapport à celle totale. Plus la variance intra-classe

sera faible et celle inter-classe forte, plus la valeur de la trace de Pillai-Bartlett sera grande. Chaque classe est évaluée en une contre les autres et ces valeurs permettent ainsi de vérifier quel découpage est le plus adapté pour chaque classe de la typologie considérée.

Enfin, le dernier critère d'évaluation que nous utilisons est l'incertitude liée à l'attribution d'une classe à une maille. Comme précédemment pour la classification des images satellite, les classes attribuées sont celles dont la surface qu'elles représentent au sein d'une maille est la plus importante. Si cette incertitude est grande alors la maille comprend différents tissus ce qui est signe d'un mauvais découpage du fait de son hétérogénéité. Afin d'évaluer cette incertitude un indice permettant de mesurer la diversité est utilisé, celui de Simpson ( $I_S$ ) modifié pour être indépendant du nombre de classes présentes (Salat, 2011) dont l'équation est :

$$I_S = \sum_{i=1}^N P_i^2 \times \frac{N}{N-1} \quad (6.1)$$

où  $P_i$  représente la part de la surface de chaque classe au sein d'une maille et  $N$  le nombre de classes présentes dans la maille. Cet indice varie de 0 pour une maille complètement homogène à 1 pour une maille où les classes de la base de référence sont également distribuées.

En conclusion cinq critères d'évaluation ont été retenus pour déterminer l'unité d'analyse la plus adaptée à la cartographie de la structure urbaine : (1) la fragmentation des éléments, (2) la variance intra-classe, (3) la variance inter-classe, (4) la MANOVA et (5) l'incertitude. La section suivante présente les résultats de cette analyse.

#### 6.4.2 Évaluation des mailles pour l'analyse de la structure urbaine

Les cinq critères d'évaluation ont été classés du meilleur au moins bon et des moyennes de rang ont été réalisées pour les huit unités d'analyse (tableau 6.4). Pour la fragmentation des classes d'occupation du sol, chacun des découpages a été classé selon qu'il fragmente plus ou moins chaque classe et la moyenne pour l'ensemble calculée pour chacun d'eux. Le classement pour l'analyse univariée des variables descriptives par les variances intra et inter classes représente le rang moyen de l'ensemble de ces variables pour chaque unité d'analyse. De plus, les critères de variance intra-classe, MANOVA et incertitude ont tout d'abord été calculé par classe issue de la typologie puis moyennés pour chaque unité d'analyse. Les résultats par classes d'occupation du sol sont disponibles en annexe D.

Il est aussi apparu intéressant de faire une différenciation entre les classes urbaines et non urbaines afin de vérifier si certains découpages étaient plus ou moins adaptés à l'analyse de ces deux types d'espace. Les classes urbaines sont celles correspondant aux

territoires artificialisés du niveau 1 de la BDOCS et les non urbaines aux trois autres classes de ce même niveau (annexe B.3).

Enfin, une ACP a été réalisée avant la réalisation de la MANOVA pour s'affranchir des problèmes de multicollinéarité entre les variables. L'ensemble des composantes n'a pas été utilisé pour cette analyse, mais seules celles dont la valeur propre était supérieure à 1 ont été conservées comme Kaiser (1960) le propose du fait de sa simplicité. Selon les découpages, de 27 à 36 composantes ont ainsi été retenues expliquant entre 75 et 78 % de la variance totale des données.

### (1) La fragmentation des éléments

L'analyse de la fragmentation des éléments gris et verts montre un avantage certain au découpage en îlot (Tableau 6.5). En effet, l'ensemble des classes d'occupation du sol est moins fragmenté avec ce découpage. Peu de ces éléments se superposent aux réseaux linéaires ayant servi à la construction des îlots exceptés les surfaces en eau et particulièrement les cours d'eau qui sont eux tous fragmentés à cause de l'utilisation du réseau hydrographique lors de la construction de cette unité d'analyse. Ces éléments se trouvent donc moins fragmentés avec le découpage en parcelles morphologiques.

Ces parcelles sont la deuxième unité d'analyse permettant de limiter la fragmentation des éléments. Il est toutefois important de noter que ce découpage ne représente pas l'ensemble de l'agglomération, mais seulement ceux à proximité des bâtiments. Ainsi la fragmentation des éléments, excepté les bâtiments, est minimisée puisqu'elle représente la part de chaque classe dans son ensemble. La fragmentation des bâtiments devrait aussi normalement être nulle, car le diagramme de Voronoï a été réalisé à partir d'eux ce qui montre une approximation liée à la construction de la tessellation qui passe par une représentation matricielle.

Enfin, les différents carroyages proposés montrent une fragmentation relativement importante des éléments gris et verts. Assez logiquement, plus les mailles sont grandes moins ces éléments sont fragmentés. On remarque néanmoins que les grilles hexagonales ont tendance, pour une même taille, à moins les fragmenter.

**Tableau 6.5** – Part d'éléments fragmentés par classe d'occupation du sol et rang moyen par unité d'analyse proposée

Maille	Bâtiment	Arbre	Herbe	Eau	Parcelle	Rang moyen
Parcelle morphologique	0,6 %	7,4 %	12,1 %	25,1 %	20,8 %	2,6
Îlot	0,1 %	4,7 %	7,4 %	27,0 %	8,4 %	1,2
Carrée (125 m)	29,6 %	19,0 %	24,6 %	76,2 %	96,2 %	7,8
Carrée (250 m)	16,0 %	10,3 %	15,1 %	60,4 %	79,4 %	6,0
Carrée (500 m)	8,3 %	5,5 %	9,9 %	46,5 %	50,0 %	3,6
Hexagonale (125 m)	28,8 %	18,2 %	23,9 %	77,7 %	96,0 %	7,2
Hexagonale (250 m)	15,6 %	9,8 %	14,7 %	60,2 %	78,3 %	5,0
Hexagonale (500 m)	8,2 %	5,3 %	9,7 %	44,5 %	48,4 %	2,6

## (2) La variance intra et (3) inter-classe

Les quatre autres critères d'évaluation utilisés dans cette recherche montrent des résultats différents (Tableau 6.6). La variance intra-classe montre que les différents carroyages permettent d'obtenir en moyenne des classes plus compactes dans l'espace des données que les îlots ou les parcelles et ce qu'elles soient urbaines ou non. Plus ces mailles sont grandes et plus cette variance est faible. À l'inverse, la variance inter-classe a tendance à permettre une meilleure différenciation entre les différents types d'occupation du sol lorsque les mailles sont de petites dimensions. L'augmentation de la taille de la grille a donc pour effet d'uniformiser l'espace ce qui se traduit par certes une meilleure ressemblance entre les mailles d'un même type, mais aussi entre les différents types.

## (4) La MANOVA

Le critère d'évaluation issu de la MANOVA et de la trace de Pillai-Bartlett montre des résultats différents de ceux précédents. Tout d'abord, les parcelles morphologiques ne permettent pas de faire une bonne différenciation entre les différents types urbains ou non. Le fait que chacune ne contient qu'un seul bâtiment ne permet donc pas de rendre compte de l'organisation des tissus. Pour les autres découpages, les rangs moyens sont plus semblables avec un léger avantage pour les grilles de 125 m, carrée ou hexagonale, lorsque l'ensemble des classes est pris en compte. Toutefois, ce critère met en évidence des performances plus contrastées quand ces classes sont séparées en urbaines et non urbaines. Dans le cas des classes urbaines, le découpage en îlot apparaît être le plus adapté pour rendre compte des spécificités de chacun des tissus. Par contre pour les classes non urbaines ce sont les carroyages les plus fins qui permettent de mieux identifier les types associés.

## (5) L'incertitude

Les conclusions de la MANOVA se retrouvent sur notre dernier critère d'évaluation, l'incertitude lors de l'attribution des classes aux unités de chaque découpage. En effet, même si dans l'ensemble celui en îlot permet d'obtenir des performances meilleures que précédemment et qu'il est très bien adapté aux classes urbaines, il entraîne une forte incertitude lors de l'attribution des classes non urbaines. Dans les espaces urbains, le réseau linéaire de transport est très dense créant ainsi de petits îlots dont la majorité se retrouve assez homogène en terme de classes issues de la base de données d'occupation des sols. Par contre dans ceux non urbains, les îlots sont beaucoup plus grands créant ainsi des îlots mixtes. Les grilles les plus fines se retrouvent donc moins incertaines pour ce type d'espace puisqu'elles sont régulières sur l'ensemble de l'espace d'étude, mais restent plus diverses pour les zones urbaines.

L'étude de ces différents critères d'évaluation montre que le choix d'une maille permettant de décrire et de caractériser l'ensemble de l'espace d'étude est difficile. En effet selon les critères, mais aussi l'appartenance des mailles aux espaces urbains ou non urbains, les performances diffèrent. Si la différence est faite entre les types d'espaces, ils permettent de montrer que le découpage en îlot est certainement celui qui serait le plus adapté à l'étude de ceux urbains. Pour les espaces non urbains par contre les grilles de 125 m sont certainement celles les plus adaptées sans que la forme hexagonale ou carrée n'apparaissent être significativement meilleure. Il serait donc intéressant de pouvoir combiner ces deux types de découpage afin de retranscrire le plus fidèlement possible l'ensemble de la structure urbaine de l'agglomération de Strasbourg.

**Tableau 6.6** – Rang moyen par critère d'évaluation et par unité d'analyse

Critère	Classe	Parcelle	Îlot	Grille carrée			Grille hexagonale		
				125 m	250 m	500 m	125 m	250 m	500 m
Variance intra-classe	Ensemble	6,0	4,7	4,6	4,2	3,7	4,6	4,0	3,8
	Urbaines	6,2	4,7	4,8	4,1	3,5	4,8	3,9	3,7
	Non urbaines	5,2	4,2	4,0	4,6	4,3	4,3	4,1	4,2
Variance inter-classe	Ensemble	3,5	5,0	3,5	4,4	5,4	3,7	4,5	6,0
MANOVA	Ensemble	7,0	4,3	3,7	4,2	4,3	3,7	4,2	4,3
	Urbaines	7,3	3,8	4,1	4,0	3,9	4,1	4,1	4,5
	Non urbaines	6,0	5,4	2,4	4,7	5,3	2,7	4,7	4,4
Incertitude	Ensemble	3,1	2,2	3,0	5,3	7,2	3,1	4,9	6,8
	Urbain	2,2	1,2	3,4	5,6	7,6	3,5	5,3	7,0
	Non urbaines	5,7	5,3	2,0	4,4	6,0	2,0	3,9	6,3

## 6.5 Proposition d'un nouveau maillage pour la cartographie de la structure urbaine

Les résultats de la section précédente ont montré que selon le type d'espace considéré, urbain ou non urbain, les unités d'analyse utilisées pouvaient avoir une influence importante sur les performances des critères utilisés pour les évaluer. Le découpage en îlot pour les espaces urbains et le carroyage pour ceux non urbains sont utilisés. Pour le maillage régulier qui est utilisé pour représenter les espaces non urbains, nous avons choisi celui hexagonal qui est intéressant pour l'analyse de la trame verte urbaine par sa facilité à considérer la connectivité entre les mailles et permet de plus une meilleure visualisation cartographique (Carr *et al.*, 1992). Toutefois, il est nécessaire au préalable de pouvoir différencier ces deux types d'espaces. Le découpage en parcelles morphologiques est donc ici utilisé.

En effet lors de sa construction une zone a tout d'abord été déterminée par une succession de zones tampons pour donner une extension maximale aux parcelles (§ 6.3.2). Cette surface se rapproche ainsi de la construction de la tache urbaine proposée par Lorient (2008) à la différence qu'ici toutes les zones autour des bâtiments ont été conservées et



non uniquement les taches urbaines les plus importantes et que les trous n'ont pas été remplis. Afin de construire ce découpage, un modèle a été réalisé avec le logiciel ArcGis et dont les grandes étapes vont être présentées.

Tout d'abord, les îlots majoritairement inclus dans la tache urbaine définis par les parcelles morphologiques ont été conservés sans modifications. Pour les autres, différents traitements ont eu lieu. L'opération précédente a laissé certains éléments de la tache urbaine et donc certains bâtiments y appartenant non inclus à des îlots. Ainsi, lorsque les différents patchs restants sont composés de sept bâtiments ou plus alors un nouvel îlot est formé. Ce seuil a été choisi, car il correspond au mode du nombre de bâtiments dans les îlots appartenant à une classe urbaine. Lorsque ce seuil n'est pas atteint, alors les parcelles morphologiques restantes sont fusionnées, si cela est possible, à l'îlot avec lequel elle partage la plus longue bordure en commun. L'ensemble de cette procédure permet ainsi de délimiter les mailles urbaines. Toutefois certaines parcelles morphologiques peuvent encore rester si elles ne sont pas contiguës à un îlot urbain et sont alors considérées dans l'espace non urbain.

La procédure pour le maillage des espaces non urbain est plus simple. Le carroyage de forme hexagonale a été utilisé et intersecté avec les limites des îlots qui appartiennent à ces espaces. L'objectif de cette intersection est de pouvoir conserver les îlots délimités par un linéaire de réseau très dense et qui correspondent donc principalement à ceux des infrastructures de transports. Lié à cette intersection, de nombreuses mailles de petites dimensions peuvent être créées et comme lors de la construction des îlots un seuil de 1500 m<sup>2</sup> a été utilisé pour les fusionner à ceux voisins (§ 6.3.3). La différenciation entre les types d'espaces a été conservée afin d'obtenir pour le site de Strasbourg 4834 mailles urbaines et 16 281 mailles non urbaines (figure 6.7). L'ensemble des traitements a été intégré dans un modèle afin d'être réappliqué sur le site de Rennes permettant pour ce site d'obtenir 3141 de mailles urbaines et de 14 096 mailles non urbaines (figure 6.8).

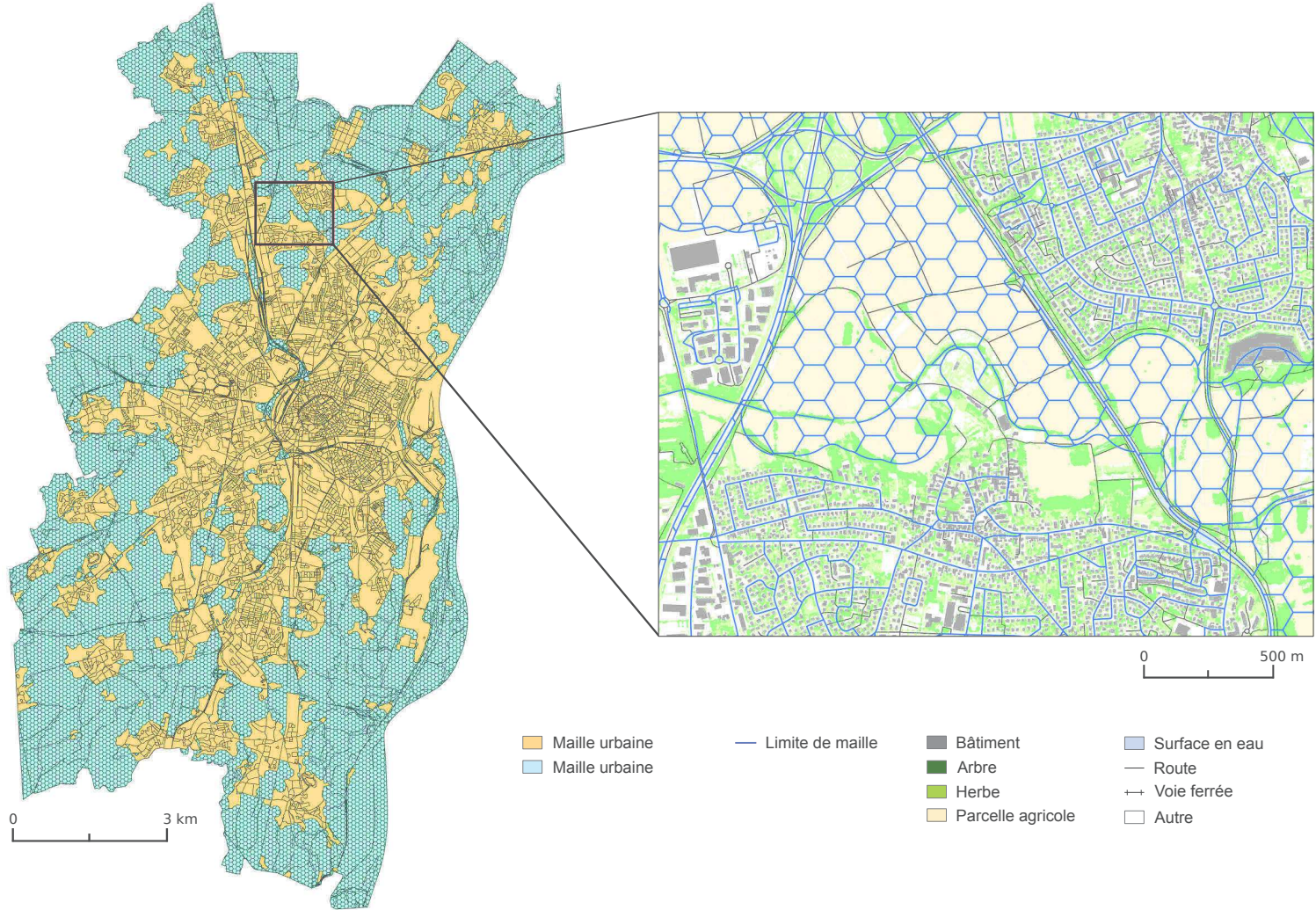


Figure 6.7 – Découpage construit pour l'analyse de la structure urbaine de Strasbourg

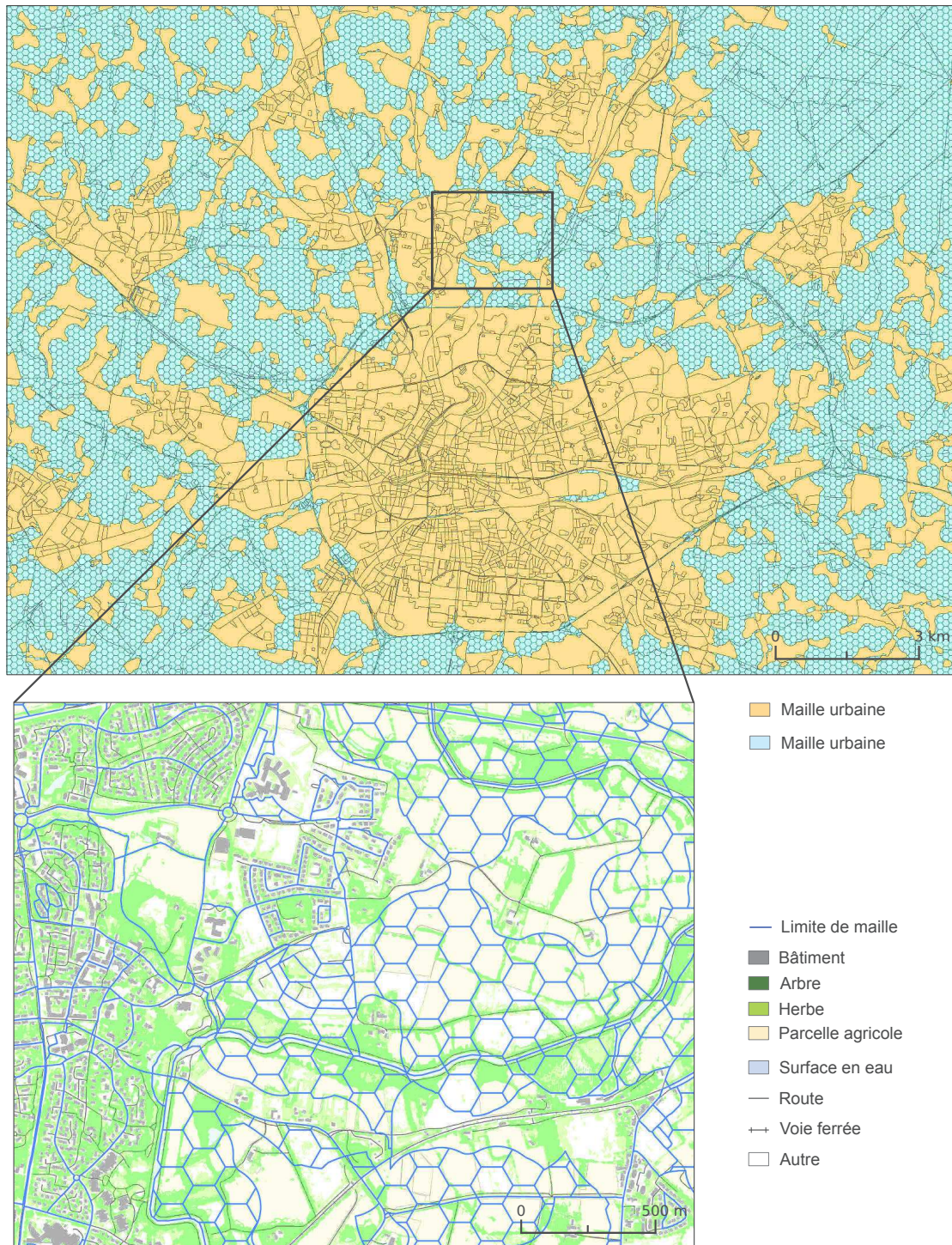


Figure 6.8 – Découpage construit pour l'analyse de la structure urbaine de Rennes

## Conclusion

Ce chapitre a permis de constituer une base de données à grande échelle et multi-sources complétée par des données auxiliaires pour analyser quantitativement et cartographier la structure urbaine. Elle s'appuie sur une description des éléments d'occupation du sol qui viennent de sources différentes. Pour les éléments gris, les bâtiments sont issus de la BD TOPO® et de données cadastrales, les réseaux routiers et ferrés proviennent de la BD TOPO®. Les éléments verts et bleu proviennent eux de la classification d'image Pléiades pour les végétations arborée et herbacée, du RPG pour les parcelles agricoles et de la BD TOPO® pour les surfaces en eau. À ces différents éléments d'occupation du sol, des caractéristiques sociodémographiques ont été ajoutées provenant des données carroyées sur la population de l'INSEE.

Plus de 100 indicateurs ont été dérivés. Ceux-ci sont pour la plupart issus de l'urbanisme ou de l'écologie du paysage et permettent de transcrire la morphologie des éléments ainsi que leur organisation spatiale au sein d'une maille. D'autres indicateurs ont été proposés afin d'aller au-delà de la caractérisation morphologique des tissus. Ils concernent l'accessibilité et la visibilité des bâtiments, les caractéristiques sociodémographiques des tissus ainsi que des informations contextuelles dérivées de mesure de centralité de graphe pour mieux localiser les unités d'analyse au sein de nos agglomérations.

Un travail sur la recherche de l'unité d'analyse la plus adaptée a également été effectué. La parcelle morphologique, l'îlot, et six carroyages de formes (carrée et hexagonale) et de tailles différentes (125, 250 et 500 m) ont été testés et évalués à travers cinq critères d'évaluation (la fragmentation des éléments, l'analyse de variance intra et inter-classe, l'analyse de variance multi-dimensionnelles et l'incertitude). Les résultats de cette évaluation ont montré que l'unité d'analyse adaptée doit être différente selon le type d'espace considéré : urbain ou non urbain. Ainsi un nouveau découpage a été proposé permettant de tenir compte de cette spécificité. L'îlot est choisi pour décrire les espaces urbains et la maille hexagonale de 125 m pour les espaces non urbains.

Le chapitre suivant s'attache donc à utiliser les informations produites dans ce chapitre afin de proposer une méthode reproductible et la plus automatique possible afin de cartographier de manière opérationnelle les tissus urbains.

*Valorisations scientifiques associées à ce chapitre :*

Nabucet J., **Rougier S.**, Deniau J., Vétillard L., Hanson E., Benarchid O., Wolff E., Hubert-Moy L., Puissant A., 2015, « Multi-scale methodology to map grey and green structures in urban areas using Pléiades images and existing geographic data », *Revue Française de Photogrammétrie et de Télédétection*, n° 209, pp. 95–101

**Rougier S.**, Nabucet J., Thomas A., Hanson E., Wolff E., Hubert-Moy L., Puissant A., 2014, « Grey and green infrastructure in urban areas based on Pléiades image analysis combined with existing vector database », *Pléiades Days*, Toulouse, France, 1–3 avril 2014

# Chapitre 7

## Application de méthodes d'extraction de connaissances

### Sommaire

---

Introduction . . . . .	204
7.1 Analyse des relations entre les variables . . . . .	205
7.2 Méthodes de regroupement des données . . . . .	208
7.2.1 Sélection d'une méthode de clustering . . . . .	209
7.2.2 Le problème du nombre de clusters . . . . .	215
7.3 Classification de la structure urbaine de Strasbourg . . . . .	216
7.3.1 Classification des indicateurs et interprétation des variables synthétiques . . . . .	216
7.3.2 Classification non-supervisée des mailles fondée sur les variables synthétiques . . . . .	225
7.3.3 Évaluation de la classification de Strasbourg . . . . .	238
7.4 Classification de la structure urbaine de Rennes . . . . .	244
7.4.1 Application des paramètres des classifications de Strasbourg sur Rennes . . . . .	244
7.4.2 Évaluation de la classification des mailles de Rennes . . . . .	245
Conclusion . . . . .	250

---

## Introduction

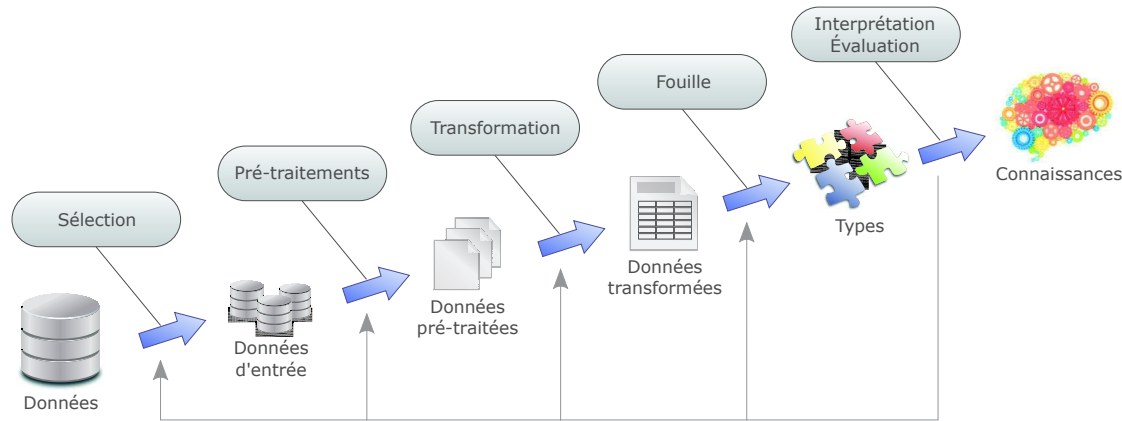
L'objectif est de cartographier de manière opérationnelle la structure urbaine de Strasbourg et de Rennes à partir de la base de données multi-sources et multi-indicateurs produite au chapitre précédent et fondée sur un découpage original combinant un maillage en îlots pour décrire les espaces urbains et un carroyage hexagonal de 125 m pour les espaces non urbains. Des méthodes d'extraction de connaissances dans des bases de données aussi appelées *Knowledge Discovery in Databases* (KDD) sont mises en œuvre dans ce chapitre pour répondre à cet objectif.

Ces méthodes permettent d'analyser des bases de données qui sont de plus en plus grandes, composées de nombreux individus et de nombreuses variables, avec l'idée sous-jacente que la plupart des informations qui y sont contenues sont cachées. Leur objectif est d'extraire à partir de ces données des connaissances qui doivent être « valides », « nouvelles », potentiellement utiles et compréhensibles facilement par un expert du domaine auquel ce processus est appliqué (Fayyad *et al.*, 1996). La validité signifie que les connaissances extraites ne sont pas dues au hasard ou à une anomalie, mais peuvent être retrouvées dans d'autres bases de données de même type et donc considérées comme pertinentes.

L'extraction de connaissances à partir de données est un processus complexe d'opérations qui peut se résumer en cinq grandes étapes (figure 7.1). Elle est souvent confondue avec la fouille de données (*Data Mining*). La fouille consiste à analyser de grands jeux de données, d'identifier des types (*pattern*), des règles ou des tendances dans les données en utilisant différentes méthodes telles que des algorithmes de classification, supervisée ou non supervisée (*clustering*) (Tan *et al.*, 2005). Elle ne représente ainsi qu'une partie d'un processus d'extraction de connaissances (Fayyad *et al.*, 1996). Parmi les autres étapes de ce processus se trouvent la sélection des données, leur préparation, l'intégration de connaissances existantes ainsi qu'une interprétation adéquate des types issus des techniques de fouilles. Ce type de processus n'est pas linéaire, mais est itératif et peut nécessiter d'être repris tout ou partie selon les objectifs de l'étude ou les connaissances de l'expert. Cette méthodologie associée à des techniques de statistique exploratoire est particulièrement bien adaptée à l'analyse de la structure des données que ce soit entre les variables ou entre les individus, nos mailles, et peut ainsi permettre de découvrir de nouveaux types à partir de la base de données qui a été constituée lors du chapitre précédent.

Appliquée aux bases de données spatiales, la recherche de connaissances se déroule pour Behnisch et Ultsch (2015) en cinq étapes. La première étape consiste à analyser, de manière univariée, les variables sélectionnées puis d'étudier les relations entre ces différentes variables par une analyse multivariée. Ce type de processus passe par l'analyse des corrélations, la définition de mesures de distances et des techniques de visualisation des variables et des individus pour évaluer leur proximité. La troisième phase est de trouver des types par des méthodes de clustering qui pourront ensuite être présentés

et expliqués par l'intermédiaire de méthodes d'analyse statistique. Enfin, ces classes nécessitent une interprétation réalisée par l'utilisateur qui peut passer par l'analyse de leur distribution spatiale, la transcription des résultats de l'étape précédente, la labélisation des groupes et enfin aboutir sur l'acquisition de nouvelles connaissances.



**Figure 7.1** – Processus d'extraction de connaissances dans des bases de données (d'après Fayyad *et al.* (1996))

Les méthodes qui peuvent être utilisées pour analyser les liens entre les variables et de clustering sont rappelées dans les sections 7.1 et 7.2 et les méthodes choisies y sont justifiées. Elles sont testées sur le site de Strasbourg (§ 7.3) et validées sur le site de Rennes (§ 7.3). Les méthodes sélectionnées et utilisées dans l'ensemble de ce chapitre sont appliquées avec le logiciel R. Ce logiciel présente l'avantage d'être libre permettant une forte reproductibilité du processus mis en place et de disposer de très nombreux packages développés par des spécialistes de divers domaines offrant ainsi une grande diversité de choix de méthodes.

## 7.1 Analyse des relations entre les variables

Dans une analyse exploratoire des données, spécifiquement lorsqu'elles sont de grandes dimensions (avec de nombreuses variables), une étape importante est d'étudier les relations entre les variables, de détecter d'éventuelles redondances ou des variables non informatives (Lebart *et al.*, 2006). Les méthodes d'analyse factorielle sont des outils privilégiés très souvent utilisés pour analyser ces relations. La méthode la plus utilisée est l'ACP qui est spécifiquement dédiée à l'analyse de variables quantitatives. Son objectif est de construire de nouvelles variables, les composantes ou axes factoriels, à partir de celles d'origine qui peuvent être corrélées. Par combinaison linéaire de ces différentes variables, l'ACP construit des facteurs dont l'objectif est de maximiser l'information contenue dans les données mesurée par la variance totale des individus dans l'espace des données. Chaque composante doit être orthogonale par rapport aux autres et le premier

axe est celui qui permettra donc de maximiser l'information, les suivantes en expliquant successivement de moins en moins. Par la construction d'axes orthogonaux maximisant la variance, elle permet aussi la visualisation des individus dans une projection en deux dimensions de données au départ en grandes dimensions et donc l'analyse visuelle des proximités entre ces individus.

Cette méthode a très largement été utilisée notamment pour étudier la morphologie urbaine afin de décrire les relations entre les variables, réduire la dimension des données et permettre de caractériser les structures urbaines. À partir de 46 indicateurs utilisés pour décrire les espaces résidentiels, Sahraoui *et al.* (2014) utilisent cette analyse pour ne retenir que quatre facteurs pour l'agglomération de Besançon et trois pour celle de Metz. Colaninno *et al.* (2011), grâce à cette méthode, en retiennent trois permettant de rendre compte respectivement de la densité, de l'efficacité et la complexité des formes bâties pour classer les tissus de Barcelone à l'échelle des bâtiments. Enfin Bonhomme (2013) utilise le premier axe de l'ACP obtenu à partir d'une sélection de 33 variables morphologiques décrivant les bâtiments, les espaces ouverts et la végétation arborée. Cette composante est combinée à la densité brute des surfaces bâties dans une analyse par maille, pour réaliser un arbre de décision simple lui permettant d'identifier cinq types de tissu urbain dans l'agglomération parisienne : pavillonnaires continus et discontinus, immeubles continus et discontinus et centre ancien.

Toutefois si cette analyse factorielle est très efficace pour réduire la dimensionnalité des données et discerner les relations entre les variables, l'analyse et l'interprétation des composantes latentes obtenues peut être difficile notamment lorsque le nombre de variables est très important (Vigneau et Qannari, 2003). La classification de variables a été proposée comme une alternative afin de simplifier l'interprétation de ces méthodes. L'objectif est de réaliser au préalable un regroupement entre les variables qui se ressemblent afin de créer des clusters homogènes de variables. Les méthodes les plus simples se basent sur la transformation d'une matrice de corrélation en une matrice de dissimilarité. Ces matrices de corrélation peuvent être mises au carré afin que soient regroupées dans un même cluster les variables positivement et négativement corrélées. À partir de ces mesures de similarités entre les différentes variables, différents algorithmes de classification non supervisés traditionnels peuvent être utilisés tels que les k-médoïdes ou un arbre de classification hiérarchique. Ces différentes méthodes de clustering sont davantage détaillées dans la section 7.2.1.

Le clustering de variables fait également l'objet de recherches et des méthodes de classification spécifiques ont été développées. Différents algorithmes peuvent être utilisés tels que les k-moyennes qui nécessitent de connaître le nombre de groupe *a priori* (Dhillon *et al.*, 2003) ou des arbres hiérarchiques (Vigneau et Qannari, 2003). La plus connue est certainement la procédure VARCLUS disponible dans le logiciel SAS qui se base sur un arbre hiérarchique (SAS Institute Inc., 2015). L'avantage d'utiliser des arbres pour la classification de variables est, pour l'utilisateur, de pouvoir déterminer le nombre de groupes présents dans ces données par l'analyse du dendrogramme. Vigneau *et al.* (2015) et Chavent *et al.* (2012) ont proposé chacun une méthode dédiée au clustering de



variables qui s'appuient toutes deux sur les algorithmes des k-moyennes ou des arbres hiérarchiques.

Chacune de ces deux méthodes a des avantages qui leur sont propres, mais toutes deux réalisent une analyse factorielle sur chacun des groupes de variables. Seule la première composante est utilisée pour résumer l'information contenue dans le groupe formé afin de constituer la variable synthétique. L'objectif pour ces deux méthodes est de maximiser à chaque étape de regroupement, dans le cas où un arbre hiérarchique est utilisé, un critère qui rend compte des relations entre les variables et leur composante latente associée. La méthode de Vigneau *et al.* (2015), tout d'abord, permet de prendre en compte deux méthodes pour évaluer la proximité entre les variables. La première, dite « directionnelle », ne prend pas en compte le sens des relations entre les variables, c'est-à-dire que peuvent se retrouver dans un même groupe des variables positivement ou négativement corrélées. Le critère à maximiser dans ce cas est le coefficient de corrélation au carré entre les variables observées et synthétiques. Le deuxième cas, dit « local », utilise simplement le coefficient de corrélation permettant ainsi de dissocier les groupes de variables négativement corrélées.

La méthode de Chavent *et al.* (2012) présente l'avantage de pouvoir être utilisé avec des variables quantitatives et qualitatives. Toutefois, elle ne permet pas de faire la différenciation entre méthode locale et directionnelle et seuls les carrés des coefficients de corrélation sont utilisés dans le cas des variables quantitatives. Pour les variables qualitatives, le rapport de corrélation qui mesure la part de la variance de la composante expliquée par les modalités de la variable qualitative est utilisé. Bien que cette méthode ait été proposée récemment, elle a déjà été utilisée régulièrement pour différentes thématiques telles que l'étude de la prise en compte de l'environnement par les agriculteurs (Kuentz-Simonet *et al.*, 2013), le lien entre épidémie de malaria et pisciculture en Amérique du Sud (Reis *et al.*, 2015), la cartographie à l'échelle mondiale de la densité d'arbre (Crowther *et al.*, 2015) ou pour construire des indicateurs de qualité de vie à l'échelle communale et étudier leur évolution (Kuentz-Simonet *et al.*, 2016).

Cette dernière méthode est proposée par ses auteurs dans R par l'intermédiaire du package `ClustOfVar`. Ce package offre de plus la possibilité d'utiliser les paramètres des analyses factorielles de chaque groupe de variables afin de les appliquer sur un autre jeu de données. Cette fonctionnalité est, pour cette recherche, très intéressante. En effet, cette méthode, appliquée sur le site de Strasbourg, peut être transférée sur celui de Rennes avec les mêmes paramètres afin de vérifier si cette classification et son interprétation peuvent être étendues et si les performances obtenues sur le premier site peuvent être équivalentes sur le deuxième. Dans ce cas, l'analyse et l'interprétation des résultats peuvent être facilitées du fait qu'elles ont déjà été menées sur un premier site et ainsi diminuer le temps nécessaire à leur réalisation. Ce dernier avantage a été un choix déterminant dans le choix de cette méthode pour l'analyse des relations et la réduction du nombre de variables.

Nous utilisons ici les avantages des deux algorithmes de classifications présents dans ce package afin de réaliser nos clusters de variables. La première que nous réalisons est l'arbre de classification hiérarchique. Cette méthode présente l'avantage d'être déterministe, c'est-à-dire que les résultats sont toujours les mêmes à la différence d'autres méthodes telle que les k-moyennes dont le résultat dépend des premiers centres choisis (Lebart *et al.*, 2006). De plus, l'utilisation d'un arbre permet d'apprécier le nombre de classes par l'étude du dendrogramme résultant. Comme il a été mentionné, cette méthode utilise le carré des coefficients de corrélation entre les variables et la première composante de l'ACP comme critère d'homogénéité pour les regrouper. Avec l'utilisation d'un arbre hiérarchique et à chaque étape de sa construction, deux variables ou deux groupes de variables sont fusionnés. Les deux variables ou classes qui sont fusionnées sont celles qui permettent de minimiser la perte d'homogénéité. La dernière étape de cet algorithme est atteinte lorsque l'ensemble des variables du jeu de données est dans un seul et unique groupe. L'analyse des résultats permet de définir le nombre de groupes pertinents afin qu'ils soient interprétables.

L'arbre hiérarchique ne permet toutefois pas de s'assurer que l'inertie expliquée est maximale. Ainsi, il peut être intéressant de réaliser un nouveau clustering par la méthode des k-moyennes à partir du découpage de l'arbre en un nombre de classes définies et des variables associées à chaque groupe. L'utilisation de cette procédure de consolidation permet d'augmenter l'inertie interclasses et donc d'optimiser le découpage réalisé à partir de l'arbre hiérarchique (Lebart *et al.*, 2006).

## 7.2 Méthodes de regroupement des données

Le regroupement ou clustering de données fait partie des méthodes de fouille et d'analyse exploratoire et peut être utilisé afin d'analyser la structure des données, trouver des groupes « naturels » dans ces données ou les compresser (Jain, 2010). Ce sont des méthodes de classification non supervisées dont l'objectif est de définir des ensembles d'individus homogènes, représentés par un ensemble de caractéristiques (Jain *et al.*, 1999). L'homogénéité est définie par une mesure de similarité entre les individus et ainsi chaque individu doit être plus proche des autres individus du même groupe que de ceux des autres groupes. L'avantage d'utiliser des méthodes de clustering est de rendre plus facilement interprétables les données par l'analyse des différentes classes obtenues (Lebart *et al.*, 2006).

Pour Jain *et al.* (1999) un processus de clustering se déroule en cinq étapes. La première est la représentation des individus qui passe par la sélection des variables permettant de les décrire (§ 6.2) et les prétraitements qui sont réalisés (§ 7.1). Une mesure de proximité entre les individus est ensuite définie qui est dépendante du domaine ou de la méthode de clustering utilisée. La troisième est le regroupement des données avec un algorithme qui aura été sélectionné (§ 7.2.1) qui peut nécessiter de choisir *a priori* le nombre de groupe (§ 7.2.2). Les deux étapes suivantes consistent à interpréter les groupes formés

(§ 7.3.2.2) et à évaluer leur validité (§ 7.3.3.1). Cette validation peut être réalisée, soit à partir de l'analyse des différents clusters par l'utilisateur, soit à partir d'indices internes purement statistiques qui vérifient la pertinence des groupes ou des indices externes qui vérifient l'adéquation du regroupement par rapport à une vérité terrain.

### 7.2.1 Sélection d'une méthode de clustering

Le clustering de données pour trouver des groupes naturels a donné naissance à de très nombreux algorithmes. Ces méthodes sont généralement séparées en deux grandes familles : les arbres hiérarchiques et les méthodes de partitionnement direct (Lebart *et al.*, 2006). Pour Nagpal *et al.* (2013) un troisième groupe existe qui intègre une grande diversité de méthodes telles que celles basées sur la densité, des grilles ou des modèles. Barbakh *et al.* (2009) utilisent comme précédemment ces deux familles, auxquels sont ajoutées les méthodes de partitionnement qui conservent la topologie. Ces dernières peuvent néanmoins être associées à celles de partitionnement et des méthodes hybrides combinant approches hiérarchique et de partitionnement ont aussi été proposées (Lebart *et al.*, 2006 ; Candillier, 2006).

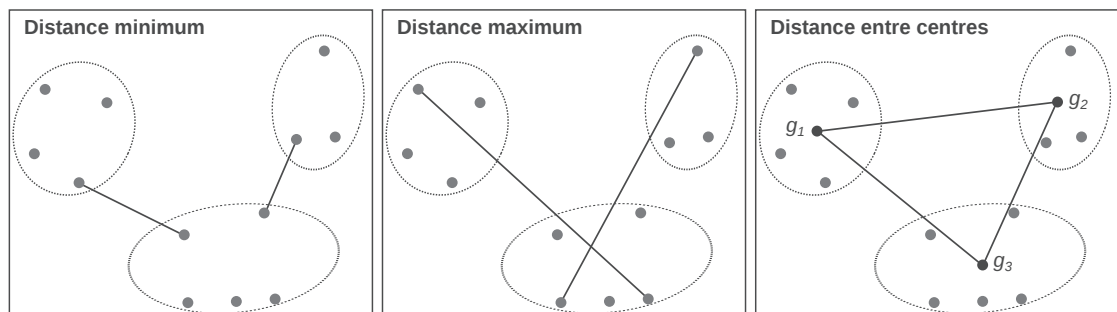
#### *Les classifications hiérarchiques*

Les méthodes hiérarchiques de regroupement d'individus permettent de créer un arbre qui est la suite d'un processus d'agrégation dans le cas d'une démarche ascendante entre deux éléments, individus ou groupes, jusqu'à ce qu'il ne reste plus qu'un groupe. À l'inverse, un algorithme hiérarchique descendant procède à des divisions successives partant donc d'un seul groupe jusqu'à ce que chaque observation forme une classe. Les classifications ascendantes hiérarchiques (CAH) sont les plus diverses et les plus utilisées et sont donc les seules présentées ici.

Afin de constituer des groupes homogènes, il est nécessaire de mesurer la dissimilarité, la distance entre les éléments qu'ils soient des individus ou des groupes. Ainsi la première étape de ces méthodes est de calculer une matrice de dissimilarité entre les individus. Cette matrice est généralement construite à partir de la distance euclidienne entre les individus, mais de nombreuses autres mesures de distances ou de pseudo-distance peuvent être utilisées selon le type de données (quantitatives, qualitatives, histogramme, etc.) ou le type d'application (Jain *et al.*, 1999).

Le premier regroupement réalisé est donc celui des deux individus les plus proches, mais le problème se pose lorsqu'il s'agit par la suite de regrouper un individu et un groupe ou deux groupes. Plusieurs critères peuvent être utilisés (figure 7.2) tels que le saut minimal qui utilise la distance minimale entre les éléments de deux groupes, le saut maximal où c'est la distance maximale entre les individus qui est utilisée, la distance moyenne entre l'ensemble des individus des groupes, le centre de gravité des groupes ou le critère de Ward qui vise à maximiser la variance inter-groupe (Lebart *et al.*, 2006).

L'avantage de ce type de méthode est de créer une hiérarchie qui permet par l'analyse de l'arbre et des sauts lors de la fusion des éléments de fixer *a posteriori* le nombre de classes. Toutefois, ces méthodes ne sont pas adaptées aux grands tableaux de données du fait de leurs complexités puisque les distances entre chaque paire d'éléments doivent être calculées et nécessitent toujours que l'ensemble de l'arbre soit construit (Candillier, 2006). Ce type de méthode n'est donc pas retenu dans cette thèse.



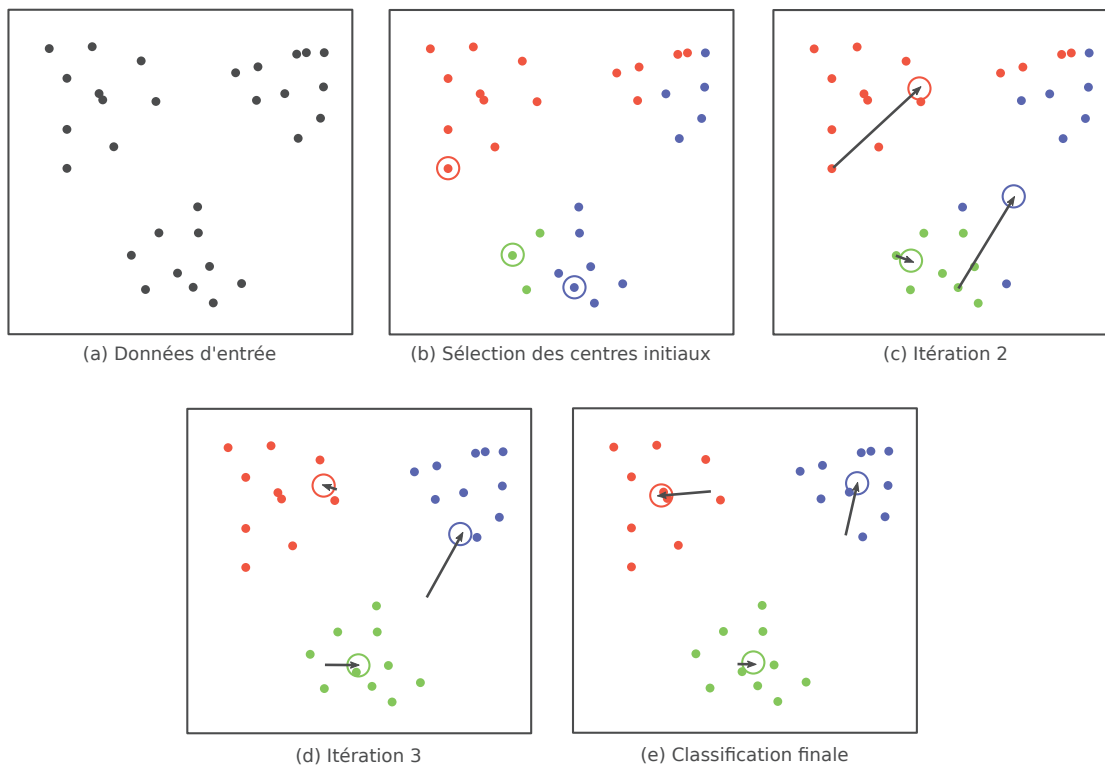
**Figure 7.2** – Exemple de critères d'agrégation d'une classification ascendante hiérarchique (de Groupe ElementR (2014))

### *Les méthodes de partitionnement direct*

Si les classifications hiérarchiques sont peu utilisées (Sahraoui *et al.*, 2014), les méthodes de partitionnement direct sont elles souvent employées pour l'analyse des espaces urbains à l'échelle du tissu (Long et Kergomard, 2005 ; Wemmert *et al.*, 2009 ; Wurm *et al.*, 2010 ; Bernabé *et al.*, 2013). Elles sont en effet bien adaptées à la recherche de groupes dans de grandes bases de données en raison de leur rapidité et de leurs faibles consommations d'espaces mémoire (Lebart *et al.*, 2006). Il existe une très grande variété d'algorithmes qui font partie des méthodes de partitionnement. L'ensemble de ces méthodes ne peut être présenté ici et seules celles des *k*-moyennes sont explicitées du fait qu'elles sont utilisées dans cette thèse pour la classification des variables et que la méthode utilisée pour la création des groupes constitue une évolution de ce type de méthode. Il est néanmoins possible de trouver dans Jain *et al.* (1999) ou dans Candillier (2006) des listes plus complètes rendant compte de la diversité des méthodes de partitionnement.

La méthode des *k*-moyennes est la méthode de classification non supervisée la plus utilisée et de nombreuses implémentations existent, mais leurs principes restent similaires (MacQueen, 1967 ; Hartigan et Wong, 1979 ; Lloyd, 1982). Elle nécessite dans un premier temps de fixer le nombre *k* de classes que l'on souhaite obtenir. La première étape de l'algorithme est de placer les *k* centres de classe de manière aléatoire dans l'espace des données. Les individus de la base sont ensuite affectés à un de ces points, celui duquel il est le plus proche en utilisant la distance euclidienne permettant ainsi de

minimiser la variance intra-classe. Les différents centres des classes sont ensuite déplacés pour devenir les centroïdes des points qui y sont associés. Les deux étapes d'affectation des points au centre de gravité puis de déplacement des centres sont ensuite itérées jusqu'à ce que l'algorithme converge, que les centres ne bougent plus ou presque plus ou qu'un nombre d'itérations maximum soit atteint. La figure 7.3 illustre le principe de fonctionnement de l'algorithme des k-moyennes. L'avantage de cette méthode est qu'elle est très simple et très efficace, mais elle est par contre très sensible à l'initialisation du fait du placement aléatoire des premiers centres de classes. Pour remédier en partie à ce défaut, il est possible de réaliser plusieurs fois le processus de regroupement par les k-moyennes afin de retenir uniquement les classes qui sont stables, dont les individus ont toujours été affectés dans les mêmes.



**Figure 7.3** – Principe de l'algorithme des k-moyennes. (a) Données d'entrée en deux dimensions avec trois clusters, (b) sélection de trois centres initiaux et affectation des points aux clusters, (c) et (d) itérations intermédiaires de mises à jour des points et des centres, (e) convergence de l'algorithme et classification finale (d'après Jain (2010))

De nombreuses autres méthodes de partitionnement utilisant le principe des k-moyennes ont été proposées. Isodata par exemple s'apparente à cette technique et permet à l'utilisateur de choisir un nombre maximum et minimum de groupes à former permettant ensuite à cette méthode de supprimer des groupes d'effectifs trop faibles ou scinder ceux dont le diamètre est trop grand (Ball et Hall, 1965). Les k-moyennes utilisent un noyau très simple pour l'affectation des individus à la classe, ici la moyenne, mais d'autres, plus

complexes, peuvent être utilisés comme pour l'algorithme des nuées dynamiques qui va utiliser un sous-ensemble d'individus de chaque classe qui seront considérés comme représentatifs (Diday, 1971). Des méthodes similaires ont aussi été développées afin de pouvoir être utilisées avec d'autres mesures de distances telles que celle des k-médoïdes qui sélectionne l'individu le plus central de chaque classe et ainsi les individus les plus proches de ce centre seront affectés à cette classe (Kaufmann et Rousseeuw, 1987). Tout comme les k-moyennes cet algorithme est itéré jusqu'à convergence et à l'avantage d'être moins sensible aux valeurs aberrantes.

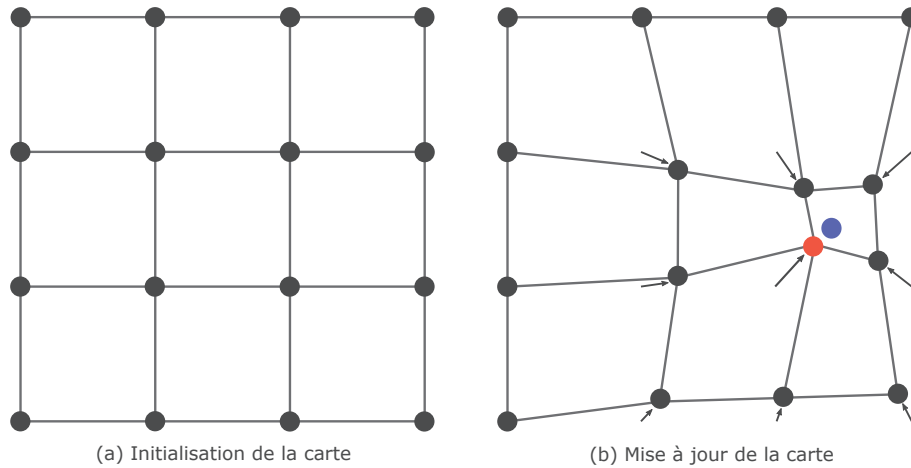
### *Les méthodes de partitionnement conservant la topologie*

Les méthodes de partitionnement permettant de conserver la topologie sont aussi très utilisées pour ce type d'analyse. La méthode la plus connue est celle des cartes auto adaptatives de Kohonen (SOM – *Self-Organizing Maps* ; Kohonen, 1982). Les SOM font partie de la famille des réseaux de neurones artificiels et sont une variante de l'algorithme des k-moyennes qui permettent de conserver la topologie entre les neurones, c'est-à-dire une notion de voisinage entre les classes.

Une carte représentant l'espace des données est tout d'abord créée. Cette carte est généralement matérialisée en une ou deux dimensions et les mailles sont de forme carrée et déformable au cours du processus, mais il est possible d'utiliser des mailles de forme hexagonale, utiliser une carte cylindrique et non plane, de la représenter avec une troisième dimension (Kohonen, 2013). Chaque maille de cette carte correspond à un neurone du réseau et les relations de voisinage entre ces différents neurones sont considérées et conservées pendant toute la réalisation de cette méthode. Le nombre de neurones ainsi que leur configuration sont fixés *a priori*, chaque neurone représentant une classe en sortie. Les observations composant le jeu de données vont être présentées une à une à cette carte. Cet exemple va être attribué au neurone dont il est le plus proche dans l'espace des données et va permettre de déplacer ce neurone qui verra ses valeurs changées afin qu'il se rapproche de ce nouvel exemple comme le ferait un centre de l'algorithme des k-moyennes. Par contre pour les SOM, les autres neurones sont aussi déplacés, mais de manière moins importante et décroissante selon la distance de ces neurones à celui où il a été attribué (figure 7.4). L'algorithme est itéré et le poids attribué pour le déplacement des neurones est diminué au cours des itérations afin que la carte et ses neurones restent stables et permettent de couvrir l'ensemble de l'espace des données.

Les avantages de cette méthode sont qu'elle permet de visualiser directement les proximités entre les classes, qu'elle peut être utilisée avec tout type de variable en permettant aussi leur représentation graphique pour chaque neurone (Rousset *et al.*, 2005). Notamment pour des données géographiques les SOM surpassent la méthode des k-moyennes et sont moins influencés par l'initialisation et permettent donc d'obtenir des résultats plus stables (Bação *et al.*, 2005b). Toutefois, le fait que le voisinage entre les neurones est

fixe et les relations ne peuvent être modifiées ou cassées durant le processus pour rendre compte de données discontinues et que les SOM nécessitent des connaissances *a priori* sur la structure topologique des données apparaît comme des limites à l'utilisation de cette méthode.



**Figure 7.4** – Principe de l'algorithme SOM. (a) Création de la carte initiale de forme carrée, (b) un exemple est proposé au réseau de neurones (point bleu), le gagnant (point rouge) et les neurones proches sont déplacés vers ce nouvel exemple (d'après Vesanto *et al.* (1999))

L'algorithme des gaz neuronaux (NG – *Neural Gas*) a été proposé comme une alternative aux SOM par Martinetz *et al.* (1993) afin de résoudre ce problème des liens entre les neurones en utilisant un réseau plus flexible. Le principe est le même que pour les SOM sauf que les neurones ne sont pas directement connectés. La notion de voisinage et les changements que l'ajout de nouveaux exemples implique sur les neurones connectés sont pris en compte par un poids qui décroît en fonction de la distance relative entre eux. Si cette méthode permet de représenter la structure de données complexes, le fait que les neurones ne soient pas connectés ne permet pas de voir directement les relations entre les classes.

L'algorithme du gaz neural croissant (GNG – *Growing Neural Gas*) est une évolution de NG qui intègre par une représentation en graphe les relations topologiques entre les neurones (Fritzke, 1995). Dans cette méthode, le nombre de neurones est augmenté au cours de l'apprentissage au fur et à mesure que de nouveaux exemples sont présentés au réseau tout en permettant de fixer un nombre maximum de neurones. Mais ils peuvent aussi être supprimés lorsqu'ils sont trop isolés du réseau. La création d'un neurone se fait entre deux neurones voisins lorsque l'erreur cumulée pour représenter les données entre ces neurones est trop importante. Les connexions entre les neurones peuvent aussi être modifiées. Ils sont créés au cours de l'apprentissage quand un exemple est attribué à un neurone entre celui-ci et son concurrent direct, le deuxième le plus proche de la donnée d'entrée. Ces liens sont supprimés quand ils n'ont pas été renouvelés pendant

un certain temps. Ce réseau permet donc de rendre explicites les relations topologiques entre les classes représentées par un graphe.

Il est important de noter que ces réseaux de neurones ont connu certaines évolutions afin de donner un poids spatial aux données. Pour les cartes auto adaptatives, Bação *et al.* (2005a) ont proposé les GeoSOM. La première étape de construction de la carte va être réalisée non pas dans l'espace des données, mais en respectant la topologie géographique. Un rayon maximum de recherche est défini que chaque exemple ne pourra dépasser pour trouver le neurone dont il est le plus proche dans l'espace des données. Cette méthode a déjà été utilisée par Hamaina *et al.* (2012) afin de reconstruire les tissus urbains de Nantes. Plusieurs contraintes géographiques ont aussi été proposées pour l'algorithme NG qui peuvent être, comme pour GeoSOM, une notion de degré maximum entre les neurones, mais aussi un poids donné aux distances géographiques ou par le biais de matrice de contiguïté entre les objets (Hagenauer et Helbich, 2013 ; Hagenauer, 2015). Ces méthodes sont intéressantes pour obtenir des régions compactes et offrent ainsi la possibilité d'une généralisation cartographique. Elles ne sont pas considérées dans cette thèse, car elles ne permettent pas d'obtenir des classes réellement homogènes. Le tableau 7.1 résume les avantages et les inconvénients des différentes méthodes présentés dans cette section.

**Tableau 7.1** – Récapitulatif des méthodes de clustering

Méthode	Avantages	Inconvénients
CAH	<ul style="list-style-type: none"> <li>• Déterministe</li> <li>• Choix du nombre de classes par analyse du dendrogramme</li> <li>• Visualisation de la proximité entre les classes</li> </ul>	<ul style="list-style-type: none"> <li>• Très coûteuse en temps de calcul et en mémoire</li> </ul>
k-moyennes	<ul style="list-style-type: none"> <li>• Adaptée aux gros volumes de données</li> <li>• Rapidité d'exécution</li> </ul>	<ul style="list-style-type: none"> <li>• Nombre de classes fixé <i>a priori</i></li> <li>• Très sensible à l'initialisation</li> </ul>
SOM	<ul style="list-style-type: none"> <li>• Adaptée aux gros volumes de données</li> <li>• Rapidité d'exécution</li> <li>• Visualisation de la proximité entre les classes</li> </ul>	<ul style="list-style-type: none"> <li>• Nombre de neurones fixé <i>a priori</i></li> <li>• Sensible à l'initialisation</li> <li>• Connaissances sur la structure des données pour l'initialisation de la carte</li> <li>• Réseau de neurones peu flexible</li> </ul>
NG	<ul style="list-style-type: none"> <li>• Adaptée aux gros volumes de données</li> <li>• Rapidité d'exécution</li> <li>• Réseau de neurones flexible</li> </ul>	<ul style="list-style-type: none"> <li>• Nombre de neurones fixé <i>a priori</i></li> <li>• Sensible à l'initialisation</li> </ul>
GNG	<ul style="list-style-type: none"> <li>• Adaptée aux gros volumes de données</li> <li>• Rapidité d'exécution</li> <li>• Réseau de neurones flexible</li> <li>• Visualisation de la proximité entre les classes</li> </ul>	<ul style="list-style-type: none"> <li>• Nombre de neurones fixé <i>a priori</i></li> <li>• Sensible à l'initialisation</li> </ul>

Pour la classification de la structure urbaine, notre choix s'est porté sur la méthode de partitionnement direct conservant la topologie GNG qui permet de traiter rapidement de



gros volume de données et de voir les proximités entre les classes. De plus, cet algorithme a été intégré à R par l'intermédiaire du package `gmum.r` avec une implémentation optimisée qui permet ainsi son utilisation sur de grandes bases de données avec un temps d'exécution très faible (Fišer *et al.*, 2013). Avec les méthodes de partitionnement direct tel que GNG, un problème est le nombre de classes ou de neurones qu'il faut sélectionner pour refléter au mieux la structure intrinsèque des données. La section suivante s'attache à montrer la diversité des méthodes permettant de guider, par des critères statistiques, le choix du nombre de clusters.

## 7.2.2 Le problème du nombre de clusters

Le choix du nombre de groupes en partitionnement des données est un des problèmes les plus difficiles à résoudre (Sugar et James, 2003). En effet, de nombreuses méthodes nécessitent de connaître ce nombre *a priori* afin de pouvoir être réalisées, comme pour l'algorithme des k-moyennes, mais aussi le nombre de neurones pour ceux des gaz neuronaux. De nombreux indices ont donc été développés pour déterminer ce nombre plus ou moins automatiquement ou pour évaluer la qualité du partitionnement réalisé. Généralement, l'algorithme de regroupement choisi est lancé sur les données avec différents nombres de groupes et le découpage permettant d'obtenir la valeur optimale d'un indice prédéfini est considéré comme adéquat (Jain, 2010). Le package R `NbClust` par exemple regroupe 30 de ces indices (Charrad *et al.*, 2014). La plupart de ces indices vont combiner des mesures de variances intra et inter-groupes, des propriétés statistiques et géométriques des données ou des mesures de similarité/dissimilarité. Parmi les plus cités se trouvent l'indice de silhouette (Rousseeuw, 1987), celui de Davies-Bouldin (Davies et Bouldin, 1979), de Calinski et Harabasz (Calinski et Harabasz, 1974), de Dunn (Dunn, 1974) ou la statistique *gap* (Tibshirani *et al.*, 2001).

Dans cette thèse nous utilisons la statistique *gap* puisqu'elle montre de bonnes performances sur des données simulées ou réelles malgré sa tendance à surestimer le nombre de groupes présents dans les données (Dudoit et Fridlyand, 2002). Lors de l'analyse des groupes obtenus, certains peuvent en effet être fusionnés si nécessaire. Cette méthode compare le logarithme du carré de la variance intra-groupe à celle de distributions de référence obtenues par une simulation de Monte-Carlo pour chaque nombre de groupe. Ces distributions de référence sont uniformes dans le rectangle englobant l'espace des données des individus à regrouper. L'écart moyen ainsi que l'écart-type entre la distribution réelle et celles de références sont calculés. Le nombre de clusters sélectionné est le plus petit dont la valeur n'excède pas un écart-type du premier maximum local. Cette méthode est implémentée dans le package `cluster` de R.

## 7.3 Classification de la structure urbaine de Strasbourg

Cette section présente les résultats de notre démarche méthodologique, tout d'abord ceux issus de la classification et de l'analyse des variables (§ 7.3.1) La sélection du nombre de clusters ainsi que les résultats du clustering sont l'objet de la section 7.3.2. L'évaluation de la classification par des critères externes est présentée à la section 7.3.3.

### 7.3.1 Classification des indicateurs et interprétation des variables synthétiques

L'objectif de cette section est d'analyser les relations entre les indicateurs et de réduire leur nombre grâce à la méthode de classification de variables proposée par Chavent *et al.* (2012) et d'identifier des variables synthétiques permettant de rendre compte des spécificités des mailles pour l'analyse de la structure urbaine. L'analyse du dendrogramme produit par la classification ascendante hiérarchique sur les variables du package `ClustOfVar` est une aide permettant de définir le nombre de variables synthétiques adéquat (figure 7.5). La hauteur du dendrogramme correspond à la perte en homogénéité résultant de la fusion de deux clusters. Un saut important lors de la fusion de deux groupes de variables signifie donc que ces deux groupes sont très différents et qu'une découpe appropriée de l'arbre pourrait se situer avant ce regroupement. Toutefois, nous privilégions ici l'interprétabilité de chaque groupe de variables afin de leur donner du sens et non l'optimisation d'un critère statistique.

Ainsi, 14 variables synthétiques ont été retenues. Une procédure de consolidation avec la méthode des k-moyennes a ensuite été réalisée à partir des groupes d'indicateurs retenus avec le regroupement hiérarchique afin de s'assurer de maximiser l'homogénéité de chaque composante. Seule une variable a changé de groupe, la dimension fractale moyenne de la végétation herbacée (« herbe FRAC.mean »).

Les figures 7.6 et 7.8 représentent les corrélations entre les variables réelles et leur composante associée, c'est-à-dire les coordonnées de ces variables sur le premier axe de l'ACP et donc leurs contributions. Ces deux figures permettent donc d'interpréter chacune des composantes. Les figures 7.7 et 7.9 spatialisent ces variables synthétiques au niveau des mailles pour l'ensemble de l'Eurométropole de Strasbourg.

La **variable synthétique 1** montre une contribution très forte de l'aire et du périmètre des mailles ainsi que du degré d'une maille, son nombre de voisins, et dans une moindre mesure leur centralité de vecteur propre (figure 7.6 – VS1). Ces quatre variables sont aussi corrélées à la diversité des patches d'eau, tant au niveau de leurs formes que de leurs tailles. Ces patches, peu nombreux et rares, correspondent à des mailles où plusieurs de ces éléments sont présents et seules les très grandes mailles sont concernées. Cette première variable peut donc s'interpréter comme la dimension des mailles. Toutes les variables synthétiques sont centrées, ainsi une valeur positive pour ce premier groupe signifie que les mailles sont relativement grandes et fortement connectées. À l'inverse,

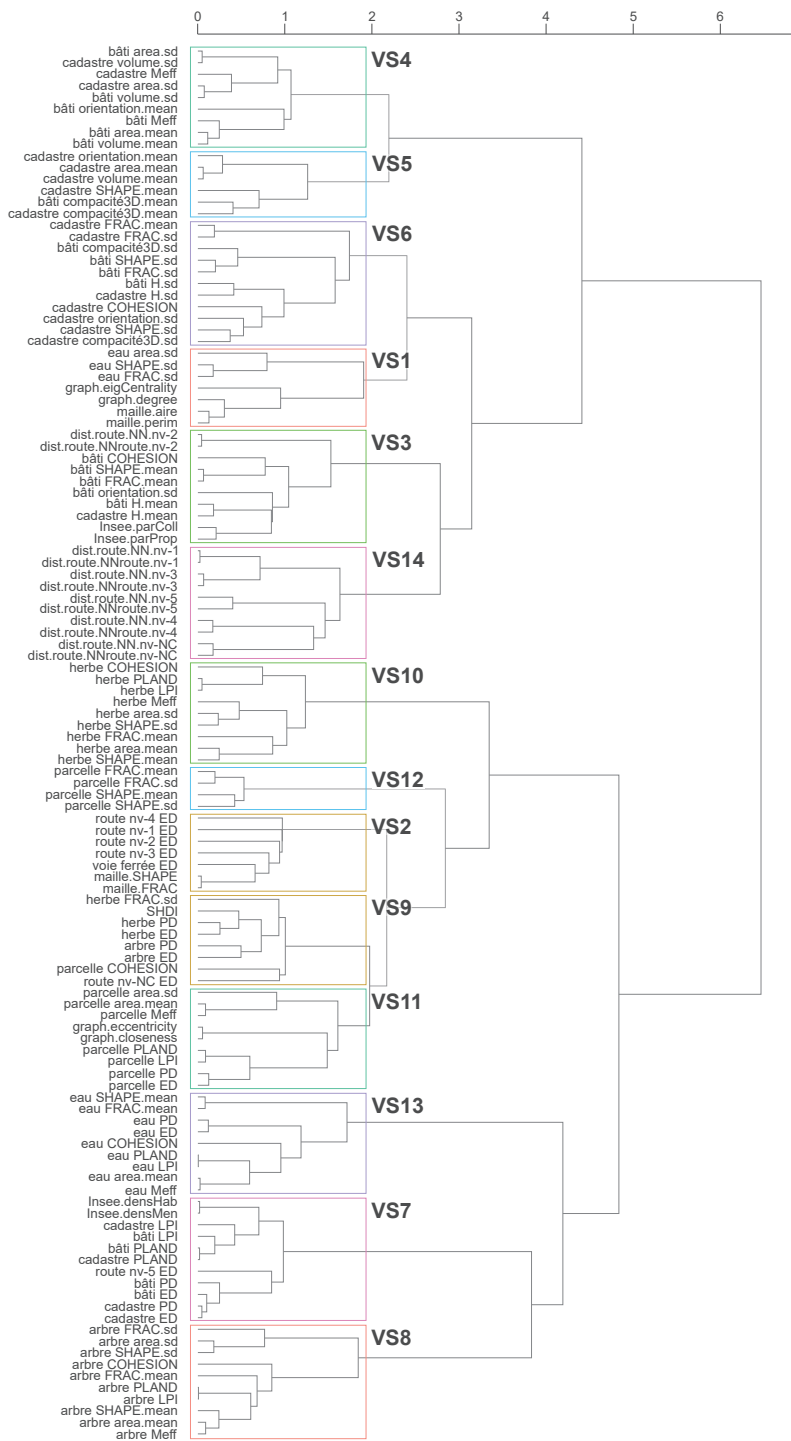


Figure 7.5 – Dendrogramme issu de la classification des 120 indicateurs

une valeur négative signifie que la dimension de la maille est petite. La cartographie de cette variable montre que les mailles les plus grandes concernent les espaces urbains, notamment celles situées sur les marges de ces espaces (figure 7.7 – VS1). Les valeurs faibles concernent principalement de petits îlots situés en centre-ville ou des mailles situées à proximité des réseaux et des bords de la zone d'étude.

La **variable synthétique 2** extraite concerne principalement la forme des mailles (figure 7.6 – VS2). Plus les valeurs de cette variable sont faibles plus les mailles apparaissent complexes. Les densités de réseaux, qu'ils soient routiers ou ferrés, y sont aussi très fortement corrélées. Les réseaux routiers de grande importance sont tous associés à cette variable synthétique, seuls les tronçons de route dont l'importance est 5 ou NC en sont absents. Cette variable est donc caractéristique des mailles de réseau ce qui apparaît de manière marquée sur la carte (figure 7.7 – VS2). Les grands îlots sont aussi, mais dans une moindre mesure, caractérisés par de faibles valeurs liées principalement à leur forme complexe et non à la densité de réseau.

La **variable synthétique 3** associe la hauteur des bâtiments agrégés et du cadastre à la complexité des formes et la cohésion des bâtiments uniquement ceux agrégés (figure 7.6 – VS3). De plus, elle permet de révéler que ces bâtiments complexes et hauts sont principalement des logements collectifs et que la part de propriétaire y est faible. En effet, s'oppose à ces variables la part de propriétaire ainsi qu'une distance importante des bâtiments au réseau de niveau 2. Cette variable synthétique présente donc des valeurs élevées pour les bâtiments hauts, complexes où la part de logement collectif est importante et des valeurs faibles où les mailles sont caractérisées par des bâtiments bas et simples, souvent occupés par des habitants propriétaires de leur logement et où la diversité de l'orientation des bâtiments est importante. Elle met en évidence une différenciation importante entre les espaces urbains, dont les plus centraux ont des valeurs les plus fortes avec une première couronne où elles sont moins élevées, et ceux périphériques caractérisés par des valeurs faibles (figure 7.7 – VS3). Les espaces non bâtis apparaissent eux dans la moyenne liés aux remplacements des valeurs manquantes par la moyenne.

La **variable synthétique 4** peut s'interpréter comme la dimension des bâtiments, principalement ceux agrégés. En effet, les variables caractéristiques de celle-ci sont l'aire et le volume des bâtiments agrégés qui sont corrélés positivement à cet axe. Les écarts-types de ces dimensions y sont importants (figure 7.6 – VS3). Les mêmes attributs pour les bâtiments du cadastre ne sont pas tous représentés sur cette composante. En effet, seuls les écarts-types de l'aire et du volume sont concernés et non les valeurs moyennes. Les bâtiments s'ils sont de grandes tailles quand ils sont agrégés ne le sont donc pas nécessairement aussi au niveau du cadastre, mais montrent dans cette base une grande diversité de taille et de volume pour les mailles concernées. La largeur effective de maille est aussi positivement corrélée montrant une fragmentation moins importante des bâtiments lorsque ceux-ci sont grands. La cartographie permet de mettre en évidence que les mailles où les valeurs de cette composante sont élevées sont principalement situées dans

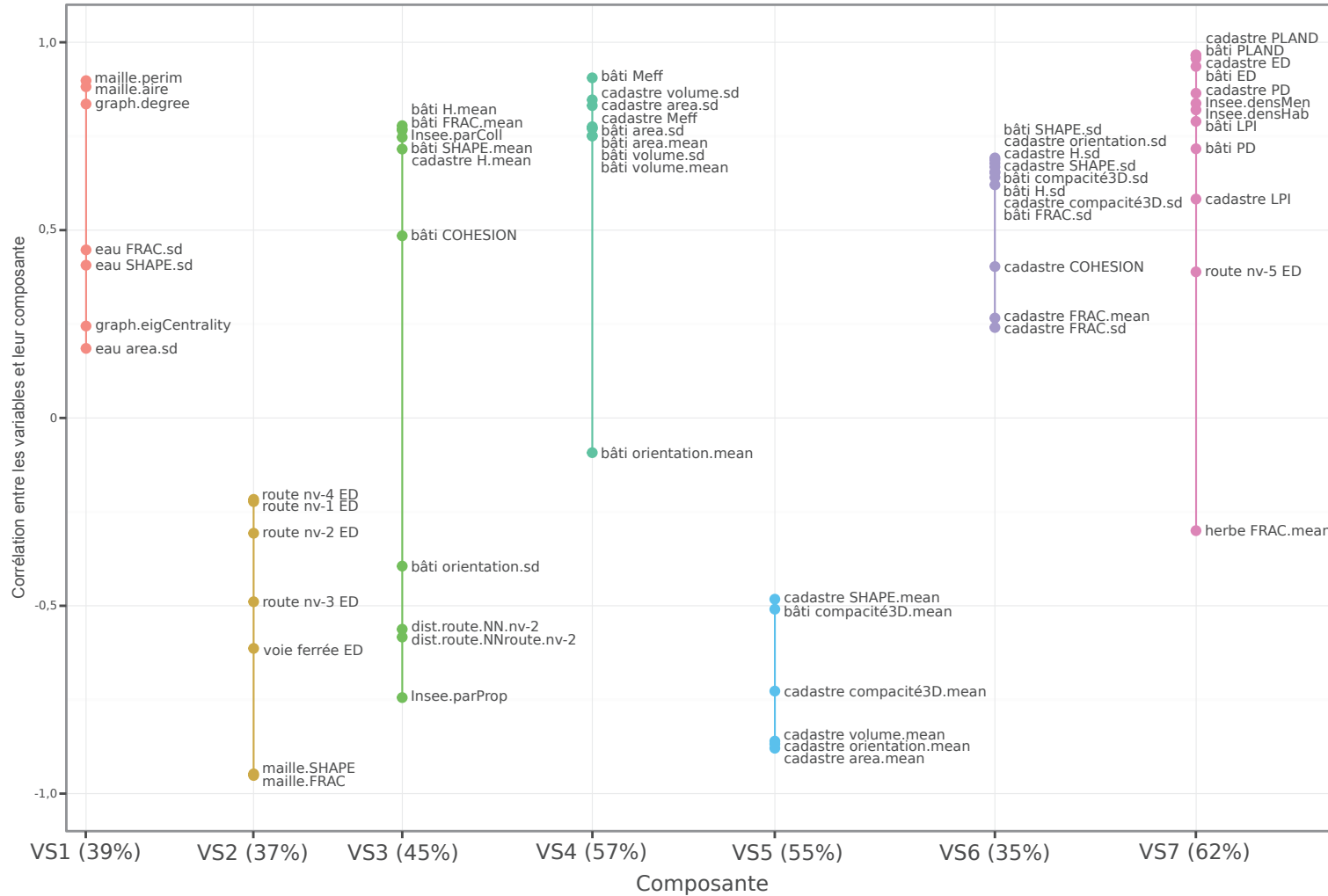
le centre, mais aussi en périphérie. Ces mailles correspondent à des zones spécialisées notamment industrielles ou commerciales (figure 7.7 – VS4).

La **variable synthétique 5** reprend en partie, mais de manière inversée, les conclusions de la précédente cette fois au niveau des bâtiments issus de la base de données cadastrale (figure 7.6 – VS5). En effet, elle permet de décrire l'aire et le volume des bâtiments du cadastre. Elle montre aussi que plus les dimensions de ces bâtiments sont grandes, plus leurs formes sont complexes avec des valeurs moyennes importantes pour les variables SHAPE et de compacité en 3 dimensions. Sur la cartographie associée, cette variable permet de montrer des valeurs faibles ou très faibles pour les mailles comprenant des espaces urbains spécialisés et des valeurs assez élevées où les bâtiments du cadastre sont de petites dimensions et simples, notamment pour les espaces périurbains où les logements individuels et pavillonnaires sont très présents (figure 7.7 – VS5).

La **variable synthétique 6** concentre de nombreuses variables montrant la diversité des bâtiments au sein d'une maille. En effet, seules la cohésion et la dimension fractale moyenne de ceux issus du cadastre ne représentent pas la dispersion de la forme, de la hauteur ou de la compacité des bâtiments dans les mailles (figure 7.6 – VS6). L'ensemble des variables est corrélé positivement à cette composante qui peut ainsi s'interpréter comme la diversité de forme et hauteur des bâtiments. La représentation spatiale de cette composante est plus difficilement interprétable que les précédentes (figure 7.7 – VS6). Si les valeurs les plus fortes sont situées au centre de l'agglomération, le centre ancien de Strasbourg est quant à lui très divers en terme de valeurs puisque certaines sont inférieures tandis que d'autres sont supérieures à la moyenne. Les valeurs les plus faibles concernent les bâtiments isolés au sein des espaces définis comme non urbains et les zones d'habitats pavillonnaires ont parfois des valeurs faibles, parfois des valeurs élevées.

La **variable synthétique 7** reflète de la dominance et de la densité du bâti au sein des mailles. Les variables rendant compte de la part du bâti totale et de son plus grand élément dans la maille ainsi que de la densité de patches et de lisière de bâtiments y sont fortement corrélées (figure 7.6 – VS7). Cette composante montre aussi que les densités d'habitants et de ménages sont très fortement corrélées à cette dominance et densité de bâti. De plus, le réseau routier de niveau 5 est aussi important lorsque les valeurs de cette composante sont élevées. Dans les mailles où la dominance du bâti est conséquente, la forme des éléments de végétation herbacée est plutôt simple comme le démontre la corrélation négative de la dimension fractale de l'herbe à cette variable synthétique. Les valeurs les plus importantes de cette variable se retrouvent dans le centre de Strasbourg et dans le quartier du Neudorf avec quelques centres de communes périphériques telles que Schiltigheim ou Bischheim (figure 7.7 – VS7). Peu de valeurs très inférieures à la moyenne sont révélées par la cartographie. En effet, l'ensemble des mailles non bâties reste compris dans la classe moyenne.

La **variable synthétique 8** concerne la dominance et la dimension des éléments arborés (figure 7.8 – VS8). La plupart des attributs permettant de décrire cette strate de végéta-



**Figure 7.6** – Corrélations entre les variables synthétiques (VS) 1 à 7 et les indicateurs associés (entre parenthèses : pourcentage d’inertie expliquée des variables réelles par la variable synthétique associée)

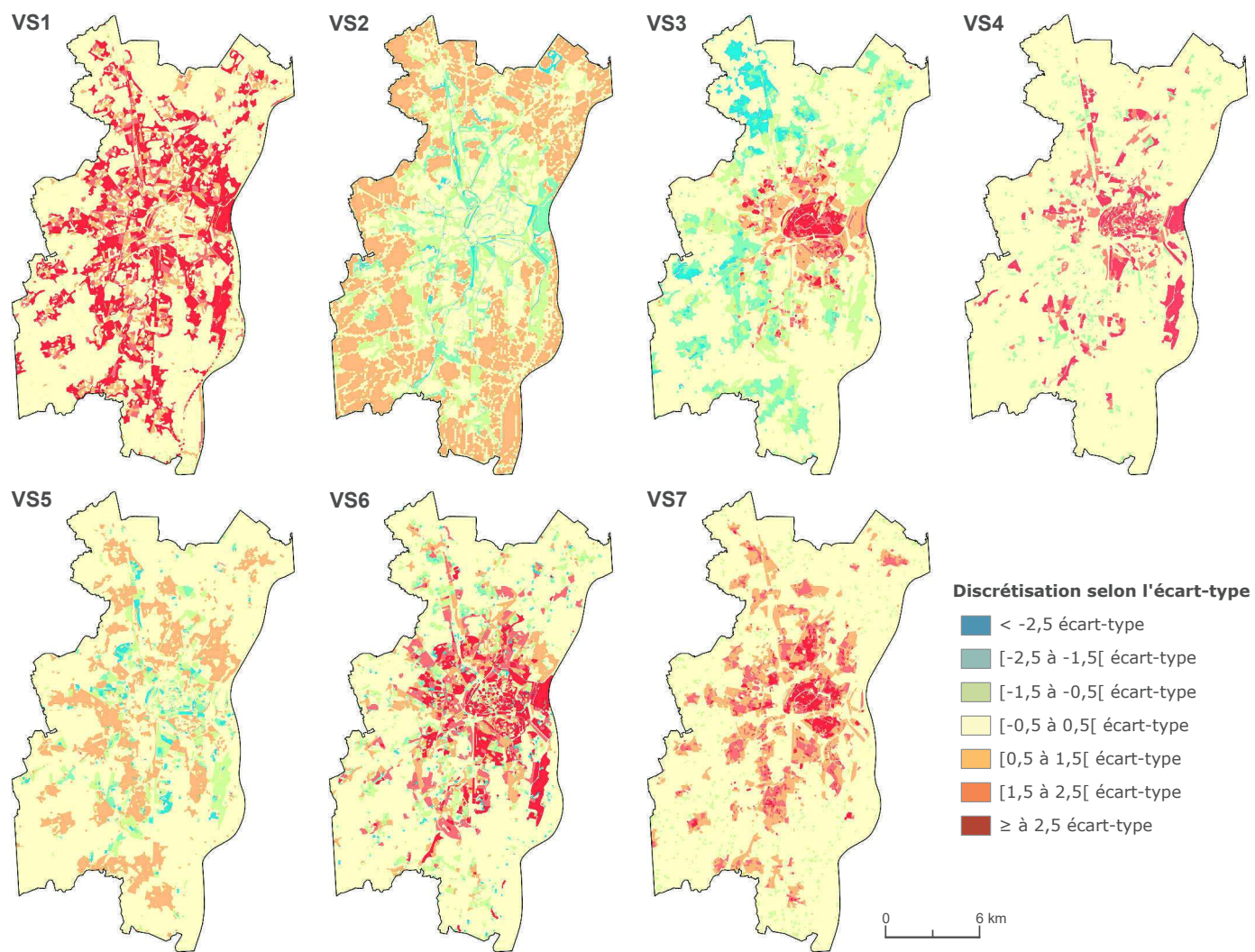


Figure 7.7 – Cartographie des variables synthétiques (VS) 1 à 7 pour l'Eurométropole de Strasbourg

tion y sont associés. Ceux dont la contribution est la plus importante sont la part du plus grand élément dans la maille, la largeur effective de maille, la part d'occupation du sol, la forme et l'aire moyenne des éléments arborés. Elle montre ainsi que, plus ces patches sont grands et dominants, plus leurs formes sont complexes. Cette composante met en évidence les grands patches forestiers présents dans l'agglomération strasbourgeoise et dans une moindre mesure d'autres espaces arborés tels que les ripisylves (figure 7.9 – VS8). Les valeurs faibles se retrouvent principalement dans l'ensemble des zones urbanisées qu'elles soient denses ou discontinues.

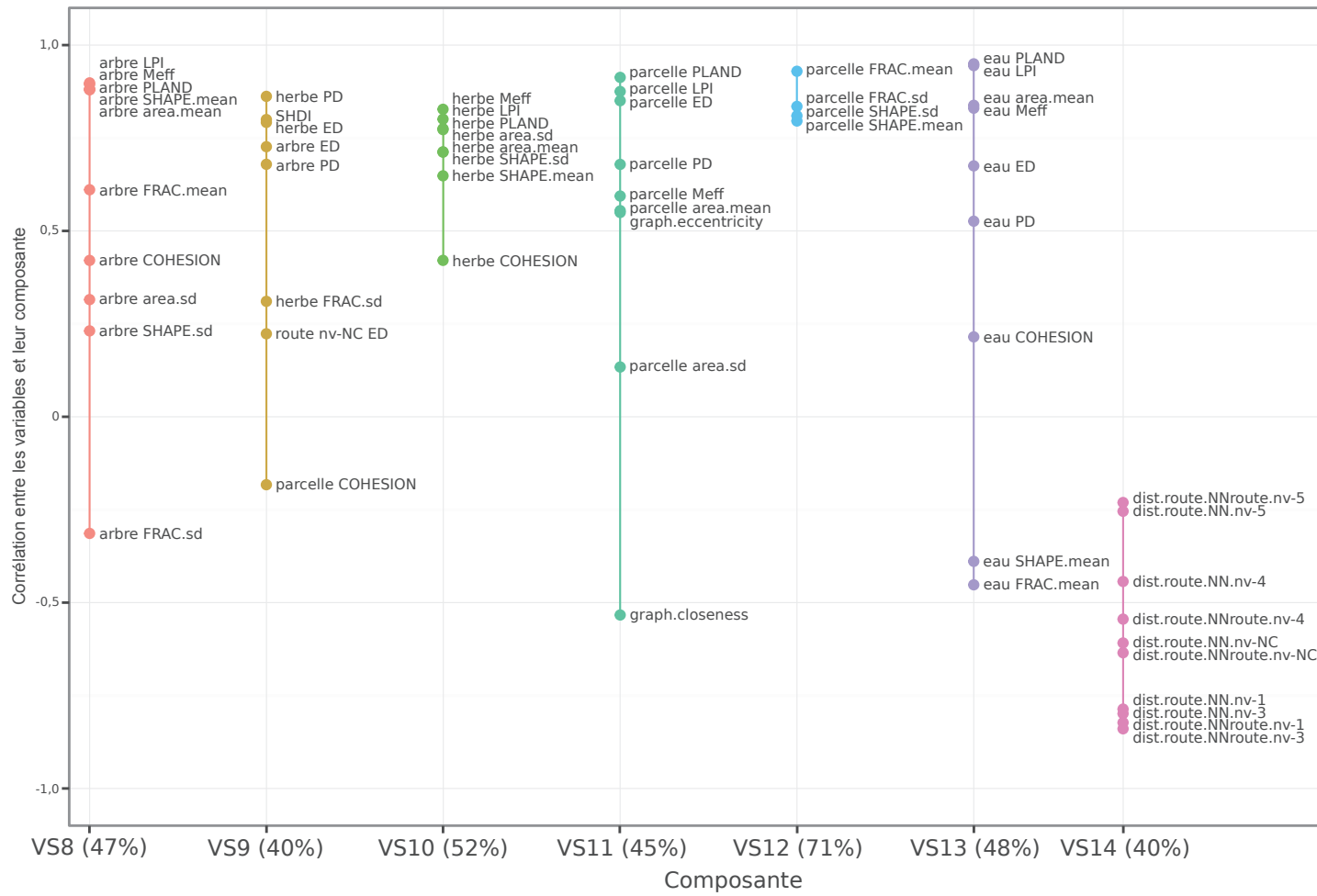
La **variable synthétique 9** rend compte de la densité des éléments de végétation que ce soit en terme de patches, mais aussi de lisière (figure 7.8 – VS9). Les densités des deux strates de végétation sont corrélées positivement à cette composante ainsi que la diversité de la dimension fractale des éléments herbacés et la densité de route de niveau NC. À l'inverse, la cohésion des parcelles agricoles est, elle, légèrement corrélée négativement. De plus, ce sont dans les mailles où la densité de végétation est importante que la diversité en terme d'occupation du sol est importante comme le traduit la corrélation positive élevée de l'indice Shannon à cette variable. Dans les espaces urbains discontinus, cette composante montre des valeurs élevées ainsi que dans certaines zones non urbaines telles que le bord de certains cours d'eau, des clarières ou des équipements sportifs (figure 7.9 – VS9). C'est au contraire dans les espaces agricoles, forestiers et en eau que les valeurs sont les plus faibles.

À l'image de la composante 8 pour la végétation arborée, la **variable synthétique 10** regroupe la majorité des variables permettant de caractériser la végétation herbacée (figure 7.8 – VS10). Cette variable synthétique permet donc d'exprimer la dominance et la dimension de la végétation herbacée au sein des mailles. La carte de celle-ci permet de mettre en évidence la dominance de l'herbe dans les espaces aéroportuaires, à proximité des réseaux ainsi que des grandes surfaces en eau (figure 7.9 – VS10).

La **variable synthétique 11** exprime quant à elle la dominance, la densité et la dimension des parcelles agricoles (figure 7.8 – VS11). Presque l'ensemble des attributs de cette classe y est regroupé excepté ceux permettant de décrire leur forme. Les attributs qui contribuent le plus à cette composante sont la part totale et du plus grand patch des parcelles agricoles ainsi que sa densité de lisière et dans une moindre mesure celle de patch. Cette composante comprend aussi deux mesures de centralité calculées grâce aux méthodes d'analyse de graphes, l'excentricité, corrélée positivement aux attributs dérivés des parcelles agricoles, qui s'oppose à la centralité de proximité, corrélée elle négativement. La cartographie de cette composante révèle l'ensemble des espaces agricoles (figure 7.9 – VS11).

Les derniers indicateurs décrivant les parcelles agricoles sont présents dans la **variable synthétique 12** (figure 7.8 – VS12). Ils concernent la forme et la dimension fractale de ces éléments, tant moyenne que leur dispersion. Les valeurs les plus élevées de cette composante se retrouvent dans des espaces agricoles parfois dominés par des prairies ou où les parcelles sont plutôt de petites dimensions, mais aussi sur les limites de la





**Figure 7.8** – Corrélations entre les variables synthétiques (VS) 8 à 14 et les indicateurs associés (entre parenthèses : pourcentage d’inertie expliquée des variables réelles par la variable synthétique associée)

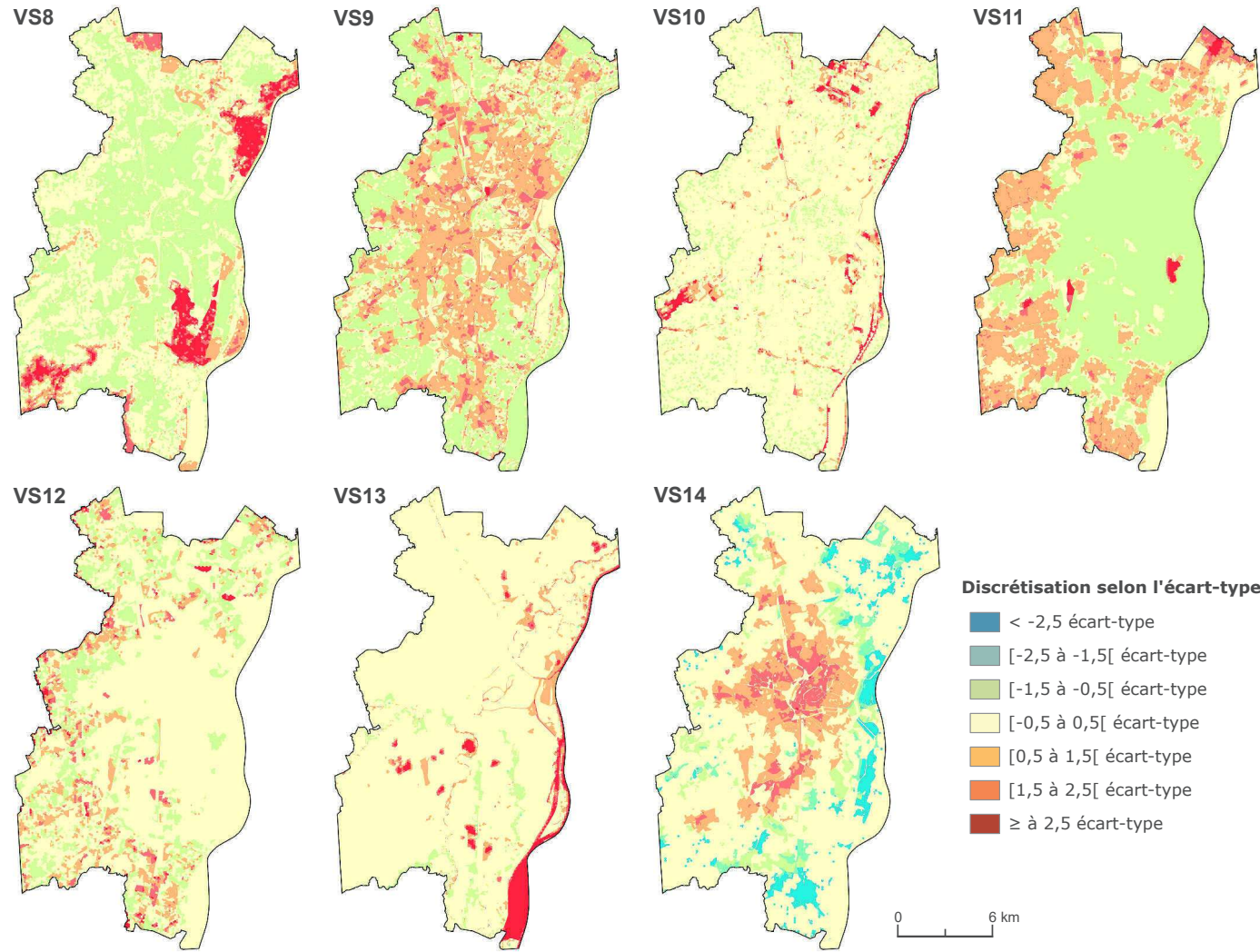


Figure 7.9 – Cartographie des variables synthétiques (VS) 8 à 14 pour l'Eurométropole de Strasbourg

zone d'étude traduisant ici un défaut lié au découpage des parcelles agricole par la zone d'étude (figure 7.9 – VS12).

La **variable synthétique 13** traduit la dominance, la dimension et la densité des surfaces en eau (figure 7.8 – VS13). Les variables les plus corrélées à cet axe sont respectivement la part totale et celle du patch le plus grand, la taille moyenne ainsi que la largeur effective de maille. Les mesures de densité de ces éléments sont aussi corrélées positivement à cette composante. À la différence de ce qui pouvait être observé pour la végétation arborée, les surfaces en eau, plus elles sont grandes et dominantes plus leurs formes sont simples, comme le montre la corrélation négative de l'indice de forme et de la dimension fractale des surfaces en eau à cette variable synthétique. Elle voit apparaître ses valeurs les plus élevées pour les surfaces en eau les plus importantes telles que le Rhin, le plan d'eau de Plobsheim ou les gravières disséminées dans l'ensemble de l'agglomération (figure 7.9 – VS13). Les autres surfaces en eau telles que celles associées à l'Ill ou au port du Rhin ont des valeurs moins importantes.

Le dernier groupe de variables rend compte de l'accessibilité des bâtiments au réseau routier. En effet, les valeurs positives de la **variable synthétique 14** signifient une proximité à ce réseau tandis que les valeurs négatives signifient un éloignement que ce soit à vol d'oiseau ou en distance par la route (figure 7.8 – VS14). L'ensemble des niveaux du réseau routier est présent, excepté ceux du niveau 2. Ceux contribuant le plus à cette composante sont ceux du niveau 1 et du niveau 3. La cartographie de cette variable met en évidence une tache urbaine centrale très accessible, tandis que certaines zones périphériques notamment à l'est de l'agglomération le sont moins (figure 7.9 – VS14). La présence du Rhin constitue ici une barrière naturelle limitant l'accès à des réseaux de transport diversifiés. Le tableau 7.2 reprend de manière synthétique les résultats l'interprétation réalisée dans cette section.

### 7.3.2 Classification non-supervisée des mailles fondée sur les variables synthétiques

À partir des variables synthétiques identifiées et interprétées dans la section précédente, l'objectif est de classer les mailles de Strasbourg. Cela nécessite de choisir le nombre de classes inhérentes à ces données (§ 7.3.2.1) puis d'analyser les classes obtenues (§ 7.3.2.2) pour enfin évaluer la performance de ce découpage (§ 7.3.3.1).

#### 7.3.2.1 Choix du nombre de classes

Lors de la création du découpage, une différenciation a été faite entre les mailles composant l'espace urbain et celles des espaces non urbains (§ 6.5). Les classifications non supervisées réalisées ont été différenciées selon ces deux types d'espaces, soit 4834 mailles urbaines et 16 281 non urbaines. La statistique *gap* a été calculée pour un nombre

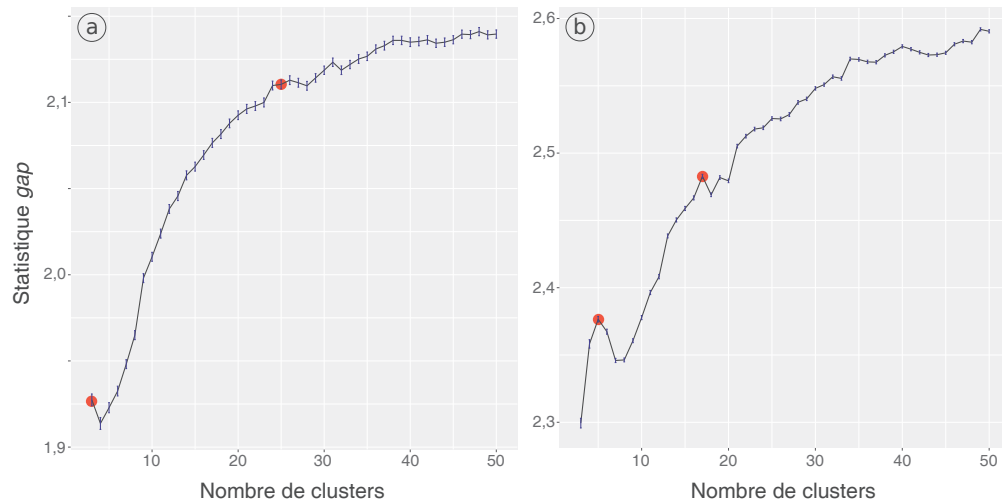
**Tableau 7.2** – Interprétation des variables synthétiques issues du clustering de variables

Variable	Label	Valeurs positives
VS1	Dimension des mailles	Grandes mailles fortement connectées
VS2	Complexité de la forme des mailles	Mailles simples où les réseaux ferré et routier sont peu denses
VS3	Hauteur et complexité du bâti	Mailles avec des bâtiments complexes et hauts avec une part importante de ménages habitants en collectif, peu de propriétaires et proches du réseau routier de niveau 2
VS4	Dimension des bâtiments agrégés	Mailles dont les bâtiments sont de grandes surfaces et souvent composées de bâtiments issus du cadastre de tailles très différentes
VS5	Dimension des bâtiments du cadastre	Mailles avec des bâtiments du cadastre de petites dimensions en surface et en volume dont la forme est simple
VS6	Diversité de forme du bâti	Maille avec une grande diversité de formes et de hauteurs des bâtiments agrégés ou du cadastre
VS7	Dominance et densité du bâti	Maille dont la part d'occupation du sol occupé par le bâti est importante, où les éléments bâtis et la population sont denses
VS8	Taille et dominance de végétation arborée	Mailles occupées par une part importante de végétation arborée peu fragmentée dont les éléments sont de grandes dimensions avec des formes complexes et diverses
VS9	Densité de végétation	Mailles où la densité de végétation arborée et herbacée est importante ainsi que la diversité de l'occupation du sol
VS10	Taille et dominance de végétation herbacée	Mailles occupées par une part importante de végétation herbacée peu fragmentée dont les éléments sont de grandes dimensions avec des formes complexes et diverses
VS11	Dominance et densité des parcelles agricoles	Mailles excentrées où la part de parcelle agricole, leur densité et leur surface sont importantes
VS12	Complexité et diversité de formes des parcelles agricoles	Mailles dont la forme des parcelles agricoles est complexe et diversifiée
VS13	Taille et dominance des surfaces en eau	Mailles occupées par une part importante de surfaces en eau de grandes dimensions
VS14	Accessibilité au réseau routier	Mailles dont les bâtiments sont proches des réseaux routiers de tout niveau (excepté le niveau 2)

de groupes partant de 3, qui correspond au minimum de neurones que permet l'implémentation de la méthode GNG, à 50 afin d'offrir un choix assez large pour la sélection du nombre de classes sans toutefois proposer trop de valeurs du fait de la complexité de cette méthode. Le nombre de simulations de Monte-Carlo retenu pour les différents tests qui sont réalisés ici a été fixé à 500 tel que conseillé par les auteurs du package `cluster` assurant ainsi des résultats précis et stables (Maechler *et al.*, 2016). La figure 7.10 présente l'évolution de cette statistique selon le nombre de classes pour les deux types de mailles.

Pour les espaces urbains, la statistique *gap* sélectionne trois clusters comme étant représentatifs des données (figure 7.10.a). Si ce partitionnement sera analysé dans la section

suivante, il ne permet pas d'obtenir une représentation suffisamment détaillée de la structure urbaine. Le second maximum local a aussi été sélectionné proposant ainsi 25 clusters pour décrire les espaces urbains. Même si ce nombre apparaît *a priori* trop important, ces types pourront être fusionnés lors de l'analyse. Pour les espaces non urbains, cette statistique propose un premier découpage en cinq types (figure 7.10.b). Afin de permettre une analyse à deux niveaux comme pour les espaces urbains le deuxième maximum local est aussi sélectionné et indique un partitionnement en 17 clusters (Tableau 7.3).



**Figure 7.10** – Estimation du nombre de classes pour les espaces urbains (a) et non urbains (b) par la statistique *gap* (les points rouges représentent le nombre de groupes retenus et les barres d'erreur les écarts-types de la statistique *gap* obtenues par une simulation de Monte-Carlo)

**Tableau 7.3** – Nombre de classes sélectionnées par la statistique *gap* par niveau pour les espaces urbains et non urbains

	Espaces urbains	Espaces non urbains
Niveau 1	3	5
Niveau 2	25	17

### 7.3.2.2 Analyse des types urbains et non-urbains en deux niveaux de classification

Dans cette section ainsi que lors de la suivante, nous employons les termes « type » pour les différents clusters extraits des données avec GNG, de « groupe » pour les regroupements de types qui ont été réalisés lors de l'analyse et de « classe » pour les différents modes d'occupation ou d'utilisation du sol extraits des bases de données qui constituent les vérités terrain.

*Interprétation du premier niveau de classification*

Le premier partitionnement des mailles proposé par l'algorithme GNG et la statistique *gap* a permis d'obtenir 3 types de mailles différentes pour les espaces urbains et 5 pour les espaces non urbains. Afin d'analyser ces clusters, leur représentation cartographique (figure 7.11) ainsi que les valeurs tests avec un seuil de significativité de 1 % ont été utilisées. La valeur test permet de comparer les moyennes entre les différents types en prenant compte de la variabilité de leurs valeurs. Elle est normalement utilisée sur des variables passives, qui n'ont pas été utilisées pour le clustering, mais est régulièrement utilisée sur les variables actives pour caractériser les types et constitue ainsi une mesure de similarité entre ces variables et les différents clusters (Lebart *et al.*, 2006). L'ensemble des moyennes des clusters et leur significativité sont résumés dans le tableau 7.4.

Les espaces urbains sont découpés en trois types appelés UA, UB, UC selon un gradient de densité. Le **type UA** correspond aux mailles les plus denses en terme de bâti. Les bâtiments sont hauts, complexes, de formes très diverses et occupent une part très importante des mailles. La végétation y est peu dense et leur dominance est peu importante qu'il s'agisse de végétation arborée ou herbacée. Le **type UB** comprend des mailles petites et simples où la densité de bâti est moyenne et qui se caractérisent par des bâtiments de très faibles hauteurs et de petites dimensions tant pour ceux agrégés que ceux du cadastre. Pour ces mailles la végétation est assez dense malgré que la part qu'elles représentent dans l'occupation du sol reste assez faible. Enfin, le **type UC** est caractérisé par une densité bâtie faible. Les mailles sont généralement grandes et complexes, les bâtiments qui les composent sont assez bas et la végétation y est peu dense. Toutefois, la taille et la dominance de la végétation arborée et herbacée sont plus importantes comparées à celles des deux types urbains précédents.

Pour les mailles non urbaines, la classification a permis de révéler cinq types nommés NA, NB, NC, ND, NE. Le premier, le **type NA**, correspond aux espaces qui sont dominés par les surfaces agricoles. Le **type NB** comprend l'ensemble des mailles où les surfaces arborées sont dominantes et de très grandes dimensions telles que les différentes forêts de l'agglomération ou les ripisylves alors que le **type NC** correspond à celles où la végétation herbacée est dominante. Ces mailles se situent principalement à proximité des grands cours d'eau comprenant aussi une partie des zones aéroportuaires. L'ensemble des surfaces en eau, que se soit les cours d'eau ou les gravières et lacs, sont intégrés au **type ND**. Enfin, le **type NE** est caractérisé par une densité de végétation très importante. Ce type compose l'autre partie des emprises aéroportuaires présentes dans le type NC, mais aussi des équipements sportifs (golfs), les grands réseaux de transport et de nombreuses mailles à proximité des forêts ou des cours d'eau.

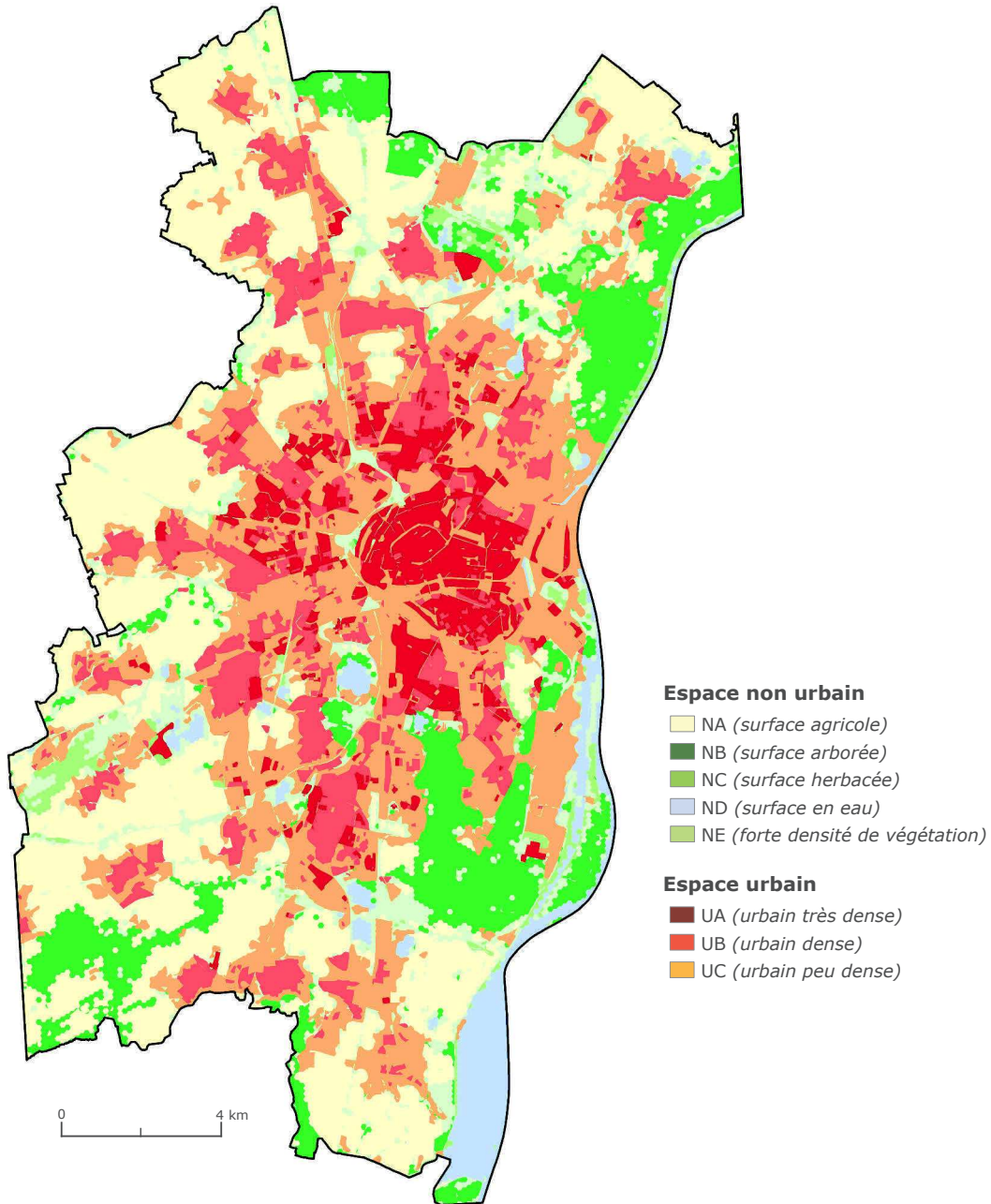


Figure 7.11 – Cartographie du premier niveau de classification réalisé avec GNG

**Tableau 7.4** – Valeur moyenne de chaque type du niveau 1 de classification des mailles de Strasbourg  
 (Les moyennes globales des VS sont de 0, celles mentionnées sont pour l'ensemble des mailles urbaine. En bleue les valeurs significativement négatives et en rouge les positives. Les plus foncées indiquent une valeur test supérieure à 10)

		VS1	VS2	VS3	VS4	VS5	VS6	VS7	VS8	VS9	VS10	VS11	VS12	VS13	VS14
Moyenne des types urbains		0,8	-0,9	0,2	0,4	-0,1	0,4	3,7	-1,4	1,3	-0,3	-1,9	0,0	-0,6	0,5
Urbain très dense	UA	0,2	0,0	5,9	3,0	-1,6	3,4	7,6	-1,6	0,4	-0,6	-2,1	0,0	-0,6	2,2
Urbain dense	UB	0,0	0,1	-2,3	-0,8	0,8	-1,4	4,4	-1,4	1,9	-0,4	-1,8	0,0	-0,7	0,7
Urbain peu dense	UC	2,4	-2,9	-0,6	0,1	-0,2	0,6	0,0	-1,1	1,1	0,0	-1,7	0,0	-0,3	-1,2
Moyenne des types non urbains		-0,2	0,3	-0,1	-0,1	0	-0,1	-1,1	0,4	-0,4	0,1	0,6	0	0,2	-0,1
Surface agricole	NA	-0,2	0,6	0,0	-0,1	0,1	-0,1	-1,1	-0,8	-1,2	-0,6	2,3	-0,1	-0,6	-0,1
Surface arborée	NB	-0,2	0,5	0,0	-0,1	0,1	-0,1	-1,1	4,4	-0,6	-0,6	-0,9	-0,1	-0,4	-0,1
Surface herbacée	NC	-0,1	0,1	-0,1	-0,2	0,1	-0,2	-1,1	-0,6	1,1	8,9	-1,0	-0,1	0,4	-0,1
Surface en eau	ND	-0,3	0,5	0,0	-0,1	0,1	-0,1	-1,2	-0,7	-1,5	-0,3	-1,2	0,0	7,3	-0,1
Forte densité de végétation	NE	-0,4	-0,6	-0,1	0,0	-0,1	-0,3	-1,0	-0,3	1,6	0,6	-0,8	0,2	-0,1	-0,3



### *Interprétation du deuxième niveau de classification*

Les résultats de partitionnement obtenus avec l'algorithme GNG ont dans un premier temps permis de réaliser des regroupements entre les 25 clusters obtenus pour les espaces urbains et les 17 pour ceux non urbains. Ces regroupements, basés sur une analyse visuelle de la cartographie obtenue et des valeurs tests réalisées avec un risque d'erreur de 1 %, ont révélé neuf groupes pour les espaces urbains et huit pour les non urbains. L'ensemble des moyennes des clusters ainsi que leur significativité peut être trouvé dans le tableau 7.5 et sont cartographiés par la figure 7.12.a. L'ensemble des groupes constitué par ce réseau de neurones va être interprété plus en détail.

Les différents types de ce deuxième niveau de classification sont représentés dans l'annexe E.1. Elle présente la localisation des mailles appartenant à chacun des groupes et celle de leur médoïde, la maille la plus proche de la valeur moyenne de la classe, ainsi que les valeurs moyennes de chaque type pour chacune des variables synthétiques et la part moyenne de chaque classe d'occupation du sol.

Le type **type UA-1** est le premier qui compose le groupe centre-ville. Ces îlots, petits, sont ceux où la part de bâti est la plus importante sur l'ensemble de l'agglomération. La végétation est presque totalement absente excepté sur la rue. Les bâtiments agrégés sont relativement hauts et leurs surfaces sont assez grandes, mais assez proches de la moyenne pour ceux du cadastre. Cette caractéristique est assez typique des centres anciens, souvent composés de nombreux bâtiments mitoyens hauts et de petites dimensions et leur agrégation produit ainsi de grands bâtiments. De plus, les hauteurs ainsi que les formes bâties y sont relativement homogènes. Ce type se retrouve principalement dans le centre historique de Strasbourg ou à proximité de la gare et est caractéristique de quelques îlots situés dans d'autres quartiers de Strasbourg. Le **type UA-2** est lui composé de bâtiments plus grands, quel que soit le niveau de représentation. Ces bâtiments sont aussi ceux qui sont les plus hauts de l'ensemble des mailles de l'agglomération. La végétation même si elle reste peu dense y est plus présente que pour le type précédent. Ce type se situe principalement dans la Neustadt (quartiers gare – tribunal et orangerie – conseil des XV) et dans le quartier de la Krutenau. Enfin, le **type UA-3** a une composition proche du type précédent. Les bâtiments y sont toutefois moins haut, d'une hauteur moyenne équivalente à celle du type UA-1, les bâtiments sont beaucoup plus divers en terme de formes ou de hauteurs et la densité de végétation est plus importante. Ce type situé dans les mêmes lieux que précédemment se retrouve aussi dans le quartier du Neudorf et comprend aussi quelques îlots d'habitats collectifs.

Le deuxième groupe identifié à partir de l'analyse de la classification est celui des mailles bâties de type collectif. Ces mailles peuvent correspondre, soit à de l'habitat de type collectif ou des zones urbaines spécialisées dont la morphologie est similaire telles que des emprises scolaires, universitaires ou hospitalières. Cette classe est composée de trois types. Le **type UB-1** est dispersé dans l'ensemble de Strasbourg et de quelques communes limitrophes. Il se caractérise par une densité bâtie assez importante comparée

aux autres types de ce groupe, mais beaucoup plus faible que pour celui du centre-ville. Les bâtiments du cadastre y sont assez grands et leurs hauteurs sont élevées. La diversité des formes de bâtiments y est très faible. Le **type UB-2** se caractérise par une dominance plus faible du bâti que la précédente. Les bâtiments sont aussi plus hauts et plus grands et ont des formes qui sont très variées. Ce type se retrouve de manière importante à Strasbourg dans les quartiers de l'Esplanade et de l'hôpital civil et comprend majoritairement des mailles de type grands ensembles. Enfin, le **type UB-3** comprend des bâtiments dont les hauteurs sont beaucoup plus faibles que pour les deux types précédents tout en restant significativement supérieur à l'ensemble des mailles urbaines et dont la taille est dans la moyenne ressemblant ainsi à ceux du type UB-1. Ce type de collectif est celui le plus présent sur cette zone d'étude et est disséminé sur l'ensemble de Strasbourg. Pour ces trois types, la végétation n'est pas une caractéristique essentielle même si la dominance et la taille de ses éléments sont plus faibles que la moyenne.

La dominance du bâti est, pour le **type UC-1** qui compose la catégorie centre bourg, similaire à celle identifiée pour le groupe UA. Néanmoins, les bâtiments sont d'une taille relativement petite avec des hauteurs légèrement au-dessus de la moyenne. Ce type de maille est très présent dans le centre des communes de Schiltigheim et de Bisheim ainsi qu'à Strasbourg dans les quartiers de Cronembourg et de Neudorf. Il se caractérise aussi par une densité et une dominance de végétation très faibles. Le **type UC-2** est beaucoup moins dense du point de vue bâti que le précédent, mais reste nettement supérieur à ceux du groupe UB. Ce type se retrouve disséminé sur l'ensemble de Strasbourg et des communes voisines. Ici, les bâtiments sont petits, bas et de formes très peu diverses avec une végétation peu présente.

Le groupe composant le bâti de type individuel est le plus diversifié puisque composé de six types différents, mais est aussi le plus nombreux dans les mailles urbaines. Pour chacun des types composant ce groupe les bâtiments sont de petites dimensions, bas qu'elle que soit la base de données utilisée. Le premier, le **type UD-1**, est celui où la dominance et la densité du bâti sont les plus importantes. Les bâtiments sont de petites dimensions et la végétation est très dense. Avec une dominance du bâti équivalente, mais des bâtiments plus petits, plus bas et de formes beaucoup moins diversifiées, le **type UD-2** est celui dont la densité de végétation est la plus importante pour les mailles urbaines. Les mailles de ce type sont petites et simples, mais aussi très accessibles par le réseau routier. Le **type UD-3** reprend les mêmes caractéristiques que le précédent avec des bâtiments toujours hauts, une densité et une dominance du bâti faible, mais est par contre moins accessible. Ce type comme les suivants voit apparaître régulièrement dans ses mailles des parcelles agricoles. Concernant des mailles de très grandes dimensions, le **type UD-4** a une dominance du bâti extrêmement faible. Les bâtiments sont plus hauts, plus grands et de formes plus diverses que les types précédents appartenant au groupe UD. La végétation, moins dense, occupe une place importante tant celle arborée que celle herbacée. Le **type UD-5**, sur des mailles plus petites, a des valeurs similaires que le type UD-4 pour les variables caractérisant la végétation avec une dominance du bâti

**Tableau 7.5** – Valeur moyenne de chaque type du niveau 2 de classification des mailles de Strasbourg  
(Les moyennes globales des VS sont de 0, celles mentionnées sont pour l'ensemble des mailles urbaine. En bleue les valeurs significativement négatives et en rouge les positives. Les plus foncées indiquent une valeur test supérieure à 10)

	VS1	VS2	VS3	VS4	VS5	VS6	VS7	VS8	VS9	VS10	VS11	VS12	VS13	VS14	
Moyenne des types urbains	0,8	-0,9	0,2	0,4	-0,1	0,4	3,7	-1,4	1,3	-0,3	-1,9	0,0	-0,6	0,5	
Bâti de type centre-ville	UA-1	-0,5	0,1	7,1	2,2	-0,4	-1,1	12,6	-1,6	-0,8	-0,7	-2,2	0,0	-0,6	3,2
	UA-2	-0,2	0,1	12,4	3,7	-2	-0,4	9,5	-1,7	-0,2	-0,8	-2,2	0	-0,7	2,8
	UA-3	0	0,1	6,9	2,5	-0,7	7	9	-1,7	0,1	-0,9	-2,2	0	-0,7	2,7
Bâti de type collectif	UB-1	-0,7	0	5,5	-0,2	-1,1	-3,5	5,4	-1,5	1,1	-0,5	-2,1	0	-0,6	2,1
	UB-2	0,8	0	6,2	3,7	-3	12,8	5,1	-1,5	0,9	-0,4	-2,1	0	-0,6	1,4
	UB-3	0,4	0,1	3,4	0,5	-0,7	5	4,8	-1,5	1,1	-0,5	-2,1	0	-0,6	1,8
Bâti de type centre bourg – individuel dense	UC-1	-0,3	0	1,6	-0,4	1,1	1,3	9,4	-1,7	0,4	-0,9	-2	0	-0,7	2,2
	UC-2	-0,5	0,2	-1,8	-1	1,5	-2,1	6,4	-1,6	1,4	-0,7	-2	0	-0,7	1,5
Bâti de type mixte individuel	UD-1	0,8	0	-3,3	-0,8	1,1	0	4,2	-1,4	2,2	-0,3	-1,8	0	-0,7	0,7
	UD-2	-0,2	0,2	-4,3	-1,1	1,5	-2,9	4,1	-1,4	2,7	-0,4	-1,8	0	-0,7	2,1
	UD-3	-0,1	0,1	-6,7	-1,1	1,3	-2,9	3,4	-1,4	2,6	-0,3	-1,3	0	-0,6	-1,7
	UD-4	5,7	-1,1	-3,5	-0,6	0,8	0,1	0,6	-1,1	1,8	0	-0,9	0,2	-0,4	-1,6
	UD-5	0,4	0,1	-2,8	-0,6	0,6	-0,7	3	-1,2	1,5	0	-1	-0,1	-0,5	-5,4
	UD-6	2	-0,3	-2,4	-0,4	0,6	-0,1	1,7	-0,9	1,3	-0,2	-0,6	0,3	-0,5	-14,8
Bâti de type mixte individuel – collectif	UE-1	0,4	-0,1	0,2	-0,6	0,7	1,1	4,5	-1,4	1,7	-0,5	-2,1	0	-0,6	2,3
Zone spécialisée bâtie	UF-1	3,4	-0,2	2,1	49,7	-10	9,3	6	-1,6	0,2	-0,1	-2	-0,1	-0,3	-0,8
	UF-2	3,2	-0,8	2,8	14,1	-3,8	4,5	3,8	-1,4	0,3	-0,3	-2,1	-0,1	-0,5	0,5
	UF-3	-0,3	0,2	1,7	2,5	-13,9	-4,9	5,3	-1,4	0,3	-0,2	-1,9	0	-0,6	-0,2
	UF-4	2,2	-0,5	-0,6	1,1	-1,2	2,1	2,1	-1,3	1,4	-0,1	-2	-0,1	-0,5	2
	UF-5	4,9	-1,1	0,4	1,6	-1,4	5,4	1,3	-1,2	1,4	0,2	-1,8	0	-0,4	-0,3
Emprise réseau	UG-1	0,5	-10,7	-0,2	-0,2	0,2	-0,6	-1	-1,4	0,2	-0,3	-2,1	0	-0,6	0
	UG-2	0,2	-5,1	0,1	-0,2	0,1	-0,1	-0,9	-1,1	1,2	-0,1	-2,2	0	-0,3	0
Espace peu ou pas bâti	UH-1	-0,5	-0,5	0	-0,2	0	0,1	-0,7	-1	1	0,2	-2	0	-0,2	0
	UH-2	-0,1	-0,7	-0,9	-0,8	0,5	-4,2	0,2	-1	1,5	0,1	-1,9	0	-0,4	0,9
Îlot mixte de grande dimension	UI-1	18,2	-2,5	-1,9	1	0,4	2,2	0,6	-1,1	1,5	-0,1	-1,4	0,1	-0,1	-1,6
Moyenne des types non urbains	-0,2	0,3	-0,1	-0,1	0	-0,1	-1,1	0,4	-0,4	0,1	0,6	0	0,2	-0,1	
Surface agricole	NA-1	-0,2	0,6	0	-0,1	0,1	0	-1,2	-0,9	-1,4	-0,6	2,6	1,5	-0,7	0
	NA-2	-0,1	0,6	0	-0,1	0	0	-1,2	-0,9	-1,6	-0,7	2,3	-1,1	-0,7	0
	NA-3	-0,3	0,2	0	-0,1	0	0	-1,1	-0,7	0,7	0	1,4	-0,5	-0,6	0
	NA-4	-0,5	0,3	0	-0,1	0,1	0	-1,1	-0,1	-0,3	-0,3	2,2	11,4	-0,6	0
Surface arborée	NB-1	-0,2	0,5	0	-0,1	0,1	-0,1	-1,1	5,7	-0,9	-0,7	-1,2	-0,1	-0,5	0
	NB-2	-0,2	0,4	0	-0,1	0,1	0	-1,1	2,4	0	-0,4	-1,1	-0,1	-0,5	0
Surface herbacée	NC-1	-0,1	0,2	0	-0,1	0,1	-0,1	-1,1	-0,8	0,7	12,5	-1	-0,1	1,1	0
	NC-2	-0,2	-0,2	0	-0,1	0,1	-0,1	-1,1	-0,6	1,4	4,6	-1,1	-0,1	-0,4	0
	NC-3	-0,5	-0,6	0	-0,1	0	0	-1	-0,1	3,5	0,9	-1,2	-0,1	-0,4	0
Surface en eau	ND-1	-0,4	0,6	0	-0,1	0,1	0	-1,2	-0,9	-2	-0,3	-0,8	0	11,4	0
	ND-2	-0,3	0,5	0	-0,1	0,1	0	-1,2	-0,8	-1,5	-0,2	-1,4	0	6,2	0
Zone lisière	NE-1	-0,3	0,5	0	-0,1	0,1	0	-1,1	2,2	-0,4	-0,7	1,8	-0,8	-0,5	0
	NE-2	-0,2	0,1	0	-0,1	0,1	0	-1,1	0,7	1,2	-0,2	-1	-0,1	2,8	0
Emprise réseau	NF-1	-0,5	-2,5	0,1	0	0	0	-1,1	-1	-0,5	-0,1	-1,7	0	-0,7	0
Bâti isolé	NG-1	-0,2	0,2	-2,1	-0,8	0,6	-4,4	-0,8	-0,5	0,7	0,2	-0,2	-0,1	-0,3	-1,3
	NG-2	0,1	0,5	-2,6	-0,8	0	-3,1	-0,5	0,2	0,6	0,2	0,4	0,2	0	-14,3
Zone spécialisée bâtie	NH-1	-0,4	0,3	0,3	17,7	-29,9	-2,7	3,1	0,4	0,2	-0,4	-1,6	0	-0,2	0,7

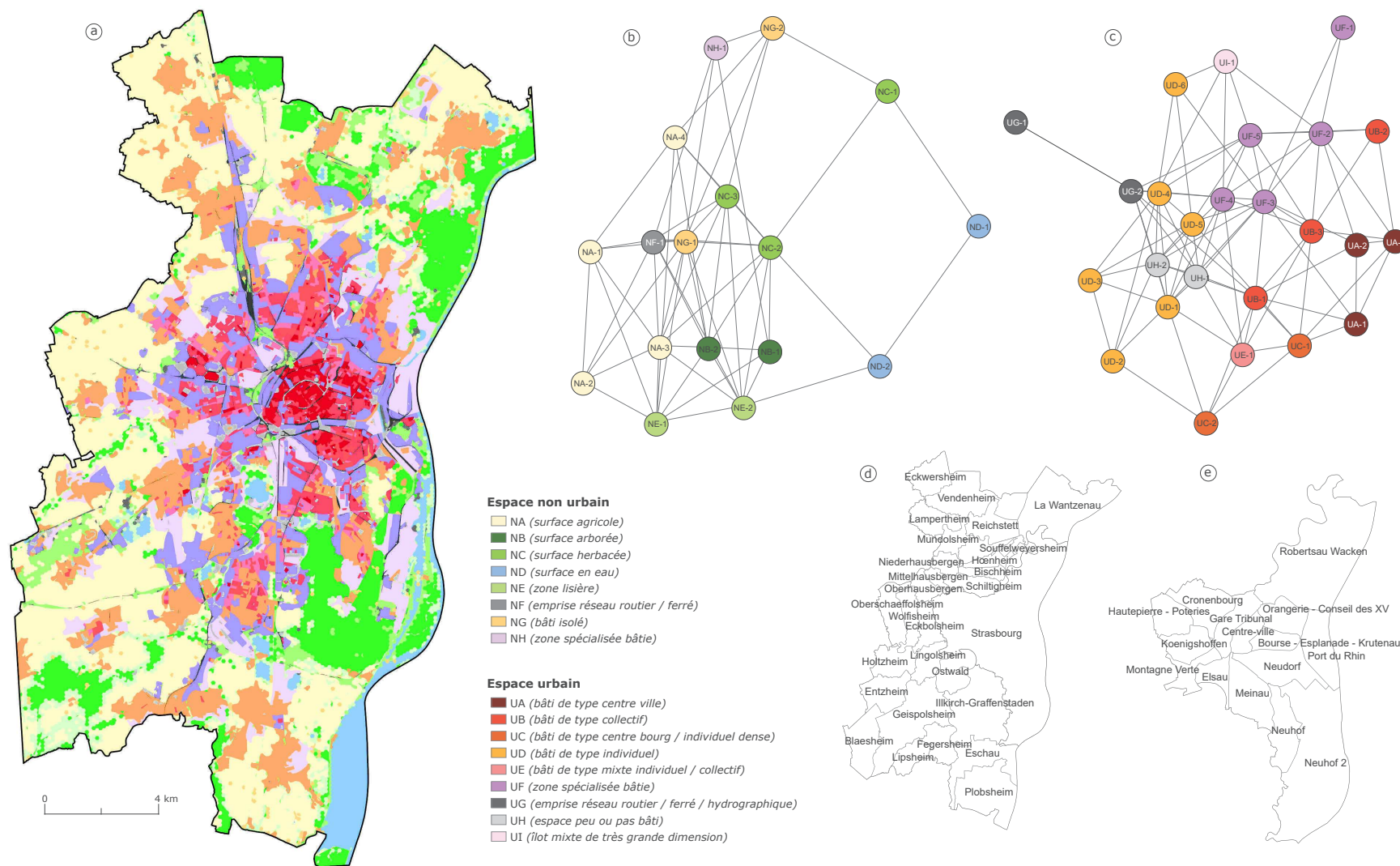


Figure 7.12 – Cartographie du deuxième niveau de classification (a) et réseaux de neurones de GNG pour les espaces non urbains (b) et urbains (c). Localisation des communes de l'agglomération (d) et de quartiers (e) de Strasbourg

beaucoup plus importante. Les bâtiments sont, comparativement aux types précédents, de plus grandes dimensions et de plus grandes hauteurs en étant néanmoins très peu accessibles. Le **type UD-6** se caractérise principalement par sa très faible accessibilité, la grande dimension de ses mailles et la présence importante, pour un type urbain, de parcelles agricoles. La densité de bâti est faible tout comme celle de la végétation avec une dominance de la végétation arborée assez importante.

Le **type UE-1** est le seul composant le groupe bâti de type mixte individuel et collectif. Ce type se caractérise par une densité bâtie qui est assez forte, proche de celle du groupe UB avec des bâtiments de formes diverses et de petites dimensions. La végétation arborée et herbacée est dense et l'accessibilité de ces mailles au réseau est bonne.

Les zones urbaines spécialisées occupent une part importante de l'Eurométropole de Strasbourg et sont elles aussi composées d'une diversité importante de types différents. Avec seulement 13 représentants, le **type UF-1** concerne des mailles imposantes dont les bâtiments sont de très grandes dimensions. La part qu'ils occupent est très importante et la diversité de leurs formes et de leurs tailles l'est tout autant. La végétation est peu présente et peu dense. Ce type concerne principalement de grands bâtiments industriels ou commerciaux ainsi que les grands hôpitaux de Strasbourg et les institutions européennes. Avec des bâtiments de moins grandes dimensions, mais en moyenne plus hauts et où leur densité est moins importante, le **type UF-2** constitue le deuxième de cette classe. Le **type UF-3** est le seul des zones spécialisées bâties à être placé sur des mailles de petite taille. Les bâtiments y prennent une part importante, sont de formes très peu diverses et une très forte différence existe entre les tailles de ceux agrégés et ceux du cadastre. En effet, s'ils restent grands comparés aux autres types urbains quand ils sont agrégés sans être exceptionnel, ceux issus du cadastre sont les plus grands rencontrés dans l'ensemble des types urbains. Les bâtiments du **type UF-4** sont plus petits et plus divers au sein des mailles qui le composent. La dominance et densité du bâti est extrêmement faible et celle de la végétation herbacée est au contraire assez élevée. Enfin, le **type UF-5** concerne de grandes mailles avec des bâtiments d'assez grandes dimensions qui occupent néanmoins une part peu importante de l'occupation du sol. La végétation aussi bien arborée qu'herbacée y est très présente.

Les différents réseaux, qu'ils soient routiers, ferrés, mais aussi hydrographiques, sont présents dans le groupe UG. Le **type UG-1** se caractérise principalement par la forme de ces mailles très complexe avec une densité de réseau très importante. La végétation dans ce type est peu dense à la différence de celui du **type UG-2** où elle est plus présente. La densité de réseau et la complexité de forme des mailles sont moins importantes dans ce dernier. Ce type peut parfois être composé de surface en eau sans que toutes les mailles qui le composent soient forcément concernées. De plus, les végétations arborées et herbacées ont une dominance plus forte que pour l'ensemble des autres mailles urbaines.

Le groupe UH correspond aux espaces urbains non ou très peu bâtis et est, pour les deux types le composant, disséminé sur l'ensemble de l'agglomération. Dans ce groupe

se retrouve l'ensemble des petites places entourées par le réseau, des ronds-points ou des parkings. Le **type UH-1** correspond à des mailles de petite taille et généralement de forme simple où la végétation arborée et herbacée occupe une place importante sans être dense. Le **type UH-2** reprend les mêmes caractéristiques, mais cette fois sur des mailles généralement plus grandes et plus complexes.

Enfin, le dernier type urbain identifié est le **type UI-1**. Il se caractérise principalement par des mailles de très grande dimension toujours situées aux limites de la zone morphologique urbaine. Seulement 70 mailles appartiennent à ce type et leur occupation du sol peut être très différente puisqu'il peut s'agir d'emprises industrielles notamment à l'est ou des mailles mixant habitat individuel et espace agricole.

La classification des mailles non urbaines a permis de révéler 17 types différents qui ont été regroupés en 8 groupes. Le **type NA-1** correspond à des mailles fortement dominées par les parcelles agricoles laissant peu de place aux autres types d'occupation du sol. La forme des parcelles y est assez complexe. Le **type NA-2** lui est très semblable, les parcelles sont très dominantes. Toutefois, leur forme est beaucoup plus simple. Ces deux classes sont très présentes sur l'ensemble du territoire de l'Eurométropole. Moins présent, le **type NA-3** n'est pas exclusivement composé de parcelles. En effet, la part des occupations du sol montre que les végétations arborée et herbacée y sont assez présentes et avec une densité assez importante. Ces mailles sont soit des espaces agricoles composées de haies, principalement dans la partie nord de l'agglomération, soit sont souvent situées à proximité des cours d'eau bordée de végétation. Enfin, le **type NA-4** se caractérise principalement par une complexité et une diversité des formes de parcelles très importante. Ce type, assez rare, se trouve principalement sur la commune de Vendenheim, ou sur les contours de la zone d'étude traduisant ici un effet de bord lié au découpage des différentes bases de données.

Le groupe forêt est composé de deux types. Le **type NB-1**, quasi exclusivement composé d'arbre, obtient des valeurs très importantes pour sa composante 8. La taille des éléments arborés est très grande et ces mailles correspondent à des forêts de grandes dimensions ou de longues ripisylves. Au contraire, le **type NB-2**, même si lui aussi est composé presque uniquement d'arbre, obtient des scores beaucoup moins élevés pour sa composante 8 traduisant des espaces forestiers de plus petites dimensions, des bois ou des mailles situées en marge des plus grandes forêts de l'agglomération.

La dominance et la taille des éléments de végétation herbacée synthétisées par la variable 3 sont significativement fortes pour trois clusters issus de la classification avec GNG. Le **type NC-1** est celui où les surfaces herbacées sont de plus grandes tailles et leur part dans les mailles est la plus importante. Les mailles qui appartiennent à ce type sont généralement situées sur les bords du Rhin ou au niveau des pistes de l'aéroport d'Entzheim. Assez similaires, situées aux mêmes endroits, mais aussi de manière plus éparse, les mailles du **type NC-2** ont une variable synthétique 3 dont les valeurs sont plus faibles et une densité de végétation qui est plus forte. Le **type NC-3** est un peu différent des deux autres. En effet, même si la dominance de la végétation herbacée est

forte comparée aux autres types des espaces non urbains elle est plus faible que pour les types NC-1 et NC-2 et celle arborée y prend une place dominante. La densité de végétation est aussi très importante dans ces mailles.

Deux types décrivent les surfaces en eau. Le **type ND-1** correspond à des surfaces qui sont exclusivement composées d'eau et du fait du découpage de l'espace par carroyage concerne uniquement les mailles situées au centre du plan d'eau de Plobsheim. Les autres surfaces en eau dont les mailles sont situées aux limites des plan ou cours d'eau et intègrent donc d'autres classes d'occupation du sol composent le **type ND-2**.

Le groupe NE, les zones lisières, est composé de deux types qui sont très différents Le **type NE-1** correspond à des zones lisères entre des espaces forestiers dont les éléments sont de grands et des espaces agricoles. Les espaces agricoles où le réseau de haies est important ne sont généralement pas intégrés à ce type, mais dans le groupe NA. Le **type NE-2** correspond à des lisières entre des zones boisées et des surfaces en eau.

Les mailles du **type NF-1** sont des réseaux de transport routier ou ferré. Elles se caractérisent uniquement par une faible moyenne de la variable synthétique 2 qui s'apparente à des mailles de fortes complexités où la densité de réseau de grande importance est élevée.

Le groupe NG correspond à des mailles où des bâtiments sont présents. Dans ces mailles les bâtiments sont majoritairement seuls, bas et petits. La part qu'occupe la végétation est généralement importante et la différence entre les deux types se fait principalement sur l'accessibilité des bâtiments au réseau. En effet, le **type NG-1**, même si les bâtiments sont éloignés du réseau routier, y est plus à proximité que le **type NG-2**.

Enfin, le **type NH-1** est caractérisé par des bâtiments agrégés de très grandes dimensions, mais surtout avec leur représentation non agrégée puisque ce type est celui qui a la moyenne la plus importante pour sa composante 5, espace urbain compris. Ces mailles sont très rares et sont occupées par de grandes surfaces commerciales situées aux marges des espaces urbains.

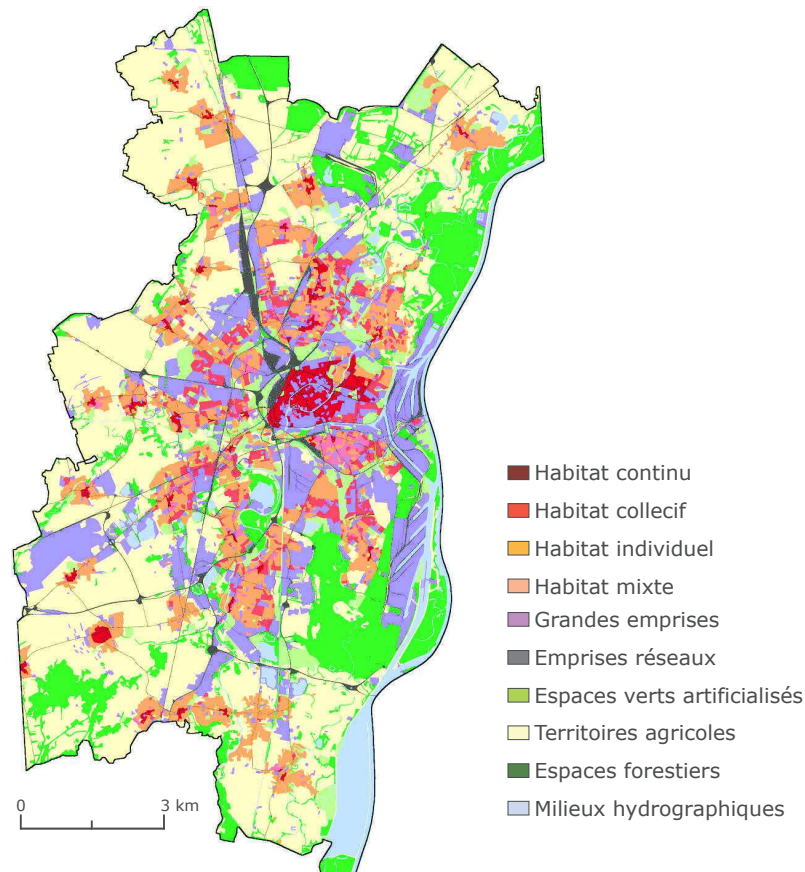
La cartographie obtenue apparaît visuellement cohérente avec la réalité du terrain malgré une surreprésentation des zones spécialisées bâties (figure 7.12.a). De plus, les graphes figurant le réseau de neurones qui représentent les distances et les liens entre chacun apparaissent assez cohérents avec les regroupements qui ont été réalisés entre les différentes classes. En effet, autant pour les espaces non urbains (figure 7.12.b) que pour les espaces urbains (figure 7.12.c) les groupes se dessinent assez bien sur les graphes excepté pour les mailles bâties de type individuel qui apparaissent scindées en deux groupes distincts. La section suivante évalue par l'intermédiaire de deux vérités terrain la classification de niveau 2 des mailles de Strasbourg.

### 7.3.3 Évaluation de la classification de Strasbourg

Cette section est dédiée à l'évaluation de la classification réalisée sur Strasbourg. Deux bases de données sont utilisées comme vérités terrain. La première est la BDOCS de la région Alsace qui permet d'obtenir une typologie proche de celle obtenue par classification des mailles de l'agglomération et la seconde est la base Urban Atlas qui elle présente l'avantage d'être disponible pour les deux sites d'étude (§ 2.2.2).

#### 7.3.3.1 Évaluation à partir de la BDOCS

Afin de valider les regroupements effectués entre les différents types obtenus lors de la classification des mailles de Strasbourg, l'évaluation réalisée à partir de la BDOCS a nécessité de fusionner certaines classes de sa nomenclature afin de pouvoir être comparées aux groupes obtenus avec la classification. Dix classes sont ainsi retenues (figure 7.13).



**Figure 7.13** – Vérité terrain issue de la base de données occupation du sol pour l'évaluation de la classification de Strasbourg (source : BDOCS2012 – CIGAL v2)



Les mesures de précisions pour le producteur et donc la F-mesure présentent deux valeurs différentes liées aux modes de calcul, un premier global et un deuxième corrigé (tableau 7.6). Le premier prend en compte l'ensemble des groupes obtenus lors de l'analyse alors que pour le deuxième tous ne sont pas considérés. En effet lors de l'analyse de la classification des mailles, certains ont pu être extraits des données, mais ne trouvent pas d'équivalents dans la base d'occupation du sol. Ils peuvent aussi être le résultat du maillage de l'espace et sont donc des mailles mixtes qui peuvent fortement biaiser les mesures de précisions. Il s'agit des groupes UC, UH et UI qui correspondent aux centres-bourg, aux espaces libres et aux îlots de grandes dimensions et les groupes NC, NE et NG, les surfaces herbacées, les mailles lisières et les bâtiments isolés. Les précisions globales sont aussi présentées selon ces deux modes de calcul. L'ensemble des matrices de confusion par groupe et par type ayant permis ce calcul peut être trouvé en annexe F.

**Tableau 7.6** – Évaluation des groupes issus de la classification de Strasbourg par la BD OCS (les précisions pour le producteur et les F-mesures sont évaluées de manière globale et corrigée lorsque l'ensemble des groupes n'est pas pris en compte. Le symbole • signifie que la classe n'a pas pu être évaluée, car sans équivalent dans la classification)

Classes	Utilisateur	Producteur		F-mesure	
		Globale	Corrigée	Globale	Corrigée
Habitat continu	0,63	0,57	0,69	0,60	0,66
Habitat collectif	0,49	0,50	0,57	0,49	0,53
Habitat individuel	0,72	0,58	0,81	0,64	0,76
Habitat mixte	0,08	0,14	0,22	0,10	0,12
Grandes emprises	0,54	0,17	0,35	0,26	0,42
Emprises réseaux	0,64	0,62	0,95	0,63	0,76
Espaces verts artificialisés	•	•	•	•	•
Territoires agricoles	0,96	0,80	0,96	0,87	0,96
Espaces forestiers et semi-naturels	0,94	0,74	0,94	0,83	0,94
Milieus hydrographiques	0,99	0,74	0,97	0,85	0,98

Les précisions obtenues pour l'ensemble des classes sont de 66 % globalement et 85 % lorsqu'elle est corrigée. Cette différence montre bien que de nombreuses mailles ont été classées dans les groupes qui ne peuvent être évalués directement à partir de cette base. Si la précision globale après correction est satisfaisante, elle cache de réelles différences entre les espaces urbains et les espaces non urbains. En effet pour les premiers elles sont de 45 % et 60 % tandis que pour les classes non urbaines les précisions obtenues sont de 72 % et 94 %.

Pour les mailles urbaines, de nombreuses confusions existent entre l'habitat continu et l'habitat collectif. C'est généralement quand la végétation devient importante ou les mailles plus lâches, dues à la présence de place ou de square, que l'habitat continu appartient au groupe UB. À l'inverse, le collectif apparaît en UA quand la densité bâtie est très importante au sein de la maille ou que les bâtiments agrégés sont fragmentés par plusieurs du cadastre. Pour l'habitat continu, l'amélioration des performances après correction se fait principalement du fait du groupe UC qui représente un habitat individuel de type centre bourg. En effet dans la base de la région toutes les communes

de l'Eurométropole ont un centre continu, alors que la morphologie de ces tissus peut être très différente du groupe UA qui décrit principalement le cœur de Strasbourg. Ce groupe UC a aussi une grande influence sur la précision pour le producteur de l'habitat de type individuel. En effet, l'utilisation de la version corrigée permet une augmentation de 23 % de cette précision pour cette classe.

L'habitat collectif voit apparaître de nombreuses erreurs du fait de confusions avec le groupe UH qui représente des espaces non construits formés la plupart du temps par de petits îlots, mais aussi avec des mailles classées UE, en habitat de type mixte. Cette classe représentée dans la base régionale est d'ailleurs celle la plus mal classée. Beaucoup de mailles classées UE sont en fait de l'habitat individuel tandis que celles appartenant à l'habitat mixte sont disséminées dans l'ensemble des groupes de type bâti. Le découpage en îlot que nous avons réalisé peut en effet créer des mailles mixtes peu présentes dans la base de la région, car elle a été photo-interprétée. Cette classe est donc dépendante de la surface minimale de représentation.

Les grandes emprises sont aussi une classe urbaine où les erreurs sont très prononcées. Toutes les erreurs ne sont pas concentrées dans les espaces urbains et de nombreuses mailles appartenant à cette classe ont été classées en NC, NA et NF (annexe F.1). Ces différentes mailles sont pour la plupart situées en limites d'espaces urbains, notamment pour celles classées en NC ou NF. Lors de la construction des mailles appartenant aux grandes emprises, des erreurs ont pu apparaître aux marges lorsque celles-ci se trouvent à une distance de plus de 50 m. Les grandes emprises classées en agricole peuvent être dans la même configuration, mais c'est principalement les espaces aéroportuaires qui ont été classés en NA. Les pistes éloignées des bâtiments appartiennent à ces espaces non urbains et ne comprennent aucun bâtiment. Selon l'aéroport, les erreurs ont des causes différentes. Pour celui d'Entzheim, les mailles sont simples et vides tandis que pour celui du Neuhof les pistes sont, dans le RPG, déclarées comme des parcelles agricoles.

Au sein des espaces urbains, les confusions sont aussi très importantes avec cette classe, mais sont beaucoup plus diversifiées et soulignent bien la difficulté de l'évaluation de cette classification. Tout d'abord, il en existe de nombreuses avec le groupe UD. C'est principalement les types UD-5 et UD-6 qui sont concernés (annexe F.2). Ces groupes étaient notamment caractérisés par un éloignement au réseau important. Certaines emprises industrielles, principalement au nord de la zone d'étude sur la commune de la Wantzenau, ont ainsi été classées dans l'un de ces deux types. Mais l'autre confusion principale se fait entre les grandes emprises et le groupe UB. Ce sont principalement les espaces urbains spécialisés, tels que définis dans base régionale, qui sont source de ces confusions puisqu'ils en représentent les deux tiers. Les emprises scolaires notamment sont presque toutes assignées à UB montrant la difficulté de les différencier d'un point de vue morphologique. Appartenant aux grandes emprises, les casernes militaires souffrent aussi du même défaut, les immeubles sont grands, régulier, dans des mailles peu denses et très semblables aux immeubles d'habitat collectifs. L'hôpital civil situé à proximité du centre de Strasbourg est assez caractéristique des difficultés d'obtenir la classe grande emprise par une étude morphologique (figure 7.14). Plusieurs groupes apparaissent dans

son enceinte, le groupe UA à l'est, UB au centre et UF à l'ouest qui est l'emplacement du nouvel hôpital civil. De plus, des mailles non bâties (UH) ou de réseau (UG) sont aussi présentes. D'un point de vue morphologique, la classification réalisée dans cette thèse paraît valable malgré des erreurs liées aux découpages qui produisent de grandes mailles où apparaissent des confusions entre habitat continu et collectif mentionnées précédemment. D'un point de vue thématique, la classification apparaît par contre fortement erronée.



**Figure 7.14** – Exemple de la difficulté d'évaluation de la morphologie urbaine par la base de données d'occupation du sol de la région Alsace : l'hôpital civil de Strasbourg représenté par l'image Pléiades (a), l'occupation du sol (b) et le résultat de la classification des mailles (c). Les photographies représentent des mailles classées en UA (1), UB (2) et UF (3)

Les espaces verts artificialisés de la BD OCS n'ont pas pu être évalués. En effet, aucun type identifié lors de la classification des mailles n'a permis de les extraire. Ces espaces se retrouvent ainsi majoritairement distribués dans les types NC-3 et NE-2 dans les mailles non urbaines et dans les groupes UF et UH des espaces urbains (annexes F.1 et F.2). Pour ces derniers, la présence de bâtiments est déterminante dans l'attribution à l'un ou l'autre groupe. Les espaces verts artificialisés, lorsqu'un bâtiment est présent, sont préférentiellement assignés à UF comme les équipements sportifs tels que les stades, les piscines et certains parcs.

L'ensemble des classes non urbaines apparaît assez bien extrait. En effet, si dans la vérité terrain les classes urbaines sont peu reliées à la morphologie, celles non urbaines sont plus directement assimilable à l'occupation du sol et peuvent donc être plus fidèlement décrites par les bases de données que nous avons utilisées ou constituées. Pour l'ensemble des classes, une nette amélioration apparaît lors que nous appliquons la correction des précisions. Qu'il s'agisse des territoires agricoles, des espaces forestiers ou des milieux hydrographiques, de nombreuses mailles appartenant à ces classes ont été attribuées aux groupes NC, les surfaces herbacées, ou NE, aux zones lisières. Plus spécifiquement pour les espaces agricoles, un certain nombre de mailles appartiennent à NG ce qui correspond à des bâtiments agricoles isolés.

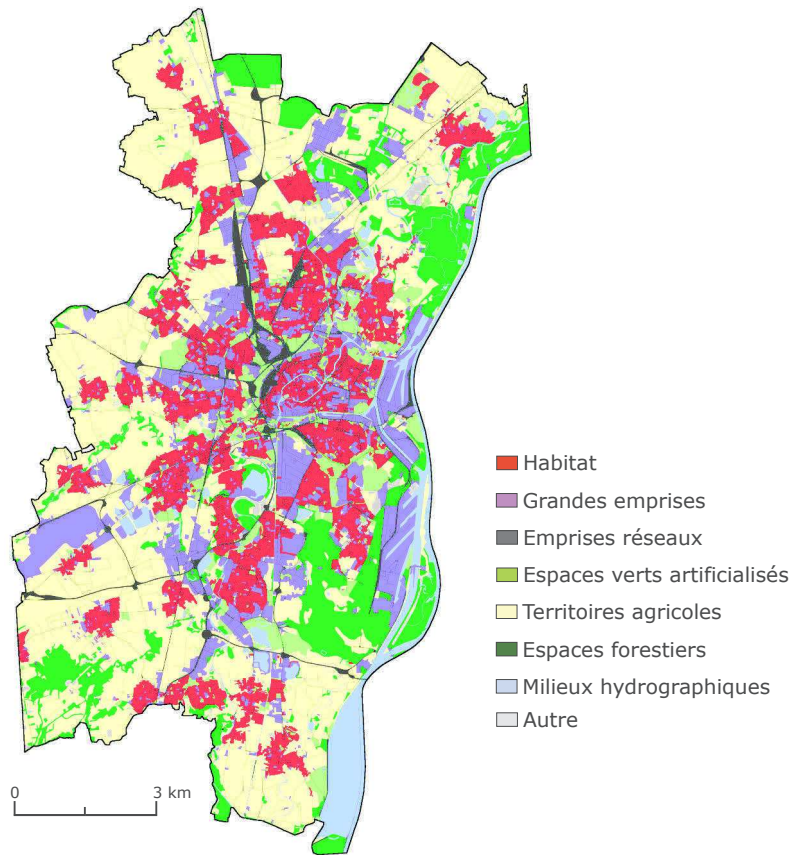
### 7.3.3.2 Évaluation à partir de la base Urban Atlas

Une deuxième évaluation est réalisée à partir de la base Urban Atlas. Même si cette base n'est pas aussi précise que celle de la région Alsace, elle est disponible pour nos deux espaces d'études pour l'année 2012 et permet donc de comparer les performances de cette classification avec celle du site de Rennes.

Moins de classes sont présentes que pour l'évaluation précédente et seulement huit ont ainsi été construites (figure 7.15). Ce sont les classes d'habitat qui sont moins détaillées que dans la base de données régionale. Dans la base Urban Atlas les espaces dont la vocation est majoritairement résidentielle sont découpés selon un gradient de densité. Les groupes obtenus avec le clustering ne permettent pas de faire ressortir ces notions de densité bâtie et les différentes classes d'Urban Atlas ont donc été fusionnées pour former une classe « habitat ». Les autres classes évaluées précédemment sont toutes présentes et équivalentes. Seule une classe « autre » a dû être définie et comprend les classes : association de végétation herbacée, structures isolées et espaces sans usage courant.

Les précisions globales sont assez similaires entre les deux bases de validation. Elle est légèrement plus optimiste avec Urban Atlas puisqu'elles sont de 70 % pour la version globale et 89 % pour la corrigée. C'est principalement pour les mailles urbaines que les précisions sont meilleures (67 % et 78 %) alors qu'elles restent équivalentes dans les espaces non urbains (71 % et 92 %). La fusion des différentes classes d'habitat en une seule peut expliquer cette nette amélioration des performances. En effet pour cette classe les F-mesures atteignent 0,86 et 0,88 (tableau 7.7). Les performances pour les autres

classes sont équivalentes et les grandes emprises apparaissent toujours aussi difficiles à comparer avec cette base de données.



**Figure 7.15** – Vérité terrain issue de la base Urban Atlas pour l'évaluation de la classification de Strasbourg (source : Urban Atlas 2012, Land Copernicus)

**Tableau 7.7** – Évaluation des groupes issus de la classification de Strasbourg avec la base Urban Atlas (les précisions pour le producteur et les F-mesures sont évaluées de manière globale et corrigée lorsque l'ensemble des groupes n'est pas pris en compte. Le symbole • signifie que la classe n'a pas pu être évaluée, car sans équivalent dans la classification)

Classes	Utilisateur	Producteur		F-mesure	
		Globale	Corrigée	Globale	Corrigée
Habitat	0,80	0,91	0,96	0,86	0,88
Grandes emprises	0,62	0,19	0,30	0,29	0,41
Emprises réseaux	0,68	0,53	0,87	0,59	0,76
Espaces verts artificialisés	•	•	•	•	•
Territoires agricoles	0,95	0,79	0,95	0,86	0,95
Espaces forestiers et semi-naturels	0,88	0,79	0,96	0,83	0,92
Milieus hydrographiques	0,99	0,72	0,96	0,83	0,97
Autre	•	•	•	•	•

## 7.4 Classification de la structure urbaine de Rennes

L'objectif de cette section est de valider la méthodologie mise en place sur l'agglomération de Strasbourg sur un autre site. Cette section se décompose en deux étapes. La première classe les mailles de Rennes à partir des classifications réalisées sur Strasbourg (§ 7.4.1) et la deuxième évalue ces résultats (§ 7.4.2).

### 7.4.1 Application des paramètres des classifications de Strasbourg sur Rennes

Les paramètres définis sur le site de Strasbourg ont ainsi été appliqués dans leur intégralité sur celui de Rennes. Dans un premier temps, ceux de la classification de variables ont été appliqués, à savoir le découpage des groupes de variables, mais aussi les paramètres des ACP réalisées pour chacun d'eux. Dans un deuxième temps, ce sont les paramètres de la classification qui ont été appliqués et donc les individus de Rennes assignés au neurone le plus proche sans que le réseau ne soit mis à jour. Comme pour Strasbourg, les mailles ont été classées selon qu'elles appartiennent ou non aux espaces urbains constituant ainsi deux jeux de données composées de 3141 de mailles urbaines et de 14096 non urbaines.

Les deux niveaux de classification ont été appliqués sur le site de Rennes. L'ensemble des 8 types qui compose ce premier niveau a été obtenu pour ce site (figure 7.16). Comparés à Strasbourg, les espaces bâtis peu denses prennent une place beaucoup plus importante au détriment du type UB qui est moins représenté pour ce site d'étude. Dans les espaces non urbains, le type NE, où est présente une forte densité de végétation, comprend de nombreuses mailles notamment dans les milieux agricoles du fait d'un réseau de haies beaucoup plus dense et qui, de temps en temps, peut être classé dans NB pour les réseaux de haies les plus importants.

Les 42 types du deuxième niveau de classification sur le site de Strasbourg ne sont pas tous représentés sur celui de Rennes (figure 7.17). Le type ND-1 n'est en effet pas présent ici. Ce type représente les mailles fortement dominées par les surfaces en eau. Le groupe UA, le bâti de type centre-ville, apparaît beaucoup plus concentré sur ce site et une dominance importante du groupe UB localisée sur la commune de Rennes se dessine. Le groupe UC est aussi beaucoup moins présent sur ce site puisque seulement 85 mailles y ont été attribuées tandis qu'elles étaient 563 à Strasbourg. Comme précédemment, cette classification apparaît cohérente avec la réalité du terrain avec une présence importante du groupe UF. Le groupe UD est ici apparaît très présent dans les espaces périphériques de Rennes.

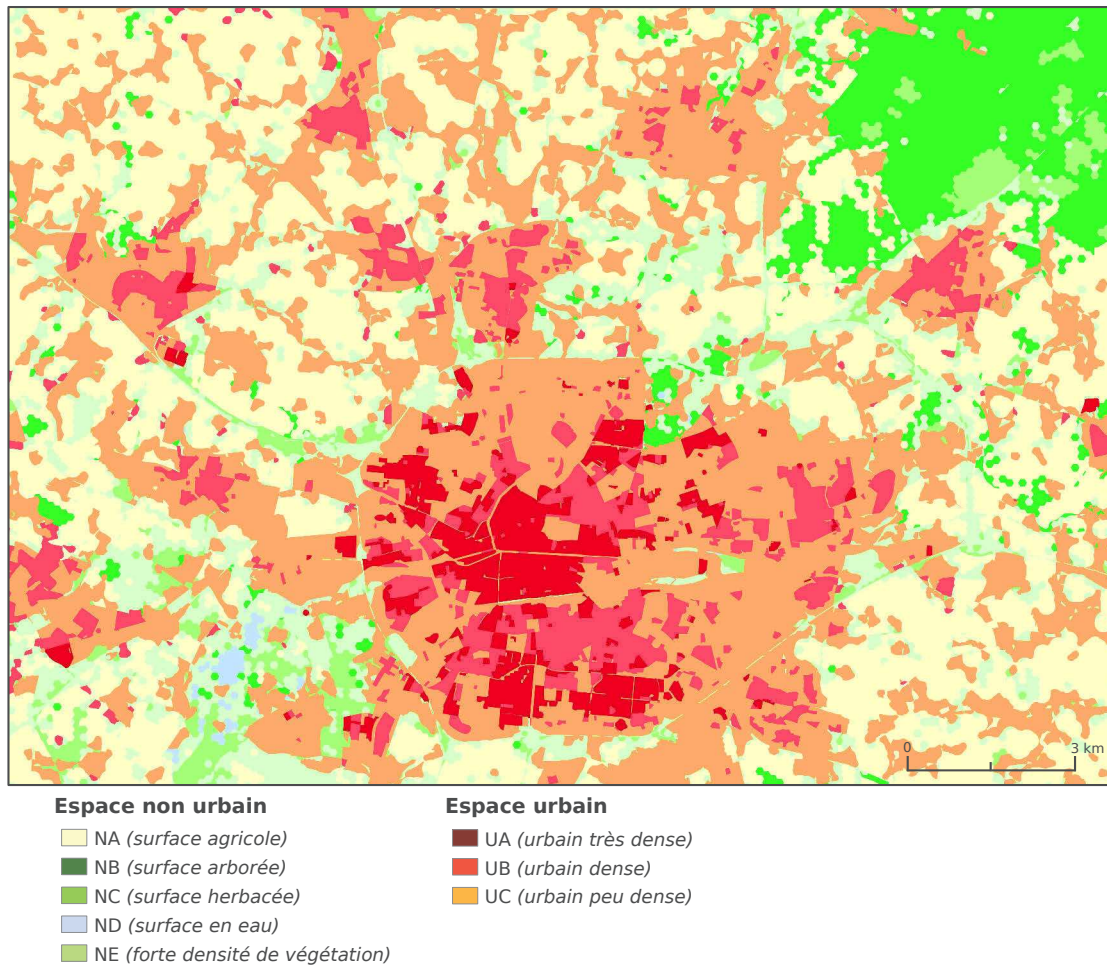
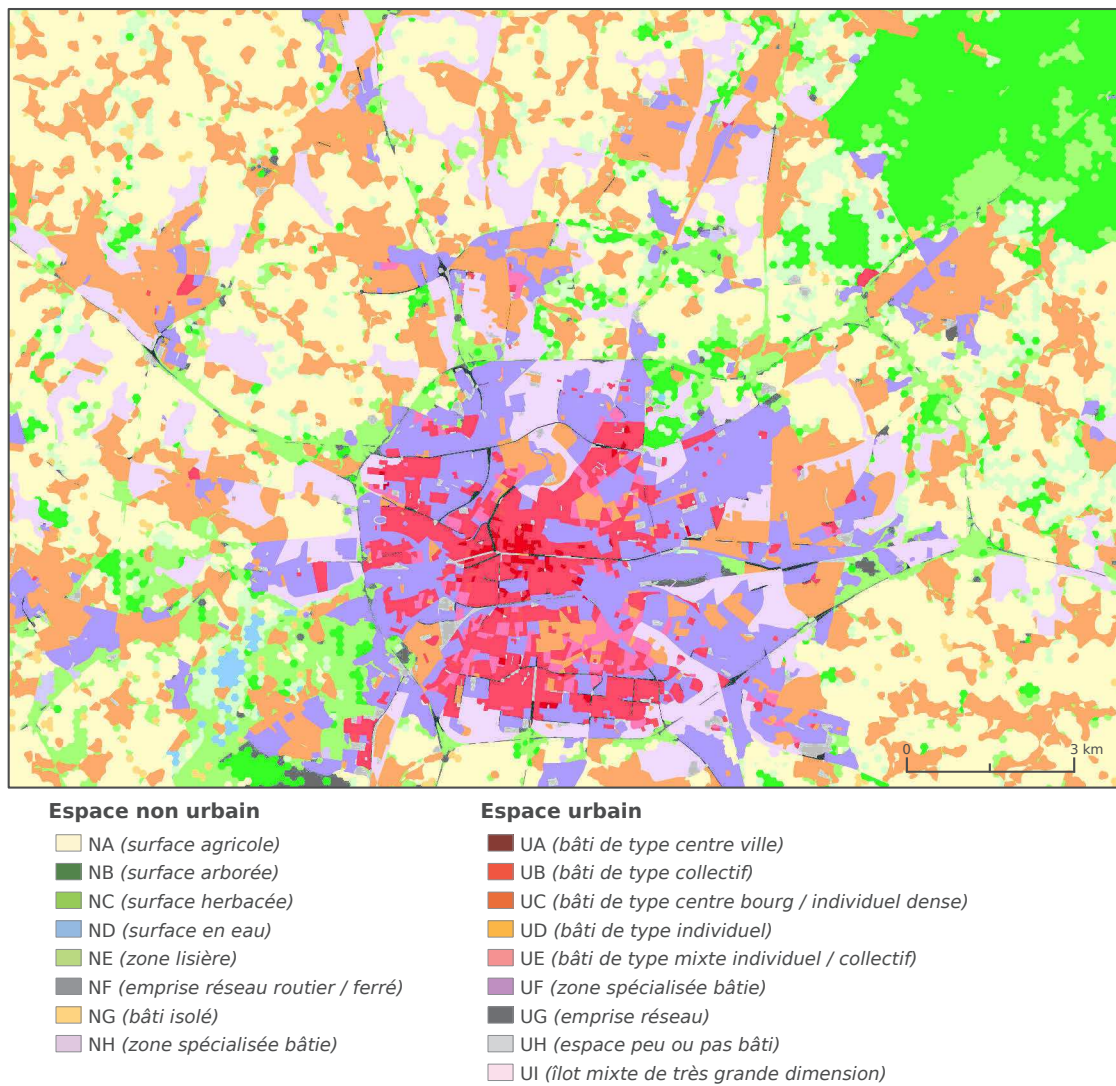


Figure 7.16 – Premier niveau de classification en 8 types du site de Rennes à partir des paramètres de Strasbourg

#### 7.4.2 Évaluation de la classification des mailles de Rennes

Une première évaluation de la classification de la structure urbaine de Rennes est réalisée à partir de la base de données Urban Atlas (figure 7.18). Les performances globales de cette classification sont, dans leurs versions corrigées, légèrement inférieures à celles obtenues à Strasbourg. Elles sont en effet de 62 % pour la précision globale et 85 % pour la corrigée. Toutefois, ces précisions cachent une véritable différence entre les espaces urbains et non urbains. Pour les mailles urbaines, elles sont de 53 % et 63 %, soit très en deçà des performances obtenues pour Strasbourg et proches pour celles non urbaines puisqu'elles atteignent 64 % et 91 %. L'amélioration très importante entre les deux modes de calcul est principalement liée aux mailles non urbaines qui sont très majoritaires et dont l'augmentation est très importante. En effet, de nombreuses mailles appartenant aux territoires agricoles dans la base Urban Atlas, mais aussi de forêts, ont



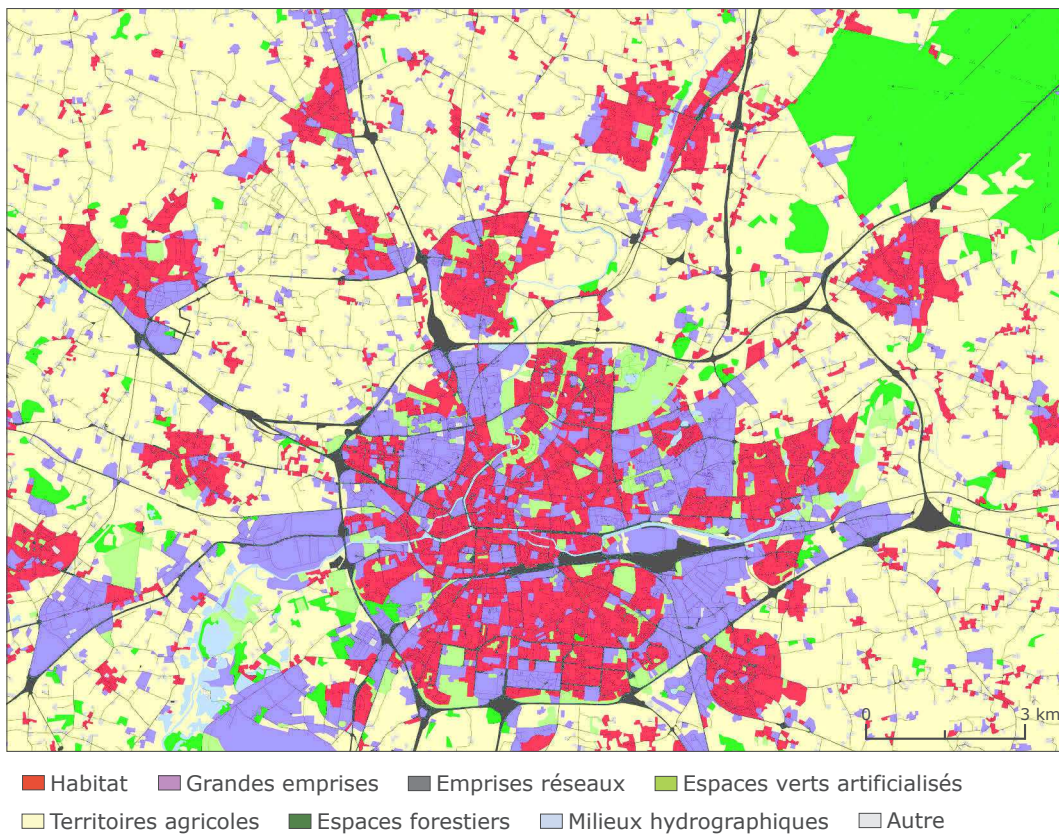
**Figure 7.17** – Deuxième niveau de classification en 42 types du site de Rennes à partir des paramètres de Strasbourg

été attribuées aux groupes NC et NE qui ne sont plus pris en compte lors du calcul de la variante corrigée de la précision globale (annexe F.5).

L'analyse par classe permet de révéler que pour les espaces urbains la classe habitat est celle dont les performances sont beaucoup moins bonnes que pour les classes urbaines de Strasbourg. Les F-mesures pour cette classe sont de 0,69 et 0,73 (tableau 7.8) alors qu'elles étaient de 0,86 et 0,88 sur l'autre site (tableau 7.7). Ce sont principalement les types appartenant à UD qui obtiennent les précisions les plus basses (annexe F.6). Les confusions de ce groupe sont principalement liées aux types UD-4, UD-5 et UD-6 avec les territoires agricoles. Comme il apparaissait sur la cartographie de la classification



où la représentation du groupe UD semblait très importante, le découpage en maille et la configuration des habitations dans les espaces périurbains peuvent en être la cause. En effet sur ce site, de nombreuses maisons « isolées » sont présentes, mais assez peu distantes les unes des autres comme c'était le cas sur le site de Strasbourg, et qui lors de la création des espaces urbains y ont été intégrés. Il serait donc nécessaire d'adapter les distances utilisées pour le site de Strasbourg à celui de Rennes ou de redéfinir ces trois types pour cette zone d'étude. Les précisions obtenues pour les grandes emprises ainsi que les réseaux de transport sont similaires entre les deux sites. Tout comme pour Strasbourg, de nombreuses mailles classées en grandes emprises par la base Urban Atlas appartiennent aux groupes UB et UD à Rennes. De même, les espaces urbains spécialisés tels que les établissements scolaires et universitaires ont tendance à appartenir à UB tandis que les confusions pour les emprises industrielles et commerciales sont préférentiellement faites avec UD.



**Figure 7.18** – Vérité terrain issue de la base Urban Atlas pour l'évaluation de la classification de Rennes (source : Urban Atlas 2012, Land Copernicus)

Dans les espaces non urbains de ce site d'étude, les précisions obtenues par classes sont aussi moins performantes que pour le précédent. Équivalentes pour les territoires agricoles, elles sont très en dessous pour les espaces forestiers. La précision pour l'utili-

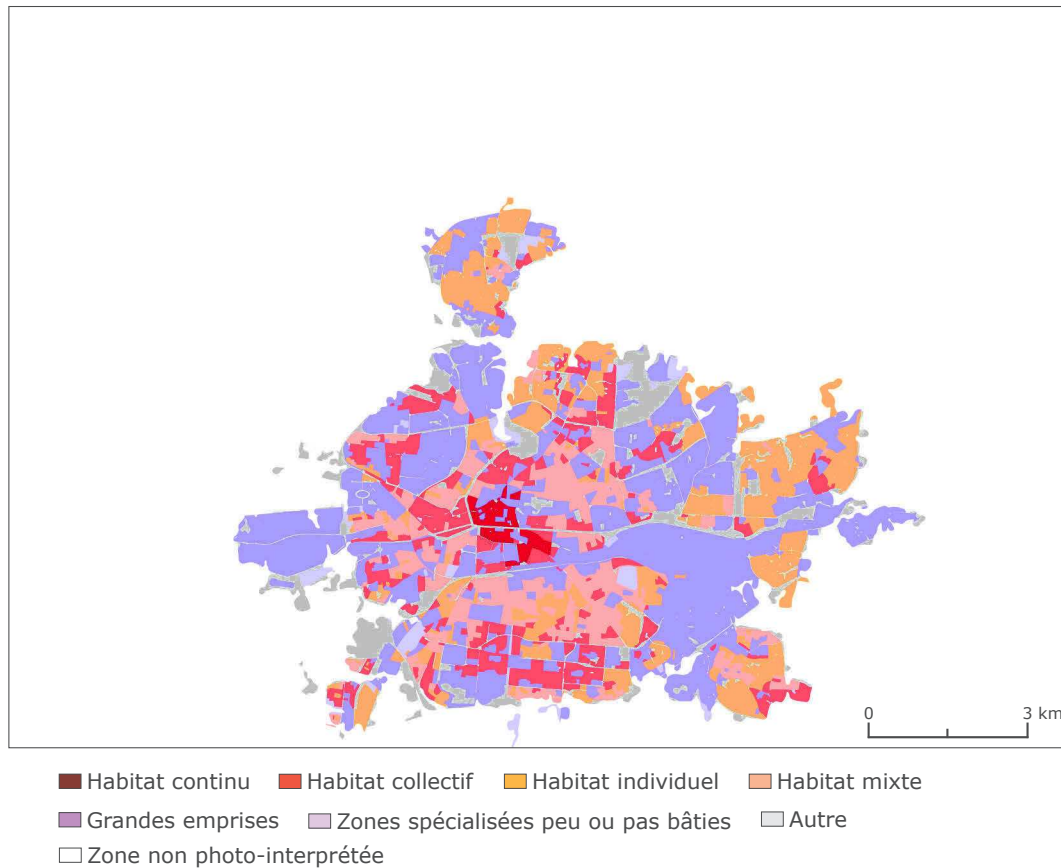
sateur est spécifiquement concernée. Sur ce site, de nombreuses mailles appartenant aux territoires agricoles ont été attribuées au groupe NB ce qui est le cas lorsque le réseau de haies est très dense. Les milieux hydrographiques aussi sont plus difficiles à extraire, mais les précisions sont très différentes lorsqu'elles sont globales ou corrigées. Contrairement à Strasbourg où de grandes surfaces en eau existent, celles de Rennes sont beaucoup plus réduites. Ainsi les mailles sont souvent mixtes et sont donc classées dans les groupes NC et NE.

**Tableau 7.8** – Évaluation des groupes issus de la classification de Rennes avec la base Urban Atlas (les précisions pour le producteur et les F-mesures sont évaluées de manière globale et corrigée lorsque l'ensemble des groupes n'est pas pris en compte. Le symbole • signifie que la classe n'a pas pu être évaluée, car sans équivalent dans la classification)

Classes	Utilisateur	Producteur		F-mesure	
		Globale	Corrigée	Globale	Corrigée
Habitat	0,63	0,77	0,87	0,69	0,73
Grandes emprises	0,55	0,21	0,33	0,30	0,41
Emprises réseaux	0,70	0,43	0,89	0,54	0,79
Espaces verts artificialisés	•	•	•	•	•
Territoires agricoles	0,98	0,67	0,89	0,80	0,93
Espaces forestiers et semi-naturels	0,68	0,77	0,97	0,72	0,80
Milieux hydrographiques	0,96	0,29	0,75	0,45	0,84
Autre	•	•	•	•	•

Une deuxième évaluation est réalisée pour ce site à partir de la base îlot photo-interprétée, utilisée et présentée dans la section 5.4.2. Cette base nous permet en effet d'évaluer en partie la classification de Rennes avec plus de détails que Urban Atlas. Toutefois, elle ne couvre pas la totalité de la zone d'étude et seule la tache urbaine centrale de l'agglomération rennaise a été photo-interprétée (figure 7.19). Elle ne permet donc pas l'évaluation des groupes appartenant aux espaces non urbains et ne livre qu'une évaluation incomplète de certains groupes urbains spécialement l'habitat individuel qui est assez peu présent dans cette zone.

Les précisions sont de manière générale moins bonnes que celles obtenues à Strasbourg. De nombreuses confusions existent toujours entre l'habitat continu et l'habitat collectif avec un nombre peu important d'îlots classés en habitat collectif qui peut expliquer les précisions moins bonnes pour cette classe. Des erreurs apparaissent aussi comme précédemment entre le groupe UB et les grandes emprises. Mais c'est principalement l'habitat individuel qui voit ses performances fortement diminuer par rapport à notre premier site d'étude. Évaluées avec cette base, les confusions du groupe UD se font principalement avec ceux d'habitat mixte dont le nombre est beaucoup plus important. La photo-interprétation a en effet été réalisée à l'échelle de l'îlot et non comme à Strasbourg par digitalisation ce qui crée une représentation importante de cette classe. Elle apparaît donc mieux classée même si les performances obtenues restent faibles.



**Figure 7.19** – Vérité terrain issue de la base îlot pour l'évaluation de la classification de Rennes (source : photo-interprétation manuelle, L. Jacquemin – Nature Par Cœur)

**Tableau 7.9** – Évaluation des groupes issus de la classification de Rennes avec la base îlots photo-interprétée (les précisions pour le producteur et les F-mesures sont évaluées de manière globale et corrigée lorsque l'ensemble des groupes n'est pas pris en compte. Le symbole • signifie que la classe n'a pas pu être évaluée, car sans équivalent dans la classification)

Classes	Utilisateur	Producteur		F-mesure	
		Globale	Corrigée	Globale	Corrigée
Habitat continu	0,46	0,46	0,47	0,46	0,47
Habitat collectif	0,35	0,52	0,64	0,42	0,45
Habitat individuel	0,61	0,50	0,59	0,55	0,60
Habitat mixte	0,48	0,25	0,29	0,33	0,36
Grandes emprises	0,58	0,31	0,49	0,41	0,53
Zones spécialisées peu ou pas bâties	•	•	•	•	•
Autres	•	•	•	•	•

## Conclusion

Ce chapitre a permis de cartographier la structure urbaine de l'agglomération de Strasbourg à partir d'une méthodologie d'extraction de connaissances puis de valider les résultats sur le site de Rennes. Cette méthodologie a été mise en place sur la base du découpage et des indicateurs proposés dans le chapitre 6 qui permettait de caractériser quantitativement les trames grises et vertes à partir de données d'occupation du sol et de statistiques sociodémographiques.

Après avoir présentées et sélectionnées les méthodes nécessaires à l'élaboration de notre démarche, la première étape a consisté à l'analyse de la structure urbaine de Strasbourg. Pour analyser les relations entre les indicateurs et réduire leur nombre, une méthode dédiée à la classification de variables a été utilisée sur la base des 120 indicateurs considérés. 14 groupes d'indicateurs ont été identifiés permettant d'obtenir autant de variables synthétiques. L'interprétation et la cartographie de ces variables synthétiques ont dans un deuxième temps permis de réaliser la classification non supervisée des mailles de Strasbourg.

Pour la réalisation de la classification avec l'algorithme GNG, il a été nécessaire de déterminer le nombre de classes inhérentes aux données selon que les mailles appartiennent ou non à ce qui a été défini comme urbain dans le chapitre 6. La statistique *gap* a été utilisée et a révélé deux niveaux de classification composés de 3 et de 25 types pour les mailles urbaines et de 5 et 17 types pour les non urbaines. Ces différents types ont été interprétés ce qui a permis de réaliser pour le niveau 2 de la classification, une typologie de la structure urbaine en 17 classes. Cette typologie a ensuite pu être évaluée grâce à la BDOCS de la région Alsace et la base Urban Atlas. Cette évaluation a montré que globalement les performances étaient bonnes avec une différence marquée entre les mailles urbaines et non urbaines. En effet, la classification des mailles non urbaines s'est révélée performante, car les classes obtenues sont plus liées à l'occupation du sol. Pour les mailles urbaines par contre les résultats sont plus contrastés. Il existe de réelles confusions entre certaines classes telles que les tissus de type centre-ville et ceux de type collectif ou entre les grandes emprises et l'habitat individuel qui sont liées, soit à un manque d'indicateurs permettant de modéliser ces tissus, soit à certains indicateurs qui peuvent être trop déterminants lors de la classification. Mais cette évaluation a aussi montré que l'utilisation de la BDOCS est difficile pour l'évaluation d'une classification morphologique des tissus notamment pour la classe grande emprise, car cette vérité terrain est plus orientée utilisation du sol pour les espaces artificialisés.

La deuxième étape de ce chapitre a été de valider la méthodologie développée sur l'agglomération rennaise. Les paramètres de la classification des indicateurs et de ceux permettant d'extraire les variables synthétiques ont été appliqués sur les mailles de Rennes tout comme ceux de la classification non supervisée réalisée avec GNG. L'avantage de transposer l'ensemble de ces paramètres est de pouvoir étendre l'interprétation faite sur le site de Strasbourg. L'évaluation de la classification de niveau 2 a montré que

celle-ci est moins performante sur ce site. Dans les espaces non urbains, c'est principalement le réseau de haies denses qui a entraîné des confusions entre les classes agricoles et forêts. Pour les mailles urbaines, c'est l'habitat individuel qui apparaît être le plus touché par la perte de performances. Pour cette classe la réparation des habitations individuelles est beaucoup plus dispersée et le choix des seuils utilisé dans le chapitre précédent déterminés sur le site de Strasbourg nécessiterait d'être adaptés aux spécificités de ce site. Ces résultats sont tout de même intéressants et montrent que si les morphologies sont assez proches la méthode est facilement reproductible, car la création de l'unité d'analyse, le calcul des indicateurs ainsi que la classification des variables et des mailles est en très grande partie automatisable.



## Conclusion de la partie III

La dernière partie de cette thèse a concerné la cartographie de la structure urbaine des agglomérations de Strasbourg et Rennes à une échelle du 1/5000<sup>e</sup> au 1/10 000<sup>e</sup> à partir de données multi-sources.

Le chapitre 6 a eu pour objectif de construire la base de données pour permettre l'analyse et la caractérisation quantitative des trames grise et verte. La première étape a donc été de sélectionner les données permettant de décrire ces deux trames. Plusieurs sources ont été nécessaires : les éléments gris proviennent de la BD TOPO® et de données cadastrales, et les éléments verts du RGP et de résultats de l'extraction de la végétation arborée et herbacée d'images Pléiades réalisée dans la partie II. Ces différentes sources ont permis de constituer une base d'occupation du sol qui a été associée à des données statistiques sur la population issues du carroyage de l'INSEE. Une série d'indicateurs ont été dérivés de ces données souvent utilisées en urbanisme et en écologie du paysage associé à d'autres permettant de décrire les caractéristiques sociodémographiques, la visibilité, l'accessibilité et la centralité des tissus. Le choix d'un découpage était nécessaire pour calculer ces indicateurs et pour rendre compte de l'organisation des tissus. Huit souvent utilisés pour ce type d'application ont été testés et évalués afin de déterminer lequel était le plus adapté à l'étude de la structure urbaine. À partir de cinq critères d'évaluation, ce chapitre a montré que des découpages différents pouvaient être adaptés selon qu'ils décrivent les espaces artificialisés ou non. Un nouveau découpage a ainsi été proposé pour tenter de prendre en compte cette particularité.

Le second volet de cette partie avait ainsi pour objectif la cartographie des trames grise et vertes de Strasbourg et de Rennes. Il s'est donc appuyé sur le découpage et les indicateurs proposés dans le chapitre précédent en utilisant une méthodologie d'extraction de connaissances dans les bases de données. Des variables synthétiques ont tout d'abord été créées afin de diminuer le nombre d'indicateurs. Sur la base de ces nouvelles variables, une classification non supervisée a été réalisée et a permis d'identifier 17 classes. Évaluée par des bases de données existantes, la classification de la structure urbaine de Strasbourg a montré globalement de bonnes performances, mais des résultats qui peuvent encore être perfectible pour les tissus urbains. L'ensemble de la méthodologie mis en place dans cette partie a été réalisé sur le site de Strasbourg puis transposé sur celui de Rennes afin de la valider. Les performances obtenues pour ce second site sont plus faibles, mais sont néanmoins encourageantes et permettent en

partie de s'affranchir de l'interprétation des types obtenus avec la classification non supervisée.







## Conclusion générale

### *Perspectives de recherche*

Cette thèse avait pour objectif de mieux comprendre la structure urbaine par l'articulation de sa trame grise et de sa trame verte. Toutefois le constat a été montré que les bases de données géographiques existantes ne permettaient d'obtenir si elles existent qu'une vision partielle de la morphologie de cette structure à l'échelle des tissus. Il était donc nécessaire de proposer une méthodologie qui permettent d'en rendre compte par sa cartographie à une échelle allant du 1/5000<sup>e</sup> au 1/10 000<sup>e</sup>. De manière générale les cartographies de la ville à cette échelle sont réalisées par interprétation visuelle ce qui représente un coût qui peut être très important. L'objectif a donc été de réaliser cette cartographie de manière la plus automatique et opérationnelle possible en s'appuyant au maximum sur des données existantes notamment celles qui permettent de représenter spatialement les objets urbains à leur niveau élémentaire. Les éléments de la trame grise, les bâtiments et les réseaux de transport, sont disponibles et représentés précisément dans ces bases au contraire des éléments de la trame verte. En effet pour les éléments verts de la structure urbaine, ces données ne sont pas exhaustives et une première étape de ce travail a donc été de cartographier la végétation au sein de la ville. L'objectif était aussi que l'ensemble des méthodologies proposées soit reproductible. Dans ce contexte, la reproductibilité de nos travaux a été validée sur deux espaces urbains aux morphologies urbaines différentes typiques des villes d'Europe occidentale.

Les images aériennes ou satellites sont des sources de données très souvent mobilisées pour extraire des informations sur l'occupation des sols. L'avènement des images satellites THRS à partir des années 2000 telles que les images Pléiades a permis grâce à leur résolution spatiale de discerner des éléments de petites dimensions comme des arbres isolés en fournissant une plus grande répétitivité ainsi qu'une meilleure qualité spectrale que les images aériennes. Toutefois de nombreux verrous méthodologiques et techniques existent toujours afin que l'extraction automatique d'informations à partir de ces images soit opérationnelle. Nous nous sommes ici placés dans le cadre d'une analyse d'image orientée objet qui permet généralement d'obtenir des résultats cartographiques fidèles à la réalité et associée à une classification supervisée adaptée à la recherche de classes connues *a priori*. Ce type d'analyse est composé de plusieurs étapes : la segmentation, le calcul d'attributs, le choix d'exemple d'apprentissage et la classification de l'image. Dans ce cadre, les développements proposés dans la partie II ont permis de lever partiellement

ces verrous. Les défis identifiés sont le choix d'une segmentation adaptée (chapitre 3), la sélection d'échantillons d'apprentissage informatifs afin de limiter l'effort pour leur création (chapitre 4) et enfin la classification d'image THRS sur de larges zones d'étude (chapitre 5).

Afin de déterminer la segmentation la plus adaptée pour identifier la végétation urbaine, deux algorithmes ont été testés, comparés et évalués : Multi-resolution et Mean Shift. Les meilleures performances pour chacune de ces méthodes sont assez similaires, mais l'ensemble des séries testées a montré une plus grande facilité d'utilisation de Multi-resolution. Plusieurs méthodes d'évaluation des segmentations ont aussi été testées, une supervisée et une non supervisée, et comparée à une évaluation par classification qui a servi de référence. L'avantage de ces deux méthodes est de pouvoir sélectionner rapidement les paramètres de segmentation, car elles limitent les temps de traitements. Toutefois les résultats obtenus ont montré que les paramètres optimaux sont très différents selon l'évaluation utilisée et que les performances sont moins bonnes par rapport à celle de référence ce qui rend difficile l'utilisation d'une telle méthode pour le choix d'une segmentation adaptée. Enfin ce chapitre a permis de montrer qu'une sur segmentation est préférable pour obtenir une identification précise de la végétation urbaine.

Le chapitre 4 a proposé des méthodes d'apprentissage actifs multi-classes par régions afin de sélectionner uniquement les échantillons les plus informatifs et les plus divers et ainsi limiter le nombre d'exemples nécessaire pour l'entraînement de la méthode de classification. Trois méthodes ont ainsi été proposées : MCSAL qui sélectionne uniquement les exemples les plus informatifs, MCSAL<sub>D</sub> qui permet d'intégrer un critère de diversité et enfin MCSAL<sub>S</sub> qui offre la possibilité d'intégrer une carte de stratification et donc de connaissances de l'utilisateur. Ces trois méthodes ont montré globalement de bons résultats comparés à un échantillonnage aléatoire stratifié notamment MCSAL<sub>S</sub> qui a permis de réduire fortement les temps d'exécution pour cette sélection d'échantillons. Toutefois les résultats par classes se sont révélés moins convaincants. En effet la transformation du mode de représentation des données, du vecteur au raster par les centres de gravité, crée des pixels vides qui perturbent la sélection d'échantillons. Dans les espaces agricoles notamment cet effet est assez marqué et limite la sélection de nouvelles régions dans ces espaces qui sont pourtant source de nombreuses confusions notamment entre les parcelles de culture et la végétation herbacée. L'apport de différentes sources d'informations complémentaires pour l'extraction de la végétation urbaine a aussi été évalué. Les informations de hauteur sont apparues comme celles les plus utiles pour améliorer les performances, bien plus que celles issues d'images multi temporelles.

Enfin, les méthodes développées précédemment ont été généralisées sur l'ensemble des deux zones d'études dans le chapitre 5. Des stratégies ont dû être mises en place afin de rendre possible cette généralisation et un tuilage de ces deux sites a été effectué afin de permettre la segmentation, le calcul des attributs et la sélection des échantillons. Plusieurs algorithmes ont été testés et optimisés dans ce but et ont permis de réaliser la classification de ces images. L'évaluation de ces classifications après post-traitements a révélé de bonnes performances pour la classe arborée, mais plus limitée pour celle

herbacée. Elles ont permis d'obtenir une base permettant de décrire la végétation arborée et herbacée nécessaire pour décrire la trame verte de la structure urbaine et donc analyser cette structure à l'échelle des tissus.

L'analyse morphologique des tissus urbains nécessite généralement d'avoir un découpage, existant ou construit, afin de quantifier l'organisation et la forme des éléments par des indicateurs dérivés dans ces mailles. De nombreux découpages peuvent être utilisés pour cet objectif et huit identifiés dans la littérature ont été testés. Pour chacune de ces unités d'analyse, une série d'indicateurs a été construite issue principalement de deux disciplines, l'urbanisme et l'écologie du paysage, afin de rendre compte au sein des mailles de l'organisation des éléments constitutifs de ces trames. Pour ne pas seulement décrire la morphologie des tissus, d'autres indicateurs ont été proposés permettant de considérer des caractéristiques sociodémographiques, l'accessibilité et la visibilité des bâtiments et des informations contextuelles pour évaluer la centralité de chaque maille dans la ville. Ces découpages ont été évalués par l'intermédiaire de cinq critères d'évaluation basés sur l'analyse de variance pour trois d'entre eux, la fragmentation des éléments d'occupation des sols et l'incertitude dans l'attribution des classes de tissus. Cette analyse a permis de montrer que les meilleurs découpages sont différents selon le type d'espace, urbain ou non. En effet l'îlot construit à partir des réseaux est apparu le plus adapté pour l'analyse des espaces urbains tandis que, dans les autres espaces, le découpage régulier par un carroyage hexagonal de 125 m est préférable. Le réseau structure fortement le tissu urbain est il est donc logique que ce découpage soit utilisé, mais dans les espaces non artificialisés les îlots sont généralement de très grandes dimensions entraînant ainsi une mixité importante de forme dans chacune des mailles. En considérant la tache urbaine pour définir ces deux types d'espaces, un nouveau découpage a pu être proposé qui prend en compte les découpages en îlots et en mailles hexagonales.

Sur la base de ce découpage et de ces indicateurs, le chapitre 7 a permis de réaliser une classification automatique des structures urbaines de Strasbourg et de Rennes grâce à une méthodologie qui s'appuie sur des méthodes d'extraction de connaissances dans les bases de données. Cette méthodologie se décompose en deux étapes importantes. La première a été de réduire le nombre d'indicateurs et une méthode de classification de variables a été utilisée permettant de retenir 14 variables synthétiques. La deuxième étape a été de réaliser la classification des mailles de Strasbourg et de déterminer le nombre de types. Deux niveaux ont été révélés : un premier composé de 8 types et un second de 42. Après l'interprétation de ce second niveau, 17 classes ont été retenues afin de se rapprocher au plus près des attentes des utilisateurs de ce type de données définies au travers de deux typologies, une grise et une verte, mais aussi de pouvoir comparer cette cartographie à celle existantes sur ce site. Les résultats de cette comparaison ont montré une adéquation assez forte entre la classification proposée et les cartographies existantes dans les espaces non artificialisés, mais plus faibles dans les espaces urbains. L'ensemble du modèle de classification construit sur le site de Strasbourg a ensuite pu être appliqué sur celui de Rennes sans modifier ses paramètres. L'avantage est ici de

s'épargner le temps nécessaire à l'interprétation des différents types d'autant que le terrain est beaucoup moins bien connu. Comme pour Strasbourg, une comparaison a été réalisée montrant des performances en deçà, mais néanmoins assez proche de ce site.

Cette thèse a donc permis d'améliorer la représentation de la ville à deux échelles : celle des éléments urbains et celle des tissus urbains. À l'échelle des éléments, la méthodologie proposée a permis de cartographier précisément la végétation arborée et herbacée sur de grandes zones d'étude par la sélection d'une segmentation adaptée et en limitant l'effort pour la création d'échantillons d'apprentissage. Cette cartographie offre ainsi la possibilité d'étudier et d'analyser de manière plus précise la trame verte qu'avec les bases de données existantes. À l'échelle des tissus, le nouveau maillage de l'espace proposé a permis de réaliser une classification en deux niveaux dont la typologie finale est opérationnelle puisqu'elle est adaptée aux demandes des utilisateurs et qui prend en compte non seulement des critères morphologiques, mais aussi des critères sociodémographiques et contextuels. Cette thèse a aussi démontré l'intérêt de l'exploitation d'images THRS combinée à des bases de données existantes pour la cartographie de la structure urbaine. Il en ressort aussi qu'il est possible d'utiliser des méthodes issues du domaine de l'extraction des connaissances pour produire une cartographie des tissus urbains. Les résultats mettent en évidence de nombreuses possibilités tant d'un point de vue thématique que d'un point de vue méthodologique pour combler certains manques de ce travail et pour rendre plus opérationnelle la démarche méthodologique mise en place et qui constituent des perspectives de recherches à court et moyen termes.

Ici seules deux classes de végétation ont été considérées. Pourtant elle se décompose en trois types principaux : la végétation arborée, arbustive et herbacée. La différence entre celles arborée et arbustive sur des images THRS est très difficile à faire, car les valeurs radiométriques ainsi que leur texture est très similaire et il est donc nécessaire d'avoir recours à des informations sur leur hauteur qui sont les caractéristiques déterminantes permettant de les différencier. Nous en disposons uniquement pour le site Strasbourg et la résolution spatiale de cette donnée est moins bonne que celle des images ce qui peut nuire à leur différenciation. Ces informations peuvent être obtenues par des relevés Lidar de manière très précise, mais qui nécessite donc des campagnes de prises de vue qui peuvent se heurter aussi à un problème d'adéquation au niveau des dates d'acquisition. Les satellites de la constellation Pléiades ont toutefois un avantage, celui de pouvoir acquérir des images en stéréoscopie voire en tri-stéréo permettant avec une unique prise de vue obtenir des informations radiométriques et de hauteurs et donner ainsi la possibilité de différencier ces trois types de végétation. De plus dans ce travail, la végétation n'a pas été caractérisée d'un point de vue morphologique qui permettrait par exemple pour la végétation arborée de définir s'il s'agit de forêt, de haie, d'arbre isolé, etc. Cette caractérisation est importante pour les travaux concernant la trame verte plus spécifiquement dans les espaces agricoles et naturels et pourrait aussi permettre d'enrichir la classification à l'échelle des tissus notamment pour les espaces non urbains.

Sur un plan plus méthodologique, la chaîne de traitement mise en place est aussi perfectible. Tout d'abord, la méthode de segmentation Multi-Resolution utilisée est celle issue du logiciel eCognition. Ce logiciel est propriétaire et représente un coût très important qui limite fortement son usage. D'autres implémentations existent dans des logiciels libres tels que InterImage (Novack et Kux, 2010) ou la plateforme Mustic du laboratoire ICube. Les évaluer et les utiliser si elles sont performantes pourraient permettre une utilisation plus large de ce type de méthodes d'autant que l'ensemble de la méthodologie proposée s'est appuyé sur des outils libres et que la quasi-totalité des traitements en est issue. Les méthodes d'apprentissage actif pourraient aussi être améliorées en considérant une représentation vectorielle des segments, mais aussi les méthodes utilisées dans le chapitre 5 en automatisant le processus de tuilage. Ce processus crée des problèmes de cohérences entre les segments aux limites des tuiles. Bien que ces problèmes n'aient pas été considérés dans cette thèse il serait nécessaire de les prendre en compte et plusieurs méthodes ont d'ores et déjà été proposées et devraient être intégrées dans la méthodologie qui a été proposée (Happ *et al.*, 2010 ; Michel *et al.*, 2012). De plus, il serait intéressant de pouvoir utiliser les exemples déjà labélisés sur un site afin qu'ils puissent enrichir ceux d'un second. Des méthodes de transfert existent qui, combinées à des méthodes d'apprentissage actif, donnent des résultats encourageants (Jun et Ghosh, 2008 ; Persello et Bruzzone, 2012) et il serait intéressant de les développer du fait de la grande répétitivité des satellites actuels. Seules les méthodes d'analyse d'image orientée objet ont été considérées, mais de nouvelles sont en train d'apparaître : celle d'apprentissage profond (*deep learning*) qui permettent d'obtenir de bien meilleures performances pour la classification d'images et ne nécessitent pas un choix d'attributs préalable.

À l'échelle des tissus urbains, de nombreuses perspectives d'amélioration sont aussi possibles. La classification de niveau deux a permis d'obtenir de manière automatique un nombre important de types regroupés afin de répondre à un besoin opérationnel et de pouvoir être comparés aux classes présentes dans les bases de données existantes. L'ensemble des types a été seulement succinctement décrit. Il serait intéressant de pouvoir les confronter à plusieurs applications, par exemple pour évaluer la perméabilité de chacun et ainsi évaluer la connectivité de la trame verte en milieu urbain de manière plus précise ou être comparé à des données microclimatiques. Confronter la cartographie produite dans ce travail à des utilisateurs serait aussi nécessaire afin de mieux l'évaluer et d'en discerner les manques. L'analyse quantitative des différents types a aussi montré l'importance que pouvaient avoir certaines variables synthétiques dans leur construction. C'était ici particulièrement le cas pour celle qui représentait l'accessibilité au réseau de transport qui a pu perturber fortement les résultats liés à la morphologie particulière du réseau de Strasbourg lié à la présence de la frontière avec l'Allemagne. Un travail devrait donc être mené afin de déterminer quels sont les indicateurs les plus pertinents à conserver pour l'étude de la structure urbaine. Toujours au niveau des indicateurs d'autres n'ont pas pu être calculés ici du fait des données à notre disposition. Par exemple la relation du bâtiment avec la rue, aligné ou en retrait, est une caractéristique essentielle du tissu (Panerai *et al.*, 1999), mais n'a pas été intégrée ici.

De plus il serait nécessaire d'améliorer le découpage proposé et de mieux délimiter les mailles que nous avons nommées « urbaines » et « non urbaines ». Nous nous sommes basés sur les parcelles morphologiques. Ce choix, s'il n'a pas été trop préjudiciable pour la cartographie de la structure urbaine de Strasbourg, a construit sur l'agglomération de Rennes de grandes mailles urbaines du fait d'un habitat plus dispersé qui a intégré aux espaces urbains de grandes parties de territoire agricole. Il serait donc nécessaire de revoir la définition que nous avons donnée des mailles urbaines par exemple grâce à une meilleure délimitation de la zone morphologique urbaine (Loriot, 2008).

Deux sites d'études disposant de données similaires ont été considérés dans cette thèse. La méthodologie proposée est potentiellement applicable à tout type de morphologie urbaine et pour tout type de données permettant de décrire ces formes. Appliquer la démarche de cette thèse sur d'autres sites disposant de données différentes serait intéressant pour mieux vérifier son caractère reproductible et mieux identifier ses limites. De plus, une base de données composée de photographies aériennes anciennes est en cours de réalisation au LIVE en partenariat avec l'Eurométropole de Strasbourg dont les plus anciennes datent de 1932. Il apparaît intéressant de réaliser la méthodologie proposée afin d'effectuer un suivi temporel des tissus urbains et d'identifier des trajectoires d'évolutions et ainsi mieux comprendre la morphogénèse des tissus.



## Références bibliographiques

- Abe N. et Mamitsuka H. 1998**, « Query learning strategies using boosting and bagging », *Proceedings of the Fifteenth International Conference on Machine Learning. ICML'98*, p. 9.
- ADEUS. 2014**, « Les îlots de fraîcheur dans la ville », note n° 140. Agence de développement et d'urbanisme de l'agglomération strasbourgeoise.
- Adolphe L. 2001**, « A simplified model of urban morphology: application to an analysis of the environmental performance of cities », *Environment and Planning B : Planning and Design*, vol. 28, n° 2, pp. 183–200.
- Agejedad R. 2009**, *Étalement urbain et évaluation de son impact sur la biodiversité, de la reconstitution des trajectoires à la modélisation prospective. Application à une agglomération de taille moyenne : Rennes Métropole*, Thèse de Doctorat, Université de Rennes 2.
- Ahern J. 1995**, « Greenways as a planning strategy », *Landscape and Urban Planning*, vol. 33, n° 1-3, pp. 131–155.
- Allag-Dhuisme F., Amsallem J., Barthod C., Deshayes M., Graffin V., Lefeuvre C., Salles E., Barnetche C., Brouard-Masson J., Delaunay A. et al. 2010**, « Choix stratégiques de nature à contribuer à la préservation et à la remise en bon état des continuités écologiques – premier document en appui à la mise en œuvre de la Trame verte et bleue en France », Proposition issue du comité opérationnel Trame verte et bleue, ministère de l'Écologie, de l'Énergie, du Développement durable et de la Mer.
- Allain R. 2004**, *Morphologie urbaine. Géographie, aménagement et architecture de la ville*, U Géographie, Armand Colin, Paris, France.
- Antoni J.P. 2001**, « Urban sprawl modelling : A methodological approach », *Cybergeo : European Journal of Geography*, 12<sup>ème</sup> Colloque Européen de Géographie Théorique et Quantitative.
- APUR. 2007**, « Consommation d'énergie et émissions de gaz à effet de serre liées au chauffage des résidences principales parisiennes », Atelier Parisien d'Urbanisme.
- Arantes L., Marry S., Baverel O. et Quenard D. 2016**, « Efficacité énergétique et formes urbaines : Élaboration d'un outil d'optimisation morpho-énergétique », *Cybergeo : European Journal of Geography*.

- Ardila J.P., Bijker W., Tolpekin V.A. et Stein A.** 2012, « Context-sensitive extraction of tree crown objects in urban areas using VHR satellite images », *International Journal of Applied Earth Observation and Geoinformation*, vol. 15, pp. 57 – 69.
- Armson D., Stringer P. et Ennos A.** 2013, « The effect of street trees and amenity grass on urban surface water runoff in Manchester, UK », *Urban Forestry & Urban Greening*, vol. 12, n° 3, pp. 282–286.
- Arrif T., Blanc N. et Clergeau P.** 2011, « Trame verte urbaine, un rapport Nature – Urbain entre géographie et écologie », *Cybergeo : European Journal of Geography*.
- Aubrecht C., Steinnocher K., Hollaus M. et Wagner W.** 2009, « Integrating earth observation and GIScience for high resolution spatial and functional modeling of urban land use », *Computers, Environment and Urban Systems*, vol. 33, n° 1, pp. 15–25.
- AUCAME.** 2010, « Répertoire des formes urbaines résidentielles de Caen-Métropole », Agence d'études d'Urbanisme de Caen-Métropole.
- AUDIAR.** 2015, « Nature en ville à Rennes. Contribution au projet urbain Rennes 2030 », Agence d'urbanisme de Rennes, Rennes, France.
- Baatz M. et Schäpe A.** 2000, « Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation », *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg 2000*, édité par J. Strobl, T. Blaschke et G. Griesebner, Herbert Wichmann Verlag, pp. 11–23.
- Baço F., Lobo V. et Painho M.** 2005a, « The self-organizing map, the Geo-SOM, and relevant variants for geosciences », *Computers & Geosciences*, vol. 31, n° 2, pp. 155–163.
- Baço F., Lobo V. et Painho M.** 2005b, « Self-organizing maps as substitutes for k-means clustering », *International Conference on Computational Science – ICCS 2005*, édité par V.S. Sunderarm, G. van Albada, P. Sloot et J.J. Dongarra, Lecture Notes in Computer Science, vol. 3516, pp. 476–483.
- Badariotti D.** 2005, « Des fractales pour l'urbanisme ? Quelques pistes de réflexion à partir de l'exemple de Strasbourg-Kehl », *Cahiers de Géographie du Québec*, vol. 49, n° 137, pp. 133–156.
- Ball G.H. et Hall D.J.** 1965, « ISODATA, a novel method of data analysis and pattern classification », Université de Stanford, Stanford, Californie, États-unis.
- Bannari A., Morin D., Bonn F. et Huete A.R.** 1995, « A review of vegetation indices », *Remote Sensing Reviews*, vol. 13, n° 1-2, pp. 95–120.
- Banzhaf E. et Höfer R.** 2008, « Monitoring urban structure types as spatial indicators with CIR aerial photographs for a more effective urban environmental management », *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, n° 2, pp. 129–138.
- Barbakh W.A., Wu Y. et Fyfe C.** 2009, « Review of clustering algorithms », *Non-standard parameter adaptation for exploratory data analysis*, n° 249, Studies in Computational Intelligence, Springer Berlin Heidelberg, pp. 7–28.

- Barnsley M.J. et Barr S.L. 1997**, « Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system », *Computers, Environment and Urban Systems*, vol. 21, n° 3-4, pp. 209–225.
- Barnsley M.J. et Barr S.L. 2000**, « Monitoring urban land use by earth observation », *Surveys in Geophysics*, vol. 21, n° 2, pp. 269–289.
- Barnsley M.J., Steel A. et Barr S. 2003**, « Determining urban land use through an analysis of the spatial composition of buildings identified in LIDAR and multispectral image data », *Remotely sensed cities*, édité par V. Mesev, Taylor and Francis, Londres, Royaume-Uni, pp. 47–82.
- Baro J., Bonin O. et Hubert J.P. 2016**, « Elaboration d'un zonage de tissus urbains. Introduction de la structure dans un référentiel carroyé », *Revue Internationale de Géomatique*, vol. 26, n° 1, pp. 33–53.
- Barreto L., Ribeiro M., Veldkamp A., van Eupen M., Kok K. et Pontes E. 2010**, « Exploring effective conservation networks based on multi-scale planning unit analysis. A case study of the Balsas sub-basin, Maranhão State, Brazil », *Ecological Indicators*, vol. 10, n° 5, pp. 1055–1063.
- Bauer T. et Steinnocher K. 2001**, « Per-parcel land use classification in urban areas applying a rule-based technique », *GeoBIT/GIS*, vol. 6, pp. 24–27.
- Bechtel B., Alexander P., Böhner J., Ching J., Conrad O., Feddema J., Mills G., See L. et Stewart I. 2015**, « Mapping Local Climate Zones for a worldwide database of the form and function of cities », *ISPRS International Journal of Geo-Information*, vol. 4, n° 1, pp. 199–219.
- Behnisch M. et Ultsch A. 2015**, « Knowledge discovery in spatial planning data: A concept for cluster understanding », *Computational Approaches for Urban Environments*, édité par M. Helbich, J. Jokar Arsanjani et M. Leitner, Geotechnologies and the Environment, vol. 13, Springer International Publishing, pp. 49–75.
- Belgiu M. et Drăguț L. 2014**, « Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 96, pp. 67–75.
- Belgiu M., Drăguț L. et Strobl J. 2014**, « Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 205–215.
- BenDor T.K., Metcalf S.S., Fontenot L.E., Sangunett B. et Hannon B. 2006**, « Modeling the spread of the Emerald Ash Borer », *Ecological Modelling*, vol. 197, n° 1-2, pp. 221–236.
- Benz U.C., Hofmann P., Willhauck G., Lingenfelder I. et Heynen M. 2004**, « Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, n° 3-4, pp. 239–258.

- Berghauser Pont M. et Haupt P.** 2005, « The Spacemate: Density and the typomorphology of the urban fabric », *Nordic Journal of Architectural Research*, vol. 4, pp. 55–68.
- Bergoënd A., Clergeau P., Blanc N. et Cormier L.** 2013, « Stratégie de mise en œuvre », *Trames vertes urbaines. De la recherche scientifique au projet urbain*, édité par P. Clergeau et N. Blanc, Éditions du Moniteur, Paris, France, pp. 266–333.
- Bernabé A., Calmet I., Musy M., Bocher E. et Andrieu H.** 2013, « Classification automatique des tissus urbains par la méthode des nuées dynamiques », *31<sup>e</sup> Rencontre de l'AUGC*.
- Betbeder J., Nabucet J., Pottier E., Baudry J., Corgne S. et Hubert-Moy L.** 2014, « Detection and characterization of hedgerows using TerraSAR-X imagery », *Remote Sensing*, vol. 6, n° 5, pp. 3752–3769.
- Beygelzimer A., Kakadet S., Langford J., Arya S., Mount D. et Li S.** 2013, « FNN: Fast Nearest Neighbor search algorithms and applications », R package version 1.1.
- Birch C.P., Oom S.P. et Beecham J.A.** 2007, « Rectangular and hexagonal grids used for observation, experiment and simulation in ecology », *Ecological Modelling*, vol. 206, n° 3-4, pp. 347–359.
- Blanc N.** 2009, « Vers un urbanisme écologique ? », *Urbia – Les cahiers du développement durable*, n° 8, pp. 39–59.
- Blanc N., Glatron S. et Schmitt G.** 2012, « Trames vertes urbaines : recherches en sciences humaines et sociales », *Développement durable et territoires*, vol. 3, n° 2.
- Blaschke T.** 2010, « Object based image analysis for remote sensing », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, n° 1, pp. 2–16.
- Blaschke T., Hay G.J., Kelly M., Lang S., Hofmann P., Addink E., Queiroz Feitosa R., van der Meer F., van der Werff H., van Coillie F. et Tiede D.** 2014, « Geographic Object-Based Image Analysis - Towards a new paradigm », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 180–191.
- Blaschke T. et Strobl J.** 2001, « What's wrong with pixels? Some recent developments interfacing remote sensing and GIS », *GeoBIT/GIS*, vol. 6, n° 1, pp. 12–17.
- Boltz S., Debreuve E. et Barlaud M.** 2007, « kNN-based high-dimensional Kullback-Leibler distance for tracking », *Eighth International Workshop on Image Analysis for Multimedia Interactive Services. WIAMIS '07*, p. 16.
- Bolund P. et Hunhammar S.** 1999, « Ecosystem services in urban areas », *Ecological Economics*, vol. 29, n° 2, pp. 293–301.
- Bondu A. et Lemaire V.** 2008, « État de l'art sur les méthodes statistiques d'apprentissage actif », *Revue des Nouvelles Technologies de l'Information*, vol. RNTI-A-2, Apprentissage Artificiel et Fouille de Données, pp. 207–226.

- Bonhomme M.** 2013, *Contribution à la génération de bases de données multi-scalaires et évolutives pour une approche pluridisciplinaire de l'énergie urbaine*, Thèse de Doctorat, Université de Toulouse.
- Borisov A., Tuv E. et Runger G.** 2011, « Active batch learning with stochastic query-by-forest (SQBF) », *JMLR : Workshop on Active Learning and Experimental Design*, édité par I. Guyon, G. Cawley, G. Dror, V. Lemaire et A. Statnikov, vol. 16, pp. 59–69.
- Boser B.E., Guyon I.M. et Vapnik V.N.** 1992, « A training algorithm for optimal margin classifiers », *Proceedings of the Fifth Annual Workshop on Computational Learning Theory. COLT '92*, pp. 144–152.
- Boussard H. et Baudry J.** 2014, « Chloe2012: A software for landscape pattern analysis », Laboratoire d'agroécologie du paysage du département Science pour l'Action et le Développement – Institut National de la Recherche Agronomique.
- Boutefeu E., Arnould P., Toussaint J.Y. et Michoudet C.** 2008, « Quels espaces verts pour la ville de demain ? », *Cafés Géographiques de Lyon*.
- Breheny M.** 1997, « Urban compaction: feasible and acceptable? », *Cities*, vol. 14, n° 4, pp. 209–217.
- Breiman L.** 2001, « Random Forests », *Machine Learning*, vol. 45, n° 1, pp. 5–32.
- Breiman L., Friedland J., Stone C.J. et Olshen R.A.** 1984, *Classification and regression trees*, Wadsworth Statistics/Probability Series, Chapman and Hall/CRC, New York, États-Unis.
- Brennan S.P. et Schnell G.D.** 2005, « Relationship between bird abundances and landscape characteristics: The influence of scale », *Environmental Monitoring and Assessment*, vol. 105, n° 1-3, pp. 209–228.
- Brunet R., Ferras R. et Théry H.** 1993, *Les mots de la géographie, dictionnaire critique*, 3<sup>e</sup> édition, Collection Dynamiques du territoire, Reclus – La Documentation Française, Montpellier, France.
- Burel F. et Baudry J.** 1999, *Écologie du paysage – Concepts, méthodes et applications*, Lavoisier.
- Burton E.** 2000, « The compact city: Just or just compact? A preliminary analysis », *Urban Studies*, vol. 37, n° 11, pp. 1969–2006.
- Büttner G. et Kosztra B.** 2012, « CLC2012 Addendum to CLC2006 Technical Guidelines », European Environment Agency.
- Calinski T. et Harabasz J.** 1974, « A dendrite method for cluster analysis », *Communications in Statistics - Theory and Methods*, vol. 3, n° 1, p. 27.
- Campbell C., Cristianini N. et Smola A.J.** 2000, « Query learning with large margin classifiers », *Proceedings of the Seventeenth International Conference on Machine Learning. ICML '00*, pp. 111–118.
- Candillier L.** 2006, *Contextualisation, visualisation et évaluation en apprentissage non supervisé*, Thèse de Doctorat, Université Charles de Gaulle - Lille III.

- Canty M.J. et Nielsen A.A. 2008**, « Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation », *Remote Sensing of Environment*, vol. 112, n° 3, pp. 1025–1036.
- Carleer A., Debeir O. et Wolff E. 2005**, « Assessment of very high spatial resolution satellite image segmentations », *Photogrammetric Engineering & Remote Sensing*, vol. 71, n° 11, pp. 1285–1294.
- Carr D.B., Olsen A.R. et White D. 1992**, « Hexagon mosaic maps for display of univariate and bivariate geographical data », *Cartography and Geographic Information Science*, vol. 19, n° 4, pp. 228–236.
- CERTU. 2011**, « Traitements géomatiques par carreaux pour l’observation des territoires », Centre d’études sur les réseaux, les transports, l’urbanisme et les constructions publiques.
- Cervero R. et Kockelman K. 1997**, « Travel demand and the 3Ds: Density, diversity, and design », *Transportation Research Part D : Transport and Environment*, vol. 2, n° 3, pp. 199–219.
- Charrad M., Ghazzali N., Boiteau V. et Niknafs A. 2014**, « NbClust: An R package for determining the relevant number of clusters in a data set », *Journal of Statistical Software*, vol. 61, n° 1, p. 36.
- Chavent M., Kuentz-Simonet V., Liquet B. et Saracco J. 2012**, « ClustOfVar: An R package for the clustering of variables », *Journal of Statistical Software*, vol. 50, n° 13, p. 16.
- Chehata N., Guo L. et Mallet C. 2009**, « Airborne Lidar feature selection for urban classification using Random Forests », *Proceedings of the ISPRS Workshop : Laserscanning’09*, pp. 207–212.
- Chen C., Liaw A. et Breiman L. 2004**, « Using Random Forest to learn imbalanced data », Rapport technique n° 666, Université de Californie à Berkeley.
- Chen D.M. et Stow D. 2002**, « The effect of training strategies on supervised classification at different spatial resolutions », *Photogrammetric Engineering & Remote Sensing*, vol. 68, n° 11, pp. 1155–1161.
- Chi M. et Bruzzone L. 2007**, « Semisupervised classification of hyperspectral images by SVMs optimized in the primal », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, n° 6, pp. 1870–1880.
- CIGAL. 2013**, « Interprétation de l’occupation du sol en 2011-2012 et mesure de l’évolution de la consommation de l’espace sur le territoire Alsace. Rapport méthodologique », Coopération pour l’Information Géographique en Alsace.
- Claggett P.R., Jantz C.A., Goetz S.J. et Bisland C. 2004**, « Assessing development pressure in the Chesapeake Bay watershed: An evaluation of two land-use change models », *Environmental Monitoring and Assessment*, vol. 94, n° 1-3, pp. 129–146.

- Claval P.** 1986, « Les jardins et la ville : une étude géographique », *Treballs de la Societat Catalana de Geografia*, n° 7-8, pp. 45–59.
- Clergeau P.** 2007, *Une écologie du paysage urbain*, Apogée, Paris, France.
- Clergeau P. et Blanc N.** 2013, « Pourquoi des trames vertes dans la ville ? », *Trames vertes urbaines. De la recherche scientifique au projet urbain*, édité par P. Clergeau et N. Blanc, Éditions du Moniteur, Paris, France, pp. 15–23.
- Clergeau P. et Hubert-Moy L.** 2011, « L'émergence d'une problématique scientifique en relation avec une demande sociétale », *Ville et biodiversité. Enseignements d'une recherche pluridisciplinaire*, Espaces et territoires, Presses universitaires de Rennes, pp. 9–28.
- Cleve C., Kelly M., Kearns F.R. et Moritz M.** 2008, « Classification of the wildland urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography », *Computers, Environment and Urban Systems*, vol. 32, n° 4, pp. 317–326.
- Clifton K., Ewing R., Knaap G. et Song Y.** 2008, « Quantitative analysis of urban form: A multidisciplinary review », *Journal of Urbanism : International Research on Placemaking and Urban Sustainability*, vol. 1, n° 1, pp. 17–45.
- Clinton N., Holt A., Scarborough J., Yan L. et Gong P.** 2010, « Accuracy assessment measures for object-based image segmentation goodness », *Photogrammetric Engineering & Remote Sensing*, vol. 76, n° 3, pp. 289–299.
- CNES.** 2016, « Spot, Pléiades la haute résolution... Jusqu'où ? », *CNES MAG*, vol. 67, pp. 23–24.
- Coenen-Huther J.** 2007, « Classifications, typologies et rapport aux valeurs », *Revue européenne des sciences sociales*, vol. XLV-138, pp. 27–40.
- Van de Coevering P. et Schwanen T.** 2006, « Re-evaluating the impact of urban form on travel patterns in Europe and North-America », *Transport Policy*, vol. 13, n° 3, pp. 229–239.
- Colaninno N., Roca J. et Pfeffer K.** 2011, « Urban form and compactness of morphological homogeneous districts in Barcelona: Towards an automatic classification of similar built-up structures in the city », *ERSA conference papers*, p. 15.
- Comaniciu D. et Meer P.** 2002, « Mean Shift: A robust approach toward feature space analysis », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, n° 5, pp. 603–619.
- Comber A., Brunson C. et Farmer C.** 2012, « Community detection in spatial networks: Inferring land use from a planar graph of land cover objects », *International Journal of Applied Earth Observation and Geoinformation*, vol. 18, pp. 274–282.
- Combrette P.** 2016, *Contribution des technologies satellitaires Pléiades à l'étude des trames vertes urbaines. Entre maintien des connectivités écologiques et densification des espaces urbains*, Thèse de Doctorat, Université de Toulouse Jean Jaurès.

- Commissariat général au développement durable.** 2015a, « Bilan énergétique de la France pour 2014 », Ministère de l'environnement de l'énergie et de la mer.
- Commissariat général au développement durable.** 2015b, « Les émissions de gaz à effet de serre par secteur en France », Ministère de l'environnement de l'énergie et de la mer.
- Congalton R.G. et Green K.** 2008, *Assessing the accuracy of remotely sensed data: principles and practices*, 2<sup>e</sup> édition, Mapping Science, CRC Press, Boca Raton, Floride, États-Unis.
- Copa L., Tuia D., Volpi M. et Kanevski M.** 2010, « Unbiased query-by-bagging active learning for VHR image classification », *Proceedings of SPIE, Image and Signal Processing for Remote Sensing XVI*, édité par L. Bruzzone, vol. 7830, p. 8.
- Corcoran P., Winstanley A. et Mooney P.** 2010, « Segmentation performance evaluation for object-based remotely sensed image analysis », *International Journal of Remote Sensing*, vol. 31, n° 3, pp. 617–645.
- Cormier L., Lajartre A.B.D. et Carcaud N.** 2010, « La planification des trames vertes, du global au local : réalités et limites », *Cybergeo : European Journal of Geography*.
- Cornuéjols A. et Miclet L.** 2010, *Apprentissage artificiel : concepts et algorithmes*, 2<sup>e</sup> édition, Algorithmes, Eyrolles, Paris, France.
- Costanza R., d'Arge R., de Groot R., Farber S., Grasso M., Hannon B., Limburg K., Naeem S., O'Neill R.V., Paruelo J. et al.** 1997, « The value of the world's ecosystem services and natural capital », *Nature*, vol. 387, n° 6630, pp. 253–260.
- Crawford M.M., Tuia D. et Yang H.L.** 2013, « Active learning: Any value for classification of remotely sensed data ? », *Proceedings of the IEEE*, vol. 101, n° 3, pp. 593–608.
- CREAT.** 2010, « Recherche d'intérêt général et pluridisciplinaire relative aux choix et au calcul d'indicateurs de fragmentation du territoire en Région wallonne », Centre de recherches en aménagement du territoire. Université catholique de Louvain.
- Crowther T.W., Glick H.B., Covey K.R., Bettigole C., Maynard D.S., Thomas S.M., Smith J.R., Hintler G., Duguid M.C., Amatulli G. et al.** 2015, « Mapping tree density at a global scale », *Nature*, vol. 525, n° 7568, pp. 201–205.
- Cushman S.A., McGarigal K. et Neel M.C.** 2008, « Parsimony in landscape metrics: Strength, universality, and consistency », *Ecological Indicators*, vol. 8, n° 5, pp. 691–703.
- Dalla Mura M., Benediktsson J.A., Waske B. et Bruzzone L.** 2010, « Morphological attribute profiles for the analysis of very high resolution images », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, n° 10, pp. 3747–3762.
- Dalponte M., Bruzzone L. et Gianelle D.** 2012, « Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispec-



- tral/hyperspectral images and LiDAR data », *Remote Sensing of Environment*, vol. 123, pp. 258–270.
- Dalponte M., Ørka H.O., Ene L.T., Gobakken T. et Næsset E.** 2014, « Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data », *Remote Sensing of Environment*, vol. 140, pp. 306–317.
- Daniel H., Le Bot J.M., Beaujouan V., Philip F., Toublanc M., Bonin S., Montembault D., Hubert-Moy L. et Ollivier C.** 2013, « Espaces arborés dans les paysages urbains de trois agglomérations de l'ouest de la France », *Trames vertes urbaines. De la recherche scientifique au projet urbain*, édité par P. Clergeau et N. Blanc, Éditions du Moniteur, Paris, France, pp. 125–153.
- Davies D.L. et Bouldin D.W.** 1979, « A cluster separation measure », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, n° 2, pp. 224–227.
- Dearborn D.C. et Kark S.** 2010, « Motivations for conserving urban biodiversity », *Conservation Biology*, vol. 24, n° 2, pp. 432–440.
- Demir B., Minello L. et Bruzzone L.** 2014, « Definition of effective training sets for supervised classification of remote sensing images by a novel cost-sensitive active learning method », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, n° 2, pp. 1272–1284.
- Demir B., Persello C. et Bruzzone L.** 2011, « Batch-mode active-learning methods for the interactive classification of remote sensing images », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, n° 3, pp. 1014–1031.
- Deniau J.** 2013, *Apport de la télédétection à très haute résolution spatiale pour la cartographie de la trame grise et de la trame verte en milieux urbain et périurbain*, Mémoire de Master, Université de Rennes.
- Dey V., Zhang Y. et Zhong M.** 2010, « A review on image segmentation techniques with remote sensing perspective », *Proceedings of the International Society for Photogrammetry and Remote Sensing Symposium. ISPRS10*, vol. XXXVIII (Part 7A), pp. 31–42.
- Dhillon I.S., Marcotte E.M. et Roshan U.** 2003, « Diametrical clustering for identifying anti-correlated gene clusters », *Bioinformatics*, vol. 19, n° 13, pp. 1612–1619.
- Di W. et Crawford M.M.** 2012, « View generation for multiview maximum disagreement based active learning for hyperspectral image classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, n° 5, pp. 1942–1954.
- Diday E.** 1971, « Une nouvelle méthode en classification automatique et reconnaissance des formes la méthode des nuées dynamiques », *Revue de Statistique Appliquée*, vol. 19, n° 2, pp. 19–33.
- Donadieu P. et Fleury A.** 2003, « La construction contemporaine de la ville-campagne en Europe », *Revue de géographie alpine*, vol. 91, n° 4, pp. 19–29.

- Dong P.** 2008, « Generating and updating multiplicatively weighted Voronoi diagrams for point, line and polygon features in GIS », *Computers & Geosciences*, vol. 34, n° 4, pp. 411–421.
- Dopido I., Li J., Marpu P.R., Plaza A., Bioucas Dias J.M. et Benediktsson J.A.** 2013, « Semisupervised self-learning for hyperspectral image classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, n° 7, pp. 4032–4044.
- Doğrusöz E. et Aksoy S.** 2007, « Modeling urban structures using graph-based spatial patterns », *IEEE International Geoscience and Remote Sensing Symposium. IGARSS 2007*, pp. 4826–4829.
- Drăguț L., Csillik O., Eisank C. et Tiede D.** 2014, « Automated parameterisation for multi-scale image segmentation on multiple layers », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 119–127.
- Drăguț L., Tiede D. et Levick S.R.** 2010, « ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data », *International Journal of Geographical Information Science*, vol. 24, n° 6, pp. 859–871.
- Dudoit S. et Fridlyand J.** 2002, « A prediction-based resampling method for estimating the number of clusters in a dataset », *Genome Biology*, vol. 3, n° 7, p. 21.
- Dunn J.C.** 1974, « Well-separated clusters and optimal fuzzy partitions », *Journal of Cybernetics*, vol. 4, n° 1, pp. 95–104.
- Duro D.C., Franklin S.E. et Dubé M.G.** 2012a, « A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery », *Remote Sensing of Environment*, vol. 118, pp. 259–272.
- Duro D.C., Franklin S.E. et Dubé M.G.** 2012b, « Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using Random Forests », *International Journal of Remote Sensing*, vol. 33, n° 14, pp. 4502–4526.
- EEA.** 2006, « Urban sprawl in Europe – The ignored challenge », Rapport technique n° 10/2006, European Environment Agency.
- EEA.** 2007, « CLC2006 technical guidelines », Rapport technique n° 17/2007, European Environment Agency.
- Espindola G.M., Câmara G., Reis I.A., Bins L.S. et Monteiro A.M.** 2006, « Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation », *International Journal of Remote Sensing*, vol. 27, n° 14, pp. 3035–3040.
- Fahrig L.** 2003, « Effects of habitat fragmentation on biodiversity », *Annual Review of Ecology, Evolution, and Systematics*, vol. 34, n° 1, pp. 487–515.
- Fang C.F. et Ling D.L.** 2003, « Investigation of the noise reduction provided by tree belts », *Landscape and Urban Planning*, vol. 63, n° 4, pp. 187–195.
- Fayyad U., Piatetsky-Shapiro G. et Smyth P.** 1996, « From data mining to knowledge discovery in databases », *AI magazine*, vol. 17, n° 3, p. 37.

- Feng Q., Liu J. et Gong J. 2015**, « UAV remote sensing for urban vegetation mapping using Random Forest and texture analysis », *Remote Sensing*, vol. 7, n° 1, pp. 1074–1094.
- Feret J.B., Corbane C. et Alleaume S. 2015**, « Detecting the phenology and discriminating Mediterranean natural habitats with multispectral sensors - An analysis based on multiseasonal field spectra », *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, n° 5, pp. 2294–2305.
- Fernandes M.R., Aguiar F.C., Silva J.a.M., Ferreira M.T. et Pereira J.M. 2014**, « Optimal attributes for the object based detection of giant reed in riparian habitats: A comparative study between airborne High Spatial Resolution and WorldView-2 imagery », *International Journal of Applied Earth Observation and Geoinformation*, vol. 32, pp. 79–91.
- Fišer D., Faigl J. et Kulich M. 2013**, « Growing neural gas efficiently », *Neurocomputing*, vol. 104, pp. 72–82.
- Fisher P. 1997**, « The pixel: A snare and a delusion », *International Journal of Remote Sensing*, vol. 18, n° 3, pp. 679–685.
- Foltête J.C., Clauzel C., Girardet X., Tournant P. et Vuidel G. 2012**, « La modélisation des réseaux écologiques par les graphes paysagers. Méthodes et outils », *Revue internationale de géomatique*, vol. 22, n° 4, pp. 641–658.
- Foody G.M. et Mathur A. 2004**, « Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification », *Remote Sensing of Environment*, vol. 93, n° 1-2, pp. 107–117.
- Forriez M. 2010**, *Caractérisation formelle des structures multi-échelles géographiques en relativité d'échelle : exemples en géographie physique, géographie urbaine, géohistoire et géographie du peuplement*, Thèse de Doctorat, Université d'Avignon.
- Frankhauser P. 1990**, « Aspects fractals des structures urbaines », *Espace géographique*, vol. 19, n° 1, pp. 45–69.
- Fritzke B. 1995**, « A growing neural gas network learns topologies », *Advances in Neural Information Processing Systems 7 – NIPS 1994*, édité par G. Tesauro, D.S. Touretzky et T.K. Leen, pp. 625–632.
- Fukunaga K. et Hostetler L. 1975**, « The estimation of the gradient of a density function, with applications in pattern recognition », *IEEE Transactions on Information Theory*, vol. 21, n° 1, pp. 32–40.
- Gao Y., Mas J.F., Kerle N. et Navarrete Pacheco J.A. 2011**, « Optimal region growing segmentation and its effect on classification accuracy », *International Journal of Remote Sensing*, vol. 32, n° 13, pp. 3747–3763.
- Genuer R. 2010**, *Forêts aléatoires : aspects théoriques, sélection de variables et applications*, Thèse de Doctorat, Université Paris Sud - Paris XI.

- Ghosh A., Fassnacht F.E., Joshi P. et Koch B.** 2014, « A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales », *International Journal of Applied Earth Observation and Geoinformation*, vol. 26, pp. 49–63.
- Gilbert-Norton L., Wilson R., Stevens J. et Beard K.** 2010, « A meta-analytic review of corridor effectiveness », *Conservation Biology*, vol. 24, n° 3, pp. 660–668.
- Gill S., Handley J., Ennos A. et Pauleit S.** 2007, « Adapting cities for climate change: The role of the green infrastructure », *Built Environment*, vol. 33, n° 1, pp. 115–133.
- Girres J.F. et Touya G.** 2010, « Quality assessment of the French OpenStreetMap dataset », *Transactions in GIS*, vol. 14, n° 4, pp. 435–459.
- Gislason P.O., Benediktsson J.A. et Sveinsson J.R.** 2006, « Random Forests for land cover classification », *Pattern Recognition Letters*, vol. 27, n° 4, pp. 294–300.
- GMES.** 2006, « Mapping guide for an European Urban Atlas », Global Monitoring for Environment and Security.
- Groupe ElementR.** 2014, *R et espace. Traitement de l'information géographique*, Framabook.
- Guerschman J.P., Paruelo J.M., Bella C.D., Giallorenzi M.C. et Pacin F.** 2003, « Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data », *International Journal of Remote Sensing*, vol. 24, n° 17, pp. 3381–3402.
- Guo L., Chehata N., Mallet C. et Boukir S.** 2011, « Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, n° 1, pp. 56–66.
- Hagenauer J.** 2015, « Weighted merge context for clustering and quantizing spatial data with self-organizing neural networks », *Journal of Geographical Systems*, vol. 18, n° 1, pp. 1–15.
- Hagenauer J. et Helbich M.** 2013, « Contextual neural gas for spatial clustering and analysis », *International Journal of Geographical Information Science*, vol. 27, n° 2, pp. 251–266.
- Haklay M.** 2010, « How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets », *Environment and Planning B : Planning and Design*, vol. 37, n° 4, pp. 682–703.
- Hamaina R., Leduc T. et Moreau G.** 2012, « Towards urban fabrics characterization based on buildings footprints », *Bridging the Geographic Information Sciences*, édité par J. Gensel, D. Josselin et D. Vandenbroucke, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, pp. 327–346.
- Happ P., Ferreira R.S., Bentes C., Costa G. et Feitosa R.Q.** 2010, « Multiresolution segmentation: A parallel approach for high resolution image segmentation in multicore architectures », *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, n° 4/C7, p. 6.

- Haralick R.M., Shanmugam K. et Dinstein I. 1973**, « Textural features for image classification », *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, n° 6, pp. 610–621.
- Hartigan J.A. et Wong M.A. 1979**, « A k-means clustering algorithm », *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, n° 1, pp. 100–108.
- Hay G. et Castilla G. 2008**, « Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline », *Object-Based Image Analysis*, édité par T. Blaschke, S. Lang et G. Hay, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, pp. 75–89.
- Hay G., Niemann K. et McLean G. 1996**, « An object-specific image-texture analysis of H-resolution forest imagery », *Remote Sensing of Environment*, vol. 55, n° 2, pp. 108–122.
- He H. et Garcia E.A. 2009**, « Learning from imbalanced data », *IEEE Transaction on Knowledge and Data Engineering*, vol. 21, n° 9, pp. 1263–1284.
- Hecht R., Herold H., Meinel G. et Buchroithner M. 2013a**, « Automatic derivation of urban structure types from topographic maps by means of image analysis and machine learning », *26<sup>th</sup> International Cartographic Conference*, p. 18.
- Hecht R., Kunze C. et Hahmann S. 2013b**, « Measuring completeness of building footprints in OpenStreetMap over space and time », *ISPRS International Journal of Geo-Information*, vol. 2, n° 4, pp. 1066–1091.
- Hecht R., Meinel G. et Buchroithner M. 2015**, « Automatic identification of building types based on topographic databases - a comparison of different data sources », *International Journal of Cartography*, vol. 1, n° 1, pp. 18–31.
- Heinzel J. et Koch B. 2011**, « Exploring full-waveform LiDAR parameters for tree species classification », *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, n° 1, pp. 152–160.
- Hermosilla T., Palomar-Vázquez J., Balaguer-Beser A., Balsa-Barreiro J. et Ruiz L.A. 2014**, « Using street based metrics to characterize urban typologies », *Computers, Environment and Urban Systems*, vol. 44, pp. 68–79.
- Hermosilla T., Ruiz L., Recio J. et Cambra-López M. 2012**, « Assessing contextual descriptive features for plot-based classification of urban areas », *Landscape and Urban Planning*, vol. 106, n° 1, pp. 124–137.
- Herold M., Couclelis H. et Clarke K.C. 2005**, « The role of spatial metrics in the analysis and modeling of urban land use change », *Computers, Environment and Urban Systems*, vol. 29, n° 4, pp. 369–399.
- Herold M., Gardner M. et Roberts D. 2003a**, « Spectral resolution requirements for mapping urban areas », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, n° 9, pp. 1907–1919.

- Herold M., Liu X.H. et Clarke K.C. 2003b**, « Spatial metrics and image texture for mapping urban land use », *Photogrammetric Engineering & Remote Sensing*, vol. 69, n° 9, pp. 991–1001.
- Herrault P.A. 2015**, *Extraction de fragments forestiers et caractérisation de leurs évolutions spatio-temporelles pour évaluer l'effet de l'histoire sur la biodiversité : une approche multi-sources*, Thèse de Doctorat, Université Toulouse - Jean Jaurès.
- Herrault P.A., Larrieu L., Cordier S., Gimmi U., Lachat T., Ouin A., Sarthou J.P. et Sheeren D. 2016**, « Combined effects of area, connectivity, history and structural heterogeneity of woodlands on the species richness of hoverflies (Diptera: Syrphidae) », *Landscape Ecology*, vol. 31, n° 4, pp. 877–893.
- Höfle B., Hollaus M. et Hagenauer J. 2012**, « Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 134–147.
- Hofmann P., Strobl J. et Nazarkulova A. 2011**, « Mapping green spaces in Bishkek – How reliable can spatial analysis be? », *Remote Sensing*, vol. 3, n° 12, pp. 1088–1103.
- Holden E. et Norland I. 2005**, « Three challenges for the compact city as a sustainable urban form: Household consumption of energy and transport in eight residential areas in the greater Oslo Region », *Urban Studies*, vol. 42, n° 12, pp. 2145–2166.
- Hoover A., Jean-Baptiste G., Jiang X., Flynn P., Bunke H., Goldgof D., Bowyer K., Eggert D., Fitzgibbon A. et Fisher R. 1996**, « An experimental comparison of range image segmentation algorithms », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, n° 7, pp. 673–689.
- Hostein C., Jacqueminet C., Kessler F., Pradinas R. et Soulier A. 2014**, « Trame écologique agropastorale du Massif central : de l'approche cartographique globale par grands types de milieux à une approche cartographique affinée des végétations agropastorales », Conservatoire botanique national du Massif central.
- Houborg R., Fisher J.B. et Skidmore A.K. 2015**, « Advances in remote sensing of vegetation function and traits », *International Journal of Applied Earth Observation and Geoinformation*, vol. 43, pp. 1–6.
- Hu S. et Wang L. 2013**, « Automated urban land-use classification with remote sensing », *International Journal of Remote Sensing*, vol. 34, n° 3, pp. 790–803.
- Huang J., Lu X. et Sellers J.M. 2007**, « A global comparative analysis of urban form: Applying spatial metrics and remote sensing », *Landscape and Urban Planning*, vol. 82, n° 4, pp. 184–197.
- Hubert-Moy L., Nabucet J., Vannier C. et Lefebvre A. 2012**, « Cartographie des continuités écologiques : quelles données pour quelles échelles territoriales? Application à la sous-trame forestière », *Revue internationale de géomatique*, vol. 22, n° 4, pp. 619–640.
- Huberty C.J. et Olejnik S. 2006**, *Applied MANOVA and discriminant analysis*, 2<sup>e</sup> édition, Wiley Series in Probability and Statistics, John Wiley & Sons, New-York.

- Hui S.C.** 2001, « Low energy building design in high density urban cities », *Renewable Energy*, vol. 24, n° 3-4, pp. 627–640.
- Hussain E. et Shan J.** 2015, « Object-based urban land cover classification using rule inheritance over very high-resolution multisensor and multitemporal data », *GIScience & Remote Sensing*, vol. 53, n° 2, pp. 164–182.
- Hussain M., Barr R. et Chen D.** 2012, « Building-based urban land use classification from vector databases in Manchester, UK », *20<sup>th</sup> International Conference on Geoinformatics*, p. 7.
- IGN.** 2010, « BD PARCELLAIRE® Version 1.2 – Descriptif de contenu », Institut national de l'information géographique et forestière.
- IGN.** 2013, « Production BD occupation du sol à grande échelle nationale. Méthodologie utilisée pour la zone test de Maubourguet (65) », Institut national de l'information géographique et forestière.
- IGN.** 2014a, « BD Forêt® Version 2 – Descriptif de contenu », Institut national de l'information géographique et forestière.
- IGN.** 2014b, « BD Topo® Version 2.1 – Descriptif de contenu », Institut national de l'information géographique et forestière.
- INSEE.** 2013, « Données carroyées à 200 mètres, mode opératoire sous QGIS », Institut National de la Statistique et des Études Économiques.
- Iovan C., Boldo D., Cord M. et Erikson M.** 2007, « Automatic extraction and classification of vegetation areas from high resolution images in urban areas », *Image Analysis. 15<sup>th</sup> Scandinavian Conference on Image Analysis. SICA 2007*, édité par B.K. Ersbøll, et K.S. Pedersen, Lecture Notes in Computer Science, vol. 4522, pp. 858–867.
- Irons J.R., Markham B.L., Nelson R.F., Toll D.L., Williams D.L., Latty R.S. et Stauffer M.L.** 1985, « The effects of spatial resolution on the classification of Thematic Mapper data », *International Journal of Remote Sensing*, vol. 6, n° 8, pp. 1385–1403.
- Jain A.K.** 2010, « Data clustering: 50 years beyond k-means », *Pattern Recognition Letters*, vol. 31, n° 8, pp. 651–666.
- Jain A.K., Murty M.N. et Flynn P.J.** 1999, « Data clustering: a review », *ACM Computing Surveys*, vol. 31, n° 3, pp. 264–323.
- Joliveau T.** 2004, *Géomatique et gestion environnementale du territoire. Recherches sur un usage géographique des SIG. Volume 1*, Mémoire d'Habilitation à Diriger des Recherches. Université de Rouen.
- Jongman R.H., Külvik M. et Kristiansen I.** 2004, « European ecological networks and greenways », *Landscape and Urban Planning*, vol. 68, n° 2-3, pp. 305–319.
- Juel A., Groom G.B., Svenning J.C. et Ejrnæs R.** 2015, « Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data », *International Journal of Applied Earth Observation and Geoinformation*, vol. 42, pp. 106–114.

- Jun G. et Ghosh J. 2008**, « An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis », *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2008*, vol. 1, pp. 52–55.
- Jun G., Vatsavai R. et Ghosh J. 2009**, « Spatially adaptive classification and active learning of multispectral data with gaussian processes », *IEEE International Conference on Data Mining Workshops, 2009. ICDMW '09*, pp. 597–603.
- Kaiser H.F. 1960**, « The application of electronic computers to factor analysis », *Educational and Psychological Measurement*, vol. 20, n° 1, pp. 141–151.
- Kaufmann L. et Rousseeuw P. 1987**, « Clustering by means of medoids », *Statistical Data Analysis Based on the  $L_1$ -Norm and Related Methods*, édité par Y. Dodge, Elsevier Science, pp. 405–416.
- Kelly M., Tuxen K.A. et Stralberg D. 2011**, « Mapping changes to vegetation pattern in a restoring wetland: Finding pattern metrics that are consistent across spatial scale and time », *Ecological Indicators*, vol. 11, n° 2, pp. 263–273.
- Key T., Warner T.A., McGraw J.B. et Fajvan M.A. 2001**, « A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest », *Remote Sensing of Environment*, vol. 75, n° 1, pp. 100–112.
- Kim M., Madden M. et Warner T.A. 2008**, « Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery », *Object-Based Image Analysis*, édité par T. Blaschke, S. Lang et G.J. Hay, Lecture Notes in Geoinformation and Cartography, Springer Science - Business Media, pp. 291–307.
- Kohavi R. 1995**, « A study of cross-validation and bootstrap for accuracy estimation and model selection », *Proceedings of the 14<sup>th</sup> International Joint Conference on Artificial Intelligence. IJCAI'95*, vol. 2, p. 7.
- Kohonen T. 1982**, « Self-organized formation of topologically correct feature maps », *Biological Cybernetics*, vol. 43, n° 1, pp. 59–69.
- Kohonen T. 2013**, « Essentials of the self-organizing map », *Neural Networks*, vol. 37, pp. 52–65.
- Kuentz-Simonet V., Labenne A. et Rambonilaza T. 2016**, « Using ClustOfVar to construct quality of life indicators for vulnerability assessment municipality trajectories in southwest France from 1999 to 2009 », *Social Indicators Research*, p. 25.
- Kuentz-Simonet V., Lyser S., Candau J., Deuffic P., Chavent M. et Saracco J. 2013**, « Une approche par classification de variables pour la typologie d'observations : le cas d'une enquête agriculture et environnement », *Journal de la Société Française de Statistique*, vol. 154, n° 2, pp. 37–63.
- Kuhn M. et Johnson K. 2013**, *Applied predictive modeling*, Springer-Verlag, New York, États-Unis.



- Kurtz C.** 2012, *Une approche collaborative segmentation - classification pour l'analyse descendante d'images multirésolutions*, Thèse de Doctorat, Université de Strasbourg.
- Kurtz C., Passat N., Gañçarski P. et Puissant A.** 2012, « Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology », *Pattern Recognition*, vol. 45, n° 2, pp. 685–706.
- La Rosa D. et Wiesmann D.** 2013, « Land cover and impervious surface extraction using parametric and non-parametric algorithms from the open-source software R: An application to sustainable urban planning in Sicily », *GIScience & Remote Sensing*, vol. 50, n° 2, pp. 231–250.
- Laben C.A. et Brower B.V.** 2000, *Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening*, États-Unis, Brevet n° 6011875, Janvier, 2000.
- Lackner M. et Conway T.M.** 2008, « Determining land-use information from land cover through an object-oriented classification of IKONOS imagery », *Canadian Journal of Remote Sensing*, vol. 34, n° 2, pp. 77–92.
- Labiberte A., Koppa J., Fredrickson E. et Rango A.** 2006, « Comparison of nearest neighbor and rule-based decision tree classification in an object-oriented environment », *IEEE International Symposium on Geoscience and Remote Sensing. IGARSS 2006*, pp. 3923–3926.
- Langanke T.** 2013, « GIO land (GMES/Copernicus initial operations land) High Resolution Layers (HRLs) – Summary of product specifications », European Environment Agency.
- Le Néchet F.** 2015, « De la forme urbaine à la structure métropolitaine : une typologie de la configuration interne des densités pour les principales métropoles européennes de l'Audit Urbain », *Cybergeo : European Journal of Geography*.
- Le Néchet F. et Aguiléra A.** 2012, « Forme urbaine et mobilité domicile – travail dans 13 aires urbaines françaises : une analyse multiéchelle », *Recherche Transports Sécurité*, vol. 2012, n° 03-04, pp. 259–270.
- Lebart L., Piron M. et Morineau A.** 2006, *Statistique exploratoire multidimensionnelle. Visualisation et inférence en fouille de données*, 4<sup>e</sup> édition, Sciences Sup, Dunod, Paris, France.
- Lefebvre A.** 2011, *Contribution de la texture pour l'analyse d'images à très haute résolution spatiale : application à la détection de changement en milieu périurbain*, Thèse de Doctorat, Université de Rennes 2, Université Européenne de Bretagne.
- Lefebvre A., Corpetti T. et Hubert-Moy L.** 2011, « Ondelettes et théorie des évidences pour la classification orientée-objet. Caractérisation et suivi des changements d'occupation des sols de la métropole de Rennes », *Revue internationale de géomatique*, vol. 21, n° 3, pp. 297–325.
- Lévy A.** 2005, « Formes urbaines et significations : revisiter la morphologie urbaine », *Espaces et sociétés*, vol. 4, n° 122, pp. 25–48.

- Lewis D.D. et Gale W.A. 1994**, « A sequential algorithm for training text classifiers », *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '94*, pp. 3–12.
- Li X. et Shao G. 2013**, « Object-based urban vegetation mapping with high-resolution aerial photography as a single data source », *International Journal of Remote Sensing*, vol. 34, n° 3, pp. 771–789.
- Li X., Zhang C., Li W., Ricard R., Meng Q. et Zhang W. 2015**, « Assessing street-level urban greenery using Google Street View and a modified green view index », *Urban Forestry & Urban Greening*, vol. 14, n° 3, pp. 675–685.
- Liaw A. et Wiener M. 2002**, « Classification and regression by randomForest », *R News*, vol. 2, n° 3, pp. 18–22.
- Liénard S. et Clergeau P. 2011**, « Trame Verte et Bleue : utilisation des cartes d'occupation du sol pour une première approche qualitative de la biodiversité », *Cybergeo : European Journal of Geography*.
- Lin J.J. et Yang A.T. 2009**, « Structural analysis of how urban form impacts travel demand: Evidence from Taipei », *Urban Studies*, vol. 46, n° 9, pp. 1951–1967.
- Lindner M., Hese S.o., Berger C. et Schmullius C. 2011**, « An object-based multisensoral approach for the derivation of urban land use structures in the city of Rostock, Germany », *SPIE Proceedings. Earth Resources and Environmental Remote Sensing/GIS Applications II*, édité par U. Michel et D.L. Civco, vol. 8181, p. 8.
- Liu A., Jun G. et Ghosh J. 2009**, « Spatially cost-sensitive active learning », *Proceedings of the SIAM International Conference on Data Mining*, pp. 814–825.
- Liu Y., Bian L., Meng Y., Wang H., Zhang S., Yang Y., Shao X. et Wang B. 2012**, « Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 68, pp. 144–156.
- Lloyd S. 1982**, « Least squares quantization in PCM », *IEEE Transactions on Information Theory*, vol. 28, n° 2, pp. 129–137.
- Long N. 2003**, *Analyses morphologiques et aérodynamiques du tissu urbain : application à la micro-climatologie de Marseille pendant la campagne ESCOMPTE*, Thèse de Doctorat, Université des Sciences et Techniques de Lille.
- Long N., Bocher E., Leduc T. et Moreau G. 2008**, « Sensitivity of spatial indicators for urban terrain characterization », *IEEE International Geoscience and Remote Sensing Symposium. IGARSS 2008*, vol. 3, pp. 1248–1251.
- Long N. et Kergomard C. 2005**, « Classification morphologique du tissu urbain pour des applications climatologiques. Cas de Marseille », *Revue internationale de géomatique*, vol. 15, n° 4, pp. 487–512.

- Loriot P.** 2008, « Détermination d'un MOS et calcul d'une tache urbaine à partir de la BD TOPO® de l'IGN », Centre d'Etudes Techniques de l'Équipement (CETE) du Sud-Ouest.
- Lu D. et Weng Q.** 2006, « Use of impervious surface in urban land-use classification », *Remote Sensing of Environment*, vol. 102, n° 1-2, pp. 146–160.
- Lucieer A. et Stein A.** 2002, « Existential uncertainty of spatial objects segmented from satellite sensor imagery », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, n° 11, pp. 2518–2521.
- MacLean M.G. et Congalton R.G.** 2013, « PolyFrag: A vector-based program for computing landscape metrics », *GIScience & Remote Sensing*, vol. 50, n° 6, pp. 591–603.
- MacQueen J.** 1967, « Some methods for classification and analysis of multivariate observations », *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, édité par L.M. Le Cam et J. Neyman, vol. 1, pp. 281–297.
- Maechler M., Rousseeuw P., Struyf A., Hubert M. et Hornik K.** 2016, « cluster: Cluster analysis basics and extensions », R package version 2.0.4.
- Mangin D.** 2004, *La ville franchisée : formes et structures de la ville contemporaine*, Éditions de la Villette, Paris, France.
- Marceau D.** 1999, « The scale issue in the social and natural sciences », *Canadian Journal of Remote Sensing*, vol. 25, n° 4, pp. 347–356.
- Marceau D.J., Gratton D.J., Fournier R.A. et Fortin J.P.** 1994, « Remote sensing and the measurement of geographical entities in a forested environment 2. The optimal spatial resolution », *Remote Sensing of Environment*, vol. 49, n° 2, pp. 105–117.
- Martha T.R., Kerle N., Van Westen C.J., Jetten V. et Kumar K.V.** 2011, « Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, n° 12, pp. 4928–4943.
- Martilli A.** 2014, « An idealized study of city structure, urban climate, energy consumption, and air quality », *Urban Climate*, vol. 10, pp. 430–446.
- Martinetz T., Berkovich S. et Schulten K.** 1993, « Neural-gas network for vector quantization and its application to time-series prediction », *IEEE Transactions on Neural Networks*, vol. 4, n° 4, pp. 558–569.
- Masson V., Marchadier C., Adolphe L., Aguejdad R., Avner P., Bonhomme M., Bretagne G., Briottet X., Bueno B., de Munck C. et et al.** 2014, « Adapting cities to climate change: A systemic modelling approach », *Urban Climate*, vol. 10, pp. 407–429.
- Mathieu R., Freeman C. et Aryal J.** 2007, « Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery », *Landscape and Urban Planning*, vol. 81, n° 3, pp. 179–192.

- Maurel N., Pellegrini P., Macon N., Cheptou P.O. et Lizet B.** 2013, « Trame verte : les pieds d'arbres à la croisée des chemins », *Trames vertes urbaines. De la recherche scientifique au projet urbain*, édité par P. Clergeau et N. Blanc, Éditions du Moniteur, Paris, France, pp. 169–187.
- McGarigal K., Cushman S.A. et Ene E.** 2012, « FRAGSTATS v4: Spatial pattern analysis program for categorical and continuous maps », Université du Massachusetts, Amherst, Massachusetts, États-unis.
- Mehdi L., Kohler M., Blond N., Clappier A. et Weber C.** 2013, « Interaction entre végétation et climat urbain », *Trames vertes urbaines. De la recherche scientifique au projet urbain*, édité par P. Clergeau et N. Blanc, Éditions du Moniteur, Paris, France, pp. 266–333.
- Mehdi L., Weber C., Pietro F.D. et Selmi W.** 2012, « Évolution de la place du végétal dans la ville, de l'espace vert à la trame verte », *VertigO – la revue électronique en sciences de l'environnement*, vol. 12, n° 2.
- Meinel G., Hecht R. et Herold H.** 2009, « Analyzing building stock using topographic maps and GIS », *Building Research & Information*, vol. 37, n° 5-6, pp. 468–482.
- Merlin P. et Choay F.** 1988, *Dictionnaire de l'urbanisme et de l'aménagement*, Presses Universitaires de France.
- Mesev T.V., Longley P.A., Batty M. et Xie Y.** 1995, « Morphology from imagery: detecting and measuring the density of urban land use », *Environment and Planning A*, vol. 27, n° 5, pp. 759–780.
- Michel J., Grizonnet M., Jaen A., Harasse S., Hermitte L., Guinet J., Malik J. et Savinaud M.** 2012, « Open tools and methods for large scale segmentation of Very High Resolution satellite images », *Open Source Geospatial Research and Education Symposium, OGRS2012*.
- Michel J., Youssefi D. et Grizonnet M.** 2015, « Stable Mean-Shift algorithm and its application to the segmentation of arbitrarily large remote sensing images », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, n° 2, pp. 952–964.
- Millennium Ecosystem Assessment.** 2005, « Ecosystems and human well-being: Current state and trends. Volume 1 », .
- Mimet A., Maurel N., Pellissier V., Simon L. et Julliard R.** 2014, « Towards a unique landscape description for multi-species studies: A model comparison with common birds in a human-dominated French region », *Ecological Indicators*, vol. 36, pp. 19–32.
- Morancho A.B.** 2003, « A hedonic valuation of urban green areas », *Landscape and Urban Planning*, vol. 66, n° 1, pp. 35–41.
- Moser B., Jaeger J.A.G., Tappeiner U., Tasser E. et Eiselt B.** 2006, « Modification of the effective mesh size for measuring landscape fragmentation to solve the boundary problem », *Landscape Ecology*, vol. 22, n° 3, pp. 447–459.

- Myint S.W., Gober P., Brazel A., Grossman-Clarke S. et Weng Q.** 2011, « Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery », *Remote Sensing of Environment*, vol. 115, n° 5, pp. 1145–1161.
- Nabucet J., Rougier S., Deniau J., Vétillard L., Hanson E., Benarchid O., Wolff E., Hubert-Moy L. et Puissant A.** 2015, « Multi-scale methodology to map grey and green structures in urban areas using Pléiades images and existing geographic data », *Revue Française de Photogrammétrie et de Télédétection*, , n° 209, pp. 95–101.
- Nagpal A., Jatain A. et Gaur D.** 2013, « Review based on data clustering algorithms », *IEEE Conference on Information Communication Technologies. ICT 2013*.
- Neubert M., Herold H. et Meinel G.** 2008, « Assessing image segmentation quality – Concepts, methods and application », *Object-Based Image Analysis*, édité par T. Blaschke, S. Lang et G.J. Hay, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, pp. 769–784.
- Nichol J. et Wong M.S.** 2007, « Remote sensing of urban vegetation life form by spectral mixture analysis of high-resolution IKONOS satellite images », *International Journal of Remote Sensing*, vol. 28, n° 5, pp. 985–1000.
- Novack T. et Kux H.J.** 2010, « Urban land cover and land use classification of an informal settlement area using the open-source knowledge-based system InterIMAGE », *Journal of Spatial Science*, vol. 55, n° 1, pp. 23–41.
- Nowak D.** 1994., « Air pollution removal by Chicago's urban forest », Rapport technique n° NE-186, United States Department of Agriculture Forest Service.
- Oke T.** 1988, « Street design and urban canopy layer climate », *Energy and Buildings*, vol. 11, n° 1-3, pp. 103–113.
- Olson C.L.** 1974, « Comparative robustness of six tests in multivariate analysis of variance », *Journal of the American Statistical Association*, vol. 69, n° 348, pp. 894–908.
- Pal N.R. et Pal S.K.** 1993, « A review on image segmentation techniques », *Pattern Recognition*, vol. 26, n° 9, pp. 1277–1294.
- Panerai P., Depaule J.C. et Demorgon M.** 1999, *Analyse urbaine*, Architecture et Urbanisme, Parenthèses, collection eupalinos, Marseille, France.
- Pasolli E., Melgani F., Tuia D., Pacifici F. et Emery W.J.** 2014, « SVM active learning approach for image classification using spatial information », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, n° 4, pp. 2217–2233.
- Persello C., Boularias A., Dalponte M., Gobakken T., Naesset E. et Scholkopf B.** 2014, « Cost-sensitive active learning with lookahead: Optimizing field surveys for remote sensing data classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, n° 10, pp. 6652–6664.
- Persello C. et Bruzzone L.** 2012, « Active learning for domain adaptation in the supervised classification of remote sensing images », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, n° 11, pp. 4468–4483.

- Persello C. et Bruzzone L.** 2014, « Active and semisupervised learning for the classification of remote sensing images », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, n° 11, pp. 6937–6956.
- Peters J., Van Coillie F., Westra T. et De Wulf R.** 2011, « Synergy of very high resolution optical and radar data for object-based olive grove mapping », *International Journal of Geographical Information Science*, vol. 25, n° 6, pp. 971–989.
- Pham T.T.H., Apparicio P., Séguin A.M. et Gagnon M.** 2011, « Mapping the greenscape and environmental equity in Montreal: An application of remote sensing and GIS », *Mapping Environmental Issues in the City : Arts and Cartography Cross Perspectives*, édité par S. Caquard, L. Vaughan et W. Cartwright, Lecture Notes in Geoinformation and Cartography, Springer-Verlag Berlin Heidelberg, pp. 30–48.
- Pham T.T.H. et He D.C.** 2008, « How do people perceive the city's green space? A view from satellite imagery (in Hanoi, Vietnam) », *IEEE International Geoscience and Remote Sensing Symposium. IGARSS 2008*, pp. 1228–1231.
- Phinn S., Stanford M., Scarth P., Murray A.T. et Shyy P.T.** 2002, « Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques », *International Journal of Remote Sensing*, vol. 23, n° 20, pp. 4131–4153.
- Plumejeaud-Perreau C., Poitevin C., Pignon-Mussaud C. et Long N.** 2015, « Building Local Climate Zones by using socio-economic and topographic vectorial databases », *9<sup>th</sup> International Conference on Urban Climate*.
- Poncet-Montanges A., Moser G., Taubenböck H., Wurm M. et Tuia D.** 2015, « Classification of urban structural types with multisource data and structured models », *Joint Urban Remote Sensing Event. JURSE 2015*, p. 4.
- Ponti M.P.** 2013, « Segmentation of low-cost remote sensing images combining vegetation indices and Mean Shift », *IEEE Geoscience and Remote Sensing Letters*, vol. 10, n° 1, pp. 67–70.
- Popescu S.C.** 2007, « Estimating biomass of individual pine trees using airborne lidar », *Biomass and Bioenergy*, vol. 31, n° 9, pp. 646–655.
- Pouyanne G.** 2004, *Forme urbaine et mobilité quotidienne*, Thèse de Doctorat, Université Montesquieu - Bordeaux IV.
- Pu R. et Landry S.** 2012, « A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species », *Remote Sensing of Environment*, vol. 124, pp. 516–533.
- Pu R., Landry S. et Yu Q.** 2011, « Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery », *International Journal of Remote Sensing*, vol. 32, n° 12, pp. 3285–3308.
- Puissant A.** 2003, *Information géographique et images à très haute résolution, utilité et application en milieu urbain*, Thèse de Doctorat, Université Louis Pasteur – Strasbourg I.

- Puissant A.** 2014, *De l'inventaire au suivi des milieux : apport des données géo-spatiales et multi-sources. Tome 1 : texte de synthèse*, Mémoire d'Habilitation à Diriger des Recherches. Université de Strasbourg.
- Puissant A., Hirsch J. et Weber C.** 2005, « The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery », *International Journal of Remote Sensing*, vol. 26, n° 4, pp. 733–745.
- Puissant A., Lachiche N., Skupinski G., Braud A., Perret J. et Mas A.** 2011, « Classification et évolution des tissus urbains à partir de données vectorielles », *Revue internationale de géomatique*, vol. 21, n° 4, pp. 513–532.
- Puissant A., Rougier S. et Stumpf A.** 2014, « Object-oriented mapping of urban trees using Random Forest classifiers », *International Journal of Applied Earth Observation and Geoinformation*, vol. 26, pp. 235–245.
- Qi J., Chehbouni A., Huete A.R., Kerr Y.H. et Sorooshian S.** 1994, « A modified soil adjusted vegetation index », *Remote Sensing of Environment*, vol. 48, n° 2, pp. 119–126.
- Quackenbush L.J., Hopkins P.F. et Kinn G.J.** 2000, « Developing forestry products from high resolution digital aerial imagery », *Photogrammetric Engineering & Remote Sensing*, vol. 66, n° 11, pp. 1337–1346.
- R Core Team.** 2014, « R: A language and environment for statistical computing », R Foundation for Statistical Computing, Vienne, Autriche.
- Rajan S., Ghosh J. et Crawford M.** 2008, « An active learning approach to hyperspectral data classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, n° 4, pp. 1231–1242.
- Räsänen A., Rusanen A., Kuitunen M. et Lensu A.** 2013, « What makes segmentation good? A case study in boreal forest habitat mapping », *International Journal of Remote Sensing*, vol. 34, n° 23, pp. 8603–8627.
- Raynaud D.** 1999, « Forme urbaine : une notion exemplaire du point de vue de l'épistémologie des sciences sociales », *Langages singuliers et partagés de l'urbain (Actes du Colloque LOUEST)*, édité par P. Boudon, L'Harmattan, pp. 93–120.
- Reis I.C.d., Honório N.A., Barros E.S.M.d., Barcellos C., Kitron U., Camara D.C.P., Pereira G.R., Keppeler E.C., da Silva-Nunes M. et Codeço C.T.** 2015, « Epidemic and endemic malaria transmission related to fish farming ponds in the Amazon frontier », *PLoS ONE*, vol. 10, n° 9, p. 15.
- Reitberger J., Krzystek P. et Stilla U.** 2008, « Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees », *International Journal of Remote Sensing*, vol. 29, n° 5, pp. 1407–1431.
- Rempel R., Kaukinen D. et Carr A.** 2012, « Patch analyst and patch grid », Centre for Northern Forest Ecosystem Research. Ontario Ministry of Natural Resources, Thunder Bay, Ontario, Canada.

- Rhein C. et Palibrk M.** 2014, « Formes urbaines, modes d'occupation du sol et composition sociale en zone urbaine dense : des relations inattendues à Paris », *Cybergeo : European Journal of Geography*.
- Riboulot-Chetrit M.** 2015, « Les jardins privés : de nouveaux espaces clés pour la gestion de la biodiversité dans les agglomérations? », *Articulo – Journal of Urban Research*, , n° 6, Special issue : La fabrique des espaces ouverts.
- Ridd M.K.** 1995, « Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities », *International Journal of Remote Sensing*, vol. 16, n° 12, pp. 2165–2185.
- Riitters K.H., O'Neill R.V., Hunsaker C.T., Wickham J.D., Yankee D.H., Timmins S.P., Jones K.B. et Jackson B.L.** 1995, « A factor analysis of landscape pattern and structure metrics », *Landscape Ecology*, vol. 10, n° 1, pp. 23–39.
- Robert S. et Autran J.** 2012, « Décrire à grande échelle l'occupation des sols urbains par photo-interprétation. Réflexion méthodologique et expérimentation en Provence », *Sud-Ouest Européen*, , n° 33, pp. 25–40.
- Robinson D. et Stone A.** 2004, « Solar radiation modelling in the urban context », *Solar Energy*, vol. 77, n° 3, pp. 295–309.
- Rodriguez-Galiano V.F., Ghimire B., Rogan J., Chica-Olmo M. et Rigol-Sánchez J.P.** 2012, « An assessment of the effectiveness of a random forest classifier for land-cover classification », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104.
- Rougier S., Nabucet J., Thomas A., Hanson E., Wolff E., Hubert-Moy L. et Puissant A.** 2014, « Grey and green infrastructure in urban areas based on Pléiades image analysis combined with existing vector database », *Pléiades Days*.
- Rougier S. et Puissant A.** 2013, « Analyse, caractérisation et suivi des espaces urbains et péri-urbains à partir de données multi-sources », *Colloque biennal des Zones Ateliers*.
- Rougier S. et Puissant A.** 2014, « Improvements of urban vegetation segmentation and classification using multi-temporal Pléiades images », *Proceedings of the 5<sup>th</sup> Geographic Object-Based Image Analysis Conference. GEOBIA 2014*, vol. 3, pp. 409–414.
- Rougier S., Puissant A., Stumpf A. et Lachiche N.** 2016, « Comparison of sampling strategies for object-based classification of urban vegetation from Very High Resolution satellite images », *International Journal of Applied Earth Observation and Geoinformation*, vol. 51, pp. 60–73.
- Rousseaux F., Long N. et Renouard A.** 2012, « Vers une simulation de l'évolution des structures urbaines à partir d'une modélisation multi-agents », *VertigO – la revue électronique en sciences de l'environnement*, vol. 11, n° 3.
- Rousseeuw P.J.** 1987, « Silhouettes: A graphical aid to the interpretation and validation of cluster analysis », *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65.



- Rousset P., Guinot C. et Vero J. 2005**, « Utilisation des cartes d'auto-organisation dans la classification, illustration par des exemples », *Bulletin de méthodologie sociologique*, vol. 87, pp. 5–37.
- Rudnick D., Beier P. and Cushman S., Dieffenbach E., Epps C., Gerber L., Hartter J., Jenness J., Kintsch J. et Merenlender A.a.a. 2012**, « The role of landscape connectivity in planning and implementing conservation and restoration priorities », Rapport technique n° 16, Ecological Society of America, Washington, District de Columbia, États-Unis.
- Sahraoui Y., Lunardi N. et Antoni J.P. 2014**, « Essai de typologie des espaces résidentiels à partir d'indicateurs désagrégés à l'échelle du bâtiment. Application à Besançon et à Metz », *Cybergeo : European Journal of Geography*.
- Sakhy A., Madelin M. et Beltrando G. 2011**, « Les échelles d'étude de l'îlot de chaleur urbain et ses relations avec la végétation et la géométrie de la ville (exemple de Paris) », *Dixièmes Rencontres de Théo Quant*.
- Salat S. 2011**, *La ville et les formes. Sur l'urbanisme durable*, Hermann & CSTB.
- SAS Institute Inc. 2015**, « SAS/STAT® 14.1 User's guide - The VARCLUS procedure », Cary, Caroline du Nord, États-Unis.
- Schneider A. et Woodcock C.E. 2008**, « Compact, dispersed, fragmented, extensive? A comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information », *Urban Studies*, vol. 45, n° 3, pp. 659–692.
- Schöpfer E., Lang S. et Blaschke T. 2005**, « A "Green Index" incorporating remote sensing and citizen's perception of green space », *International Archives of Photogrammetry, Remote Sensing and spatial Information Sciences. Joint Symposia URBAN - URS 2005*, vol. XXXVII-5/W1, p. 6.
- Schroeder T.A., Cohen W.B., Song C., Canty M.J. et Yang Z. 2006**, « Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon », *Remote Sensing of Environment*, vol. 103, n° 1, pp. 16–26.
- Schwanen T., Dieleman F.M. et Dijst M. 2004**, « The impact of metropolitan structure on commute behavior in the Netherlands: A multilevel approach », *Growth and Change*, vol. 35, n° 3, pp. 304–333.
- Secord J. et Zakhor A. 2007**, « Tree detection in urban regions using aerial Lidar and image data », *IEEE Geoscience and Remote Sensing Letters*, vol. 4, n° 2, pp. 196–200.
- Sellin V., Magnanon S., Gourmelon F., Debaine F. et Nabucet J. 2015**, « Étude expérimentale en cartographie de la végétation par télédétection », *Cybergeo : European Journal of Geography*.
- Selmi W., Weber C. et Mehdi L. 2013**, « Multifonctionnalité des espaces végétalisés urbains », *VertigO – la revue électronique en sciences de l'environnement*, vol. 13, n° 2.

- Selmi W., Weber C., Rivière E., Blond N., Mehdi L. et Nowak D. 2016**, « Air pollution removal by trees in public green spaces in Strasbourg city, France », *Urban Forestry & Urban Greening*, vol. 17, pp. 192–201.
- Setiawan H., Mathieu R. et Thompson-Fawcett M. 2006**, « Assessing the applicability of the V-I-S model to map urban land use in the developing world: Case study of Yogyakarta, Indonesia », *Computers, Environment and Urban Systems*, vol. 30, n° 4, pp. 503–522.
- Settles B. 2010**, « Active learning literature survey », Rapport technique n° 1648, University of Wisconsin.
- Settles B., Craven M. et Friedland L. 2008**, « Active learning with real annotation costs », *Proceedings of the NIPS workshop on cost-sensitive learning*, p. 10.
- Seung H.S., Opper M. et Sompolinsky H. 1992**, « Query by committee », *Proceedings of the Fifth Annual Workshop on Computational Learning Theory. COLT '92*, pp. 287–294.
- Sheeren D., Bastin N., Ouin A., Ladet S., Balent G. et Lacombe J.P. 2009**, « Discriminating small wooded elements in rural landscape from aerial photography: a hybrid pixel/object-based analysis approach », *International Journal of Remote Sensing*, vol. 30, n° 19, pp. 4979–4990.
- Smith A. 2010**, « Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm », *Journal of Spatial Science*, vol. 55, n° 1, pp. 69–79.
- Smith D.A. et Crooks A.T. 2010**, « From buildings to cities: techniques for the multi-scale analysis of urban form and function », Centre for Advanced Spatial Analysis, University College London : Working Paper 155.
- Sordello R., Rogeon G. et Touroult J. 2014**, « La fonctionnalité des continuités écologiques – Premiers éléments de compréhension », Service du patrimoine naturel. Muséum national d'histoire naturelle, Paris, France.
- Steiniger S., Lange T., Burghardt D. et Weibel R. 2008**, « An approach for the classification of urban building structures based on discriminant analysis techniques », *Transactions in GIS*, vol. 12, n° 1, pp. 31–59.
- Stewart I.D. et Oke T.R. 2012**, « Local climate zones for urban temperature studies », *Bulletin of the American Meteorological Society*, vol. 93, n° 12, pp. 1879–1900.
- Stow D., Lopez A., Lippitt C., Hinton S. et Weeks J. 2007**, « Object-based classification of residential land use within Accra, Ghana based on QuickBird satellite data », *International Journal of Remote Sensing*, vol. 28, n° 22, pp. 5167–5173.
- Strobl C., Boulesteix A.L., Kneib T., Augustin T. et Zeileis A. 2008**, « Conditional variable importance for Random Forests », *BMC Bioinformatics*, vol. 9, n° 1, p. 11.
- Stumpf A. et Kerle N. 2011**, « Object-oriented mapping of landslides using Random Forests », *Remote Sensing of Environment*, vol. 115, n° 10, pp. 2564–2577.

- Stumpf A., Lachiche N., Malet J.P., Kerle N. et Puissant A. 2014**, « Active learning in the spatial domain for remote sensing image classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, n° 5, pp. 2492–2507.
- Sugar C.A. et James G.M. 2003**, « Finding the number of clusters in a dataset », *Journal of the American Statistical Association*, vol. 98, n° 463, pp. 750–763.
- Tan P.N., Steinbach M. et Kumar V. 2005**, *Introduction to data mining*, Addison-Wesley Longman Publishing Co., Inc., Boston, Massachusetts, États-Unis.
- Tan Q. et Wang J. 2009**, « Mapping urban tree coverage using object-oriented image analysis method: A case study », *2009 IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, pp. 342–345.
- Tansey K., Chambers I., Anstee A., Denniss A. et Lamb A. 2009**, « Object-oriented classification of very high resolution airborne imagery for the extraction of hedgerows and field margin cover in agricultural areas », *Applied Geography*, vol. 29, n° 2, pp. 145–157.
- Tathiri A., Shafri H.Z.M. et Hamedianfar A. 2014**, « Development of transferable rule-sets for urban areas using QuickBird satellite imagery », *IEEE International Conference on Aerospace Electronics and Remote Sensing Technology. ICARES 2014*, pp. 229– 233.
- Taubenböck H., Wiesner M., Felbier A., Marconcini M., Esch T. et Dech S. 2014**, « New dimensions of urban landscapes: The spatio-temporal evolution from a polynuclei area to a mega-region based on remote sensing data », *Applied Geography*, vol. 47, pp. 137–153.
- Taylor M.S., Wheeler B.W., White M.P., Economou T. et Osborne N.J. 2015**, « Research note: Urban street tree density and antidepressant prescription rates: A cross-sectional study in London, UK », *Landscape and Urban Planning*, vol. 136, pp. 174–179.
- Tibshirani R., Walther G. et Hastie T. 2001**, « Estimating the number of clusters in a data set via the gap statistic », *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, vol. 63, n° 2, pp. 411–423.
- Tigges J., Lakes T. et Hostert P. 2013**, « Urban vegetation classification: Benefits of multitemporal RapidEye satellite data », *Remote Sensing of Environment*, vol. 136, pp. 66–75.
- Tong X., Li X., Xu X., Xie H., Feng T., Sun T., Jin Y. et Liu X. 2014**, « A two-phase classification of urban vegetation using airborne Lidar data and aerial photography », *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, n° 10, pp. 4153–4166.
- Tooke T.R., Coops N.C., Goodwin N.R. et Voogt J.A. 2009**, « Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications », *Remote Sensing of Environment*, vol. 113, n° 2, pp. 398–407.
- Toublanc M. et Bonin S. 2012**, « Planifier les trames vertes dans les aires urbaines : une alliance à trouver entre paysagisme et écologie », *Développement durable et territoires*, vol. 3, n° 2.

- Traisnel J.P.** 2001, « Habitat et développement durable, bilan rétrospectif et prospectif », *Les Cahiers du CLIP*, vol. 13, pp. 5–72.
- Tran D.B., Piombini A., Ignatowicz M., Moreno D., Frigui R., Aguiléra A. et Badariotti D.** 2014, « Morphologie urbaine et mobilité dans la Communauté Urbaine de Strasbourg », *Cybergeo : European Journal of Geography*.
- Tran T.D.B., Puissant A., Badariotti D. et Weber C.** 2011, « Optimizing spatial resolution of imagery for urban form detection – The cases of France and Vietnam », *Remote Sensing*, vol. 3, n° 12, pp. 2128–2147.
- Troya-Galvis A., Gancarski P., Passat N. et Berti-Equille L.** 2015, « Unsupervised quantification of under- and over-segmentation for object-based remote sensing image analysis », *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, n° 5, pp. 1936–1945.
- Tsai Y.H.** 2005, « Quantifying urban form: compactness versus sprawl », *Urban Studies*, vol. 42, n° 1, pp. 141–161.
- Tuia D. et Munoz-Mari J.** 2013, « Learning user’s confidence for active learning », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, n° 2, pp. 872–880.
- Tuia D., Pacifici F., Kanevski M. et Emery W.** 2009a, « Classification of very high spatial resolution imagery using mathematical morphology and support vector machines », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, n° 11, pp. 3866–3879.
- Tuia D., Pasolli E. et Emery W.** 2011a, « Using active learning to adapt remote sensing image classifiers », *Remote Sensing of Environment*, vol. 115, n° 9, pp. 2232–2242.
- Tuia D., Ratle F., Pacifici F., Kanevski M. et Emery W.** 2009b, « Active learning methods for remote sensing image classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, n° 7, pp. 2218–2232.
- Tuia D., Volpi M., Copa L., Kanevski M. et Munoz-Mari J.** 2011b, « A survey of active learning algorithms for supervised remote sensing image classification », *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, n° 3, pp. 606–617.
- Tzoulas K., Korpela K., Venn S., Yli-Pelkonen V., Kaźmierczak A., Niemela J. et James P.** 2007, « Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review », *Landscape and Urban Planning*, vol. 81, n° 3, pp. 167–178.
- Uuemaa E., Mander U. et Marja R.** 2013, « Trends in the use of landscape spatial metrics as landscape indicators: A review », *Ecological Indicators*, vol. 28, pp. 100–106.
- Vacquié L. et Houet T.** 2012, « Cartographie des zones humides de montagne par télédétection. Potentialités des images à très haute résolution spatiale », *Revue internationale de géomatique*, vol. 22, n° 4, pp. 497–518.
- Van Coillie F.M., Gardin S., Anseel F., Duyck W., Verbeke L.P. et De Wulf R.R.** 2014, « Variability of operator performance in remote-sensing image interpretation: The

- importance of human and external factors », *International Journal of Remote Sensing*, vol. 35, n° 2, pp. 754–778.
- Van Delm A. et Gulinck H. 2011**, « Classification and quantification of green in the expanding urban and semi-urban complex: Application of detailed field data and IKONOS-imagery », *Ecological Indicators*, vol. 11, n° 1, pp. 52–60.
- Vanderhaegen S. et Canters F. 2010**, « Developing urban metrics to describe the morphology of urban areas at block level », *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, n° 4/C7, p. 6.
- Vannier C. et Hubert-Moy L. 2014**, « Multiscale comparison of remote-sensing data for linear woody vegetation mapping », *International Journal of Remote Sensing*, vol. 35, n° 21, pp. 7376–7399.
- Vaudour E., Noirot-Cosson P. et Membrive O. 2015**, « Early-season mapping of crops and cultural operations using very high spatial resolution Pléiades images », *International Journal of Applied Earth Observation and Geoinformation*, vol. 42, pp. 128–141.
- Vesanto J., Himberg J., Alhoniemi E. et Parhankangas J. 1999**, « Self-organizing map in Matlab: the SOM toolbox », *Proceedings of the Matlab DSP Conference*, pp. 35–40.
- Vigneau E., Chen M. et Qannari E.M. 2015**, « ClustVarLV: An R package for the clustering of variables around latent variables », *The R Journal*, vol. 7, n° 2, pp. 134–148.
- Vigneau E. et Qannari E.M. 2003**, « Clustering of variables around latent components », *Communications in Statistics - Simulation and Computation*, vol. 32, n° 4, pp. 1131–1150.
- Volpi M., Tuia D. et Kanevski M. 2012**, « Memory-based cluster sampling for remote sensing image classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, n° 8, pp. 3096–3106.
- Van de Voorde T., Jacquet W. et Canters F. 2011**, « Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data », *Landscape and Urban Planning*, vol. 102, n° 3, pp. 143–155.
- Vétillard L. 2013**, *Cartographie des corridors écologiques par télédétection THRS. Évaluation d'une image Pléiades sur la ville de Rennes*, Mémoire de Master, Université de Rennes.
- Walde I., Hese S., Berger C. et Schmullius C. 2014**, « From land cover-graphs to urban structure types », *International Journal of Geographical Information Science*, vol. 28, n° 3, pp. 584–609.
- Walker J.S. et Briggs J.M. 2007**, « An object-oriented approach to urban forest mapping in Phoenix », *Photogrammetric Engineering & Remote Sensing*, vol. 73, n° 5, pp. 577–583.

- Wan L., Tang K., Li M., Zhong Y. et Qin A.K. 2015**, « Collaborative active and semisupervised learning for hyperspectral remote sensing image classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, n° 5, pp. 2384–2396.
- Wang H.F., Qiu J.X., Breuste J., Ross Friedman C., Zhou W.Q. et Wang X.K. 2013**, « Variations of urban greenness across urban structural units in Beijing, China », *Urban Forestry & Urban Greening*, vol. 12, n° 4, pp. 554–561.
- Wania A., Bruse M., Blond N. et Weber C. 2012**, « Analysing the influence of different street vegetation on traffic-induced particle dispersion using microscale simulations », *Journal of Environmental Management*, vol. 94, n° 1, pp. 91–101.
- Wania A. et Weber C. 2007**, « Hyperspectral imagery and urban green observation », *2007 Urban Remote Sensing Joint Event*, p. 8.
- Weber C., Petropoulou C. et Hirsch J. 2005**, « Urban development in the Athens metropolitan area using remote sensing data with supervised analysis and GIS », *International Journal of Remote Sensing*, vol. 26, n° 4, pp. 785–796.
- Weber N., Haase D. et Franck U. 2014**, « Assessing modelled outdoor traffic-induced noise and air pollution around urban structures using the concept of landscape metrics », *Landscape and Urban Planning*, vol. 125, pp. 105–116.
- Weeks J.R. 2001**, « The role of spatial analysis in demographic research », *Spatially integrated social science : examples in best practice*, édité par M. Goodchild et D. Janelle, Oxford University Press, New York, États-Unis, pp. 381–399.
- Wemmert C., Puissant A., Forestier G. et Gañçarski P. 2009**, « Multiresolution remote sensing image clustering », *IEEE Geoscience and Remote Sensing Letters*, vol. 6, n° 3, pp. 533–537.
- Witharana C. et Civco D.L. 2014**, « Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2 (ED2) », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 108–121.
- Woodcock C.E. et Strahler A.H. 1987**, « The factor of scale in remote sensing », *Remote Sensing of Environment*, vol. 21, n° 3, pp. 311–332.
- Wu J. 2004**, « Effects of changing scale on landscape pattern analysis: scaling relations », *Landscape Ecology*, vol. 19, n° 2, pp. 125–138.
- Wu S., Silván-Cárdenas J. et Wang L. 2007**, « Per-field urban land use classification based on tax parcel boundaries », *International Journal of Remote Sensing*, vol. 28, n° 12, pp. 2777–2801.
- Wu S.s., Xu B. et Wang L. 2006**, « Urban land-use classification using variogram-based analysis with an aerial photograph », *Photogrammetric Engineering & Remote Sensing*, vol. 72, n° 7, pp. 813–822.

- Wulder M.A. et Boots B. 1998**, « Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic », *International Journal of Remote Sensing*, vol. 19, n° 11, pp. 2223–2231.
- Wurm M., Taubenböck H. et Dech S. 2010**, « Quantification of urban structure on building block level utilizing multisensoral remote sensing data », *SPIE Proceedings – Earth Resources and Environmental Remote Sensing/GIS Applications*, édité par U. Michel et D.L. Civco, vol. 7831, p. 12.
- Xie Y., Sha Z. et Yu M. 2008**, « Remote sensing imagery in vegetation mapping: a review », *Journal of Plant Ecology*, vol. 1, n° 1, pp. 9–23.
- Xu J., Hang R. et Liu Q. 2014**, « Patch-Based Active Learning (PTAL) for spectral-spatial classification on hyperspectral data », *International Journal of Remote Sensing*, vol. 35, n° 5, pp. 1846–1875.
- Yang J., He Y. et Weng Q. 2015**, « An automated method to parameterize segmentation scale by enhancing intrasegment homogeneity and intersegment heterogeneity », *IEEE Geoscience and Remote Sensing Letters*, vol. 12, n° 6, pp. 1282–1286.
- Youjing Z. et Hengtong F. 2007**, « Identification scales for urban vegetation classification using high spatial resolution satellite data », *IEEE International Geoscience and Remote Sensing Symposium. IGARSS 2007*, pp. 1472–1475.
- Zhan Q., Molenaar M. et Gorte B. 2000**, « Urban land use classes with fuzzy membership and classification based on integration of remote sensing and GIS », *International Archives of Photogrammetry and Remote Sensing*, vol. 33, Part B7, pp. 1751–1759.
- Zhan Q., Molenaar M., Tempfli K. et Shi W. 2005**, « Quality assessment for geo-spatial objects derived from remotely sensed data », *International Journal of Remote Sensing*, vol. 26, n° 14, pp. 2953–2974.
- Zhang B.T. et Veenker G. 1991**, « Neural networks that teach themselves through genetic discovery of novel examples », *IEEE International Joint Conference on Neural Networks*, vol. 1, pp. 690–695.
- Zhang C. et Xie Z. 2012**, « Combining object-based texture measures with a neural network for vegetation mapping in the Everglades from hyperspectral imagery », *Remote Sensing of Environment*, vol. 124, pp. 310–320.
- Zhang H., Fritts J.E. et Goldman S.A. 2008**, « Image segmentation evaluation: A survey of unsupervised methods », *Computer Vision and Image Understanding*, vol. 110, n° 2, pp. 260–280.
- Zhang X., Feng X. et Jiang H. 2010**, « Object-oriented method for urban vegetation mapping using IKONOS imagery », *International Journal of Remote Sensing*, vol. 31, n° 1, pp. 177–196.
- Zhang X., Xiao P. et Feng X. 2012**, « An unsupervised evaluation method for remotely sensed imagery segmentation », *IEEE Geoscience and Remote Sensing Letters*, vol. 9, n° 2, pp. 156–160.

- Zhang Y.** 1996, « A survey on evaluation methods for image segmentation », *Pattern Recognition*, vol. 29, n° 8, pp. 1335–1346.
- Zhang Y., Yang H.L., Prasad S., Pasolli E., Jung J. et Crawford M.** 2015, « Ensemble multiple kernel active learning for classification of multisource remote sensing data », *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, n° 2, pp. 845–858.
- Zhang Y.J.** 1995, « Influence of segmentation over feature measurement », *Pattern Recognition Letters*, vol. 16, n° 2, pp. 201–206.



## Liste des figures

1.1 Trames et structure urbaines . . . . .	12
1.2 Les trois éléments constitutifs des tissus urbains : les voies, le découpage foncier et les bâtiments . . . . .	14
1.3 Exemple de tissus urbains caractéristiques . . . . .	17
1.4 Évolution de la consommation finale énergétique en France par secteur en millions de tonnes équivalents pétrole . . . . .	18
1.5 Émissions de gaz à effet de serre en France en millions de tonnes équivalents CO <sub>2</sub> . . . . .	18
1.6 Exemple de morphologie simplifiée pour l'analyse de la consommation des bâtiments . . . . .	21
1.7 Les <i>Local Climate Zones</i> proposées par Stewart et Oke (2012) . . . . .	22
1.8 La <i>Spacemate</i> de Berghauser Pont et Haupt (2005) . . . . .	23
1.9 Typologie des tissus urbains proposée par Meinel <i>et al.</i> (2009) . . . . .	24
1.10 Exemple d'éléments de la Trame verte et bleue : réservoirs de biodiversité et types de corridors . . . . .	28
1.11 Exemple de Trames verte et bleue composées de sous-trames écologiques spécifiques . . . . .	32
2.1 Chronologie des capteurs satellites optiques et thermiques multi et hyperspectraux . . . . .	41
2.2 Les échelles d'analyse des formes urbaines . . . . .	43
2.3 Présentation du site de l'Eurométropole de Strasbourg . . . . .	44
2.4 Présentation du site de Rennes Métropole . . . . .	46
2.5 Extrait de la base de données Corine Land Cover sur l'Eurométropole de Strasbourg . . . . .	47
2.6 Extrait de la base de données Urban Atlas sur l'Eurométropole de Strasbourg . . . . .	48
2.7 Extrait de la BDOCS sur l'Eurométropole de Strasbourg . . . . .	50
2.8 Extrait de la BD TOPO® du centre-ville de Rennes . . . . .	51
2.9 Les étapes d'une démarche d'analyse d'image orientée-objet . . . . .	57
2.10 Évolution du nombre de publications traitant de la cartographie de la végétation urbaine . . . . .	62
A Typologie de la trame grise à l'échelle des tissus urbains proposée . . . . .	72

B	Typologie de la trame verte proposée . . . . .	73
C	Structuration de la démarche méthodologique de la thèse . . . . .	75
D	Processus d'analyse d'image orientée-objet par classification supervisée et structuration de la démarche méthodologique . . . . .	80
3.1	Exemple d'une image sur-segmentée et sous-segmentée . . . . .	85
3.2	Emprises des images Pléiades des sites de Strasbourg et de Rennes . . . . .	93
3.3	Localisation des zones tests dans les agglomérations de Rennes et de Strasbourg et extraits des images Pléiades associées . . . . .	94
3.4	Stratification des sites de Rennes et de Strasbourg et localisation des mailles d'entraînement et de test . . . . .	97
3.5	Distribution pour Multi-Resolution et Mean Shift de l'aire des segments sur le site d'étude de Rennes . . . . .	98
3.6	Exemple de segmentations réalisées pour Strasbourg à deux échelles avec Multi-Resolution et Mean Shift pour quatre types d'utilisation du sol . . .	100
3.7	Sélection de la segmentation optimale par l'ESP2 pour l'ensemble des stratégies . . . . .	101
3.8	Résultats de NSR et de PSE pour les séries de segmentation réalisées pour Strasbourg avec l'image d'août . . . . .	102
3.9	Sélection de la segmentation optimale par l'ED2 pour l'ensemble des stratégies . . . . .	103
3.10	Sélection de la segmentation optimale par classification avec RF pour l'ensemble des stratégies . . . . .	105
4.1	Effet de la normalisation de l'entropie sur les trois stratégies d'apprentis- sage actif . . . . .	124
4.2	Performances de MCSAL <sub>D</sub> selon le nombre de fenêtres candidates . . . . .	125
4.3	Performances de MCSAL <sub>D</sub> selon le seuil minimum d'importance des variables . . . . .	126
4.4	Comparaison des performances des méthodes d'apprentissage actif avec SRGS selon la taille des régions en nombre de fenêtres sélectionnée . . . . .	128
4.5	Comparaison des performances des méthodes d'apprentissage actif avec SRGS selon la taille des régions en temps d'exécution . . . . .	129
4.6	Précisions des classifications pour MCSAL <sub>D</sub> et MCSAL <sub>S</sub> selon le jeu de données utilisé . . . . .	131
4.7	F-mesures par classes et par strate d'urbanisation obtenues avec MCSAL <sub>S</sub> . . . . .	133
4.8	Exemple de résultats de classification réalisée avec MCSAL <sub>S</sub> par strate d'urbanisation pour le site de Strasbourg . . . . .	134
4.9	Évaluation des résultats selon l'aire des segments et la pureté de la classe attribuée à un segment . . . . .	134
5.1	Localisation des extraits des images Pléiades considérés pour les sites de Rennes et de Strasbourg . . . . .	139

5.2 Répartition par classe des segments des jeux de données d'entraînement de Rennes et de Strasbourg . . . . .	141
5.3 Résultats de la sélection des paramètres par validation croisée des quatre algorithmes de classification testés . . . . .	143
5.4 Occupation du sol de l'Eurométropole de Strasbourg et des mailles de validation . . . . .	144
5.5 Répartition par classe des segments correctement classés, des faux positifs et négatifs selon la moyenne de leur brillance, du MSAVI et de l'aire pour le site de Rennes . . . . .	147
5.6 Répartition par classe des segments correctement classés, des faux positifs et négatifs selon la moyenne de la brillance, du MSAVI et de l'aire pour le site de Strasbourg . . . . .	147
5.7 Classification post-traitée de la végétation arborée et herbacée de Rennes	149
5.8 Classification post-traitée de la végétation arborée et herbacée de Strasbourg	150
5.9 Précisions de la classification de Rennes réalisée avec RF avant et après post-traitements . . . . .	151
5.10 Précisions de la classification de Strasbourg réalisée avec SVM avant et après post-traitements . . . . .	151
5.12 Occupation du sol des sites de Strasbourg issue de la base de données de la région Alsace et de Rennes issue de la photo-interprétation des îlots . .	154
5.11 Évaluation par type d'occupation du sol issue de la base de données de la région Alsace de la classification de Strasbourg . . . . .	155
5.13 Évaluation par type d'occupation du sol issue de la photo-interprétation des îlots de la classification de Rennes . . . . .	155
5.14 Distribution pour Rennes des objets détectés, des faux négatifs et positifs selon l'aire des objets . . . . .	158
5.15 Distribution pour Strasbourg des objets détectés, des faux négatifs et positifs selon l'aire des objets . . . . .	158
5.16 Distribution de l'erreur totale pour les objets détectés et des fréquences de sur et de sous-identification par classe de végétation . . . . .	159
5.17 Classification post-traitée du site d'étude de Rennes et exemples de quatre mailles de validation . . . . .	160
5.18 Classification post-traitée du site d'étude de Strasbourg et exemples de quatre mailles de validation . . . . .	161
E Structuration de la démarche méthodologique suivie dans la partie I . . .	170
6.1 Exemples de bâtiments du centre-ville de Rennes issus de la BD TOPO® et de Strasbourg issus de la BD TOPO® et du cadastre . . . . .	173
6.2 Cartographie des éléments constitutifs des trames grise, verte et bleue de l'agglomération de Strasbourg . . . . .	177
6.3 Cartographie des éléments constitutifs des trames grise, verte et bleue de l'agglomération de Rennes . . . . .	178

6.4	Exemple du diagramme de Voronoï obtenu en espace urbain dense ou pavillonnaire pour Strasbourg . . . . .	191
6.5	Exemple d'îlots construits à partir des parcelles cadastrales et à partir des réseaux de communication . . . . .	191
6.6	Exemple de découpage en îlot obtenu en espace urbain dense ou périurbain pour Rennes . . . . .	192
6.7	Découpage construit pour l'analyse de la structure urbaine de Strasbourg	200
6.8	Découpage construit pour l'analyse de la structure urbaine de Rennes . .	201
7.1	Processus d'extraction de connaissances dans des bases de données . . . .	205
7.2	Exemple de critères d'agrégation d'une classification ascendante hiérarchique . . . . .	210
7.3	Principe de l'algorithme des k-moyennes . . . . .	211
7.4	Principe de l'algorithme SOM . . . . .	213
7.5	Dendrogramme issu de la classification des 120 indicateurs . . . . .	217
7.6	Corrélations entre les variables synthétiques 1 à 7 et les indicateurs associés	220
7.7	Cartographie des variables synthétiques 1 à 7 pour l'Eurométropole de Strasbourg . . . . .	221
7.8	Corrélations entre les variables synthétiques 8 à 14 et les indicateurs associés . . . . .	223
7.9	Cartographie des variables synthétiques 8 à 14 pour l'Eurométropole de Strasbourg . . . . .	224
7.10	Estimation du nombre de classes pour les espaces urbains et non urbains par la statistique <i>gap</i> . . . . .	227
7.11	Cartographie du premier niveau de classification réalisé avec GNG . . . .	229
7.12	Cartographie du deuxième niveau de classification et réseaux de neurones de GNG pour les espaces non urbains et urbains . . . . .	234
7.13	Vérité terrain issue de la base de données occupation du sol pour l'évaluation de la classification de Strasbourg . . . . .	238
7.14	Exemple de la difficulté d'évaluation de la morphologie urbaine par la base de données d'occupation du sol de la région Alsace : l'hôpital civil de Strasbourg . . . . .	241
7.15	Vérité terrain issue de la base Urban Atlas pour l'évaluation de la classification de Strasbourg . . . . .	243
7.16	Premier niveau de classification en 8 types du site de Rennes à partir des paramètres de Strasbourg . . . . .	245
7.17	Deuxième niveau de classification en 42 types du site de Rennes à partir des paramètres de Strasbourg . . . . .	246
7.18	Vérité terrain issue de la base Urban Atlas pour l'évaluation de la classification de Rennes . . . . .	247
7.19	Vérité terrain issue de la base îlot pour l'évaluation de la classification de Rennes . . . . .	249

## Liste des tableaux

1.1	Typologie proposée par Puissant <i>et al.</i> (2011) pour la cartographie des îlots au 1/10 000 <sup>e</sup> . . . . .	23
1.2	Typologie proposée par Sellin <i>et al.</i> (2015) pour la cartographie des formations végétales par télédétection . . . . .	33
2.1	Caractéristiques des bases de données CLC, Urban Atlas et BDOCS . . . . .	49
3.1	Vue d'ensemble des attributs calculés pour chaque segment . . . . .	96
3.2	Vue d'ensemble des paramètres de segmentation utilisés pour Strasbourg et Rennes . . . . .	98
3.3	Sélection de la segmentation optimale par série et par classe de végétation avec l'ED2 . . . . .	103
3.4	F-mesures obtenues pour les segmentations sélectionnées avec l'ESP2, l'ED2 et RF pour le site de Strasbourg . . . . .	105
3.5	F-mesures obtenues pour les segmentations sélectionnées avec l'ESP2, l'ED2 et RF pour le site de Rennes . . . . .	106
4.1	Vue d'ensemble des attributs calculés à partir des informations de hauteur et du deuxième niveau de segmentation . . . . .	121
4.2	Vue d'ensemble des différents jeux de données testés pour le site de Strasbourg . . . . .	130
4.3	Moyenne des F-mesures et du nombre d'exemples sélectionnés par classes obtenues avec MCSAL <sub>S</sub> par validation croisée en 16 plis . . . . .	132
4.4	Moyenne des F-mesures et du nombre d'exemples sélectionnés par classes obtenues avec MCSAL <sub>S</sub> . . . . .	132
5.1	F-mesures par classes en segments et en aire pour chaque algorithme testé avec ses paramètres optimaux pour le site de Rennes . . . . .	145
5.2	F-mesures par classes en segments et en aire pour chaque algorithme testé avec ses paramètres optimaux pour le site de Strasbourg . . . . .	145
5.3	Temps d'exécution en minutes pour l'entraînement et pour la prédiction de chaque algorithme testé . . . . .	146
5.4	F-mesures par strate d'artificialisation pour Strasbourg et Rennes à partir des classifications post-traitée . . . . .	152

6.1 Bases de données utilisées pour la cartographie des éléments gris . . . . .	174
6.2 Bases de données utilisées pour la cartographie des éléments verts et bleus	175
6.3 Vue d'ensemble des attributs calculés pour décrire les différentes mailles	185
6.4 Unité d'analyse testée pour la cartographie de la structure urbaine et nombre de mailles pour Strasbourg et pour Rennes . . . . .	192
6.5 Part d'éléments fragmentés par classe d'occupation du sol et rang moyen par unité d'analyse proposée . . . . .	196
6.6 Rang moyen par critère d'évaluation et par unité d'analyse . . . . .	198
7.1 Récapitulatif des méthodes de clustering . . . . .	214
7.2 Interprétation des variables synthétiques issues du clustering de variables	226
7.3 Nombre de classes sélectionnées par la statistique <i>gap</i> par niveau pour les espaces urbains et non urbains . . . . .	227
7.4 Valeur moyenne de chaque type du niveau 1 de classification des mailles de Strasbourg . . . . .	230
7.5 Valeur moyenne de chaque type du niveau 2 de classification des mailles de Strasbourg . . . . .	233
7.6 Évaluation des groupes issus de la classification de Strasbourg par la BD OCS . . . . .	239
7.7 Évaluation des groupes issus de la classification de Strasbourg avec la base Urban Atlas . . . . .	243
7.8 Évaluation des groupes issus de la classification de Rennes avec la base Urban Atlas . . . . .	248
7.9 Évaluation des groupes issus de la classification de Rennes avec la base îlots photo-interprétée . . . . .	249

# Annexes

## Table des annexes

---

<b>ANNEXE A</b> : Les trames vertes et bleues dans les SRCE adoptés en France métropolitaine.....	303
<b>ANNEXE B</b> : Nomenclatures des bases de données d'occupation du sol avec une représentation continue de l'espace .....	307
<b>ANNEXE C</b> : Bases de données communales pour décrire la végétation .....	313
<b>ANNEXE D</b> : Résultats par classes d'occupation du sol des critères d'évaluation des unités d'analyse de la structure urbaine .....	317
<b>ANNEXE E</b> : Description des types de niveau 2 issus de la classification des mailles de Strasbourg.....	321
<b>ANNEXE F</b> : Matrices de confusions des classifications des mailles de Strasbourg et de Rennes .....	329





## **Annexe A**

Les trames vertes et bleues dans les SRCE adoptés en  
France métropolitaine

**Annexe A.1 – Liste des SRCE adoptés et des sous-trames considérées**

Régions	Date d'adoption	Nombre de sous-trames	Sous-trames considérées
Alsace	22/12/2014	5	milieu forestiers ; milieux ouverts non humides ; milieux humides ; milieux aquatiques ; milieu agricoles et anthropisés
Aquitaine	24/12/2015	8	boisements de feuillus et forêts mixtes ; boisements de conifères et milieux associés ; milieux ouverts et semi-ouverts ; systèmes bocagers ; milieux rocheux d'altitude ; milieux humides ; cours d'eau ; milieux côtiers
Auvergne	07/07/2015	6	milieux aquatiques et humides ; milieux forestiers ; milieux agropastoraux ; milieux cultivés ; milieux subalpins ; milieux thermophiles
Basse-Normandie	29/07/2014	5	milieux forestiers ; milieux ouverts ; milieux humides ; milieux aquatiques ; milieux littoraux
Bourgogne	06/05/2015	5	forêts ; prairies et bocage ; pelouse sèche ; plans d'eau et zone humides ; cours d'eau et milieux humides associés
Bretagne	02/11/2015	6	forêts ; landes pelouses et tourbières ; bocage ; zones humides ; cours d'eau ; littoral
Centre-Val de Loire	16/01/2015	10	pelouses et lisières sèches sur sols calcaires ; pelouses et landes sèches à humides sur sols acides ; milieux prairiaux ; espaces cultivés ; milieux bocagers ; boisements humides ; boisements sur sols acides ; boisements sur sols calcaires ; milieux humides ; cours d'eau
Champagne-Ardenne	08/12/2015	4	milieu boisés ; milieux ouverts ; milieux humides ; milieux aquatiques
Franche-Comté	02/12/2015	7	milieux forestiers ; milieux herbacés permanents ; milieux agricoles en mosaïque paysagère ; milieux xériques ouverts ; milieux souterrains ; milieux humides ; milieux aquatiques
Haute-Normandie	18/11/2014	4	milieux aquatiques ; milieux humides ; milieux sylvo-arborées ; milieux calcicoles ; milieux silicicoles
Île-de-France	21/10/2013	4	milieux arborés ; milieux herbacés ; grandes cultures ; milieux aquatiques et humides
Languedoc-Roussillon	20/11/2015	7	Forêts ; milieux semi-ouverts ; milieux ouverts ; cultures annuelles ; cultures pérennes ; zones humides ; plans d'eau et lagunes ; Milieux littoraux
Limousin	02/12/2015	5	milieux boisés ; milieux bocagers ; milieux secs et/ou thermophiles et/ou rocheux ; milieux humides ; milieux aquatiques
Lorraine	20/11/2015	5	milieux forestiers ; milieux herbacés ; zones alluviales ou humides ; milieux thermophiles ; milieux aquatiques des cours d'eau
Midi-Pyrénées	27/03/2015	8	milieux boisés de plaine ; milieux boisés d'altitude ; milieux ouverts et semi-ouverts de plaine ; milieux ouverts et semi-ouverts d'altitude ; milieux rocheux d'altitude ; milieux cultivés ; milieux humides ; cours d'eau
Nord-Pas-de-Calais	16/07/2014	10	estuaires ; falaises et estrans rocheux ; landes et pelouses acidiphiles ; terrils et autres milieux anthropiques ; coteaux calcaires ; dunes et estrans sableux ; zones humides ; prairies et/ou bocage ; forêts ; autres milieux

Régions	Date d'adoption	Nombre de sous-trames	Sous-trames considérées
Pays de la Loire	30/10/2015	6	milieux bocagers ; milieux boisés ; milieux littoraux ; milieux humides ; milieux aquatiques ; milieux ouverts particuliers secs
Poitou-Charentes	03/11/2015	5	milieux aquatiques ; systèmes bocagers ; forêts et landes ; plaines ouvertes ; pelouses sèches calcicoles
Provence-Alpes-Côte d'Azur	26/11/2014	6	milieux forestiers ; milieux semi-ouverts ; milieux ouverts ; zones humides ; eaux courantes ; milieux littoraux
Rhône-Alpes	16/07/2014	7	milieux aquatiques et humides ; forêts et pâturages d'altitude ; continuum des zones agricoles et lisières ; zones rocheuses d'altitudes ; pelouses et landes subalpines ; milieux thermophiles secs ; continuum forestier de basse altitude



## **Annexe B**

Nomenclatures des bases de données d'occupation du  
sol avec une représentation continue de l'espace

**Annexe B.1 – Nomenclature Corine Land Cover 2012 en 4 niveaux et 50 classes**

Niveau 1	Niveau 2	Niveau 3	Niveau 4	
1 – Territoires artificialisés	11 – Zones urbanisées	111 – Tissu urbain continu 112 – Tissu urbain discontinu	1110 – Tissu urbain continu 1120 – Tissu urbain discontinu	
	12 – Zones industrielles ou commerciales et réseaux de communication	121 – Zones industrielles ou commerciales et installations publiques 122 – Réseaux routier et ferroviaire et espaces associés 123 – Zones portuaires 124 – Aéroports	1210 – Zones industrielles ou commerciales et installations publiques 1220 – Réseaux routier et ferroviaire et espaces associés 1230 – Zones portuaires 1240 – Aéroports	
	13 – Mines, décharges et chantiers	131 – Extraction de matériaux 132 – Décharges 133 – Chantiers	1310 – Extraction de matériaux 1320 – Décharges 1330 – Chantiers	
	14 – Espaces verts artificialisés, non agricoles	141 – Espaces verts urbains 142 – Équipements sportifs et de loisirs	1410 – Espaces verts urbains 1420 – Équipements sportifs et de loisirs	
	2 – Territoires agricoles	21 – Terres arables	211 – Terres arables hors périmètres d'irrigation	2111 – Terres arables hors périmètres d'irrigation 2112 – Canne à sucre
			212 – Périmètres irrigués en permanence	2120 – Périmètres irrigués en permanence 2130 – Rizières
			213 – Rizières	2130 – Rizières
		22 – Cultures permanentes	221 – Vignobles 222 – Vergers et petits fruits	2210 – Vignobles 2221 – Vergers et petits fruits 2222 – Bananeraies 2223 – Palmeraies 2224 – Cafésiers
			223 – Oliveraies	2230 – Oliveraies
		23 – Prairies	231 – Prairies et autres surfaces toujours en herbe à usage agricole	2310 – Prairies et autres surfaces toujours en herbe à usage agricole
24 – Zones agricoles hétérogènes	241 – Cultures annuelles associées à des cultures permanentes 242 – Systèmes culturaux et parcellaires complexes 243 – Surfaces essentiellement agricoles, interrompues par des espaces naturels importants 244 – Territoires agroforestiers	2410 – Cultures annuelles associées à des cultures permanentes	2410 – Cultures annuelles associées à des cultures permanentes	
		2420 – Systèmes culturaux et parcellaires complexes	2420 – Systèmes culturaux et parcellaires complexes	
		2430 – Surfaces essentiellement agricoles, interrompues par des espaces naturels importants	2430 – Surfaces essentiellement agricoles, interrompues par des espaces naturels importants	
		2440 – Territoires agroforestiers	2440 – Territoires agroforestiers	

Niveau 1	Niveau 2	Niveau 3	Niveau 4	
3 – Forêts et milieux semi-naturels	31 – Forêts	311 – Forêts de feuillus	3111 – Forêts de feuillus	
		312 – Forêts de conifères	3112 – Mangroves	
		313 – Forêts mélangées	3120 – Forêts de conifères	
	32 – Milieux à végétation arbustive et/ou herbacée	321 – Pelouses et pâturages naturels	3210 – Pelouses et pâturages naturels	3210 – Pelouses et pâturages naturels
			322 – Landes et broussailles	3220 – Landes et broussailles
			323 – Végétation sclérophylle	3230 – Végétation sclérophylle
			324 – Forêt et végétation arbustive en mutation	3240 – Forêt et végétation arbustive en mutation
	33 – Espaces ouverts, sans ou avec peu de végétation	331 – Plages, dunes et sable	3310 – Plages, dunes et sable	3310 – Plages, dunes et sable
			332 – Roches nues	3320 – Roches nues
			333 – Végétation clairsemée	3330 – Végétation clairsemée
334 – Zones incendiées			3340 – Zones incendiées	
335 – Glaciers et neiges éternelles			3350 – Glaciers et neiges éternelles	
4 – Zones humides	41 – Zones humides intérieures	411 – Marais intérieurs	4110 – Marais intérieurs	
		412 – Tourbières	4120 – Tourbières	
	42 – Zones humides côtières	421 – Marais maritimes	4210 – Marais maritimes	
		422 – Marais salants	4220 – Marais salants	
		423 – Zones intertidales	4230 – Zones intertidales	
		424 – Zones humides côtières	4240 – Zones humides côtières	
5 – Surfaces en eau	51 – Eaux continentales	511 – Cours et voies d'eau	5111 – Cours et voies d'eau	
		512 – Plans d'eau	5112 – Cours et voies d'eau temporaires	
	52 – Eaux maritimes	521 – Lagunes littorales	5210 – Plans d'eau	
		522 – Estuaires	5210 – Lagunes littorales	
		523 – Mers et océans	5220 – Estuaires	
		5230 – Mers et océans	5230 – Mers et océans	

**Annexe B.2 – Nomenclature Urban Atlas 2012 en 4 niveaux et 20 classes**

Niveau 1	Niveau 2	Niveau 3	Niveau 4	
1 – Surface artificialisée	11 – Tissu urbain	111 – Tissu urbain continu (> 80 %)	1110 – Tissu urbain continu (> 80 %)	
		112 – Tissu urbain discontinu (10 % – 80 %)	1121 – Tissu urbain discontinu dense (50 % – 80 %)	
			1122 – Tissu urbain discontinu moyennement dense (30 % – 50 %)	
			1123 – Tissu urbain discontinu peu dense (10 % – 30 %)	
			1123 – Tissu urbain discontinu très peu dense (< 10 %)	
			1130 – Structure isolée	
			113 – Structure isolée	
			121 – Emprise industrielle, commerciale, publique, militaire et privée	1210 – Emprise industrielle, commerciale, publique, militaire et privée
			122 – Réseau routier et ferré et espace associé	1221 – Réseau routier rapide et espace associé
				1222 – Autre réseau routier et espace associé
		1223 – Réseau ferré et espace associé		
		1230 – Emprise portuaire	1230 – Emprise portuaire	
		124 – Aéroport	1240 – Aéroport	
	13 – Mine, décharge et site de construction	131 – Mine et décharge	1310 – Mine et décharge	
		133 – Site de construction	1330 – Site de construction	
		134 – Espace sans usage courant	1340 – Espace sans usage courant	
	14 – Espace vert artificialisé non agricole	141 – Espace vert urbain	1410 – Espace vert urbain	
		142 – Équipements sportifs et de loisirs	1420 – Équipements sportifs et de loisirs	
2 – Espace agricole, semi-naturel et zone humide	20 – Espace agricole, semi-naturel et zone humide	200 – Espace agricole, semi-naturel et zone humide	2000 – Espace agricole, semi-naturel et zone humide	
3 – Forêt	30 – Forêt	300 – Forêt	3000 – Forêt	
5 – Surface en eau	50 – Surface en eau	500 – Surface en eau	5000 – Surface en eau	



Annexe B.3 – Nomenclature BDOCS 2012 d'Alsace en 4 niveaux et 55 classes

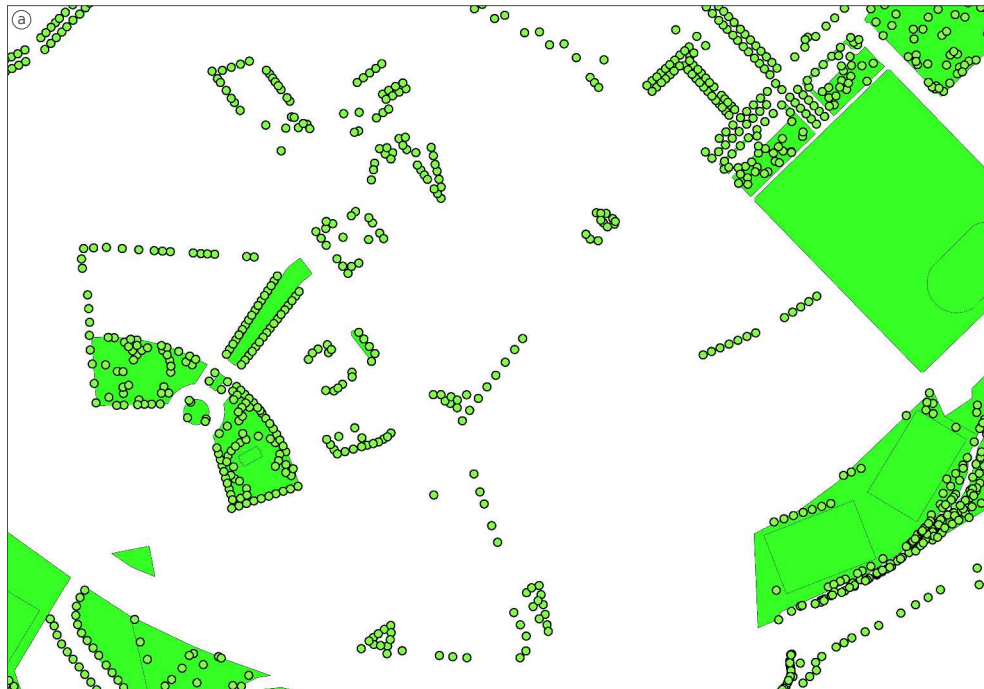
Niveau 1	Niveau 2	Niveau 3	Niveau 4	Surface minimale en ha
Territoires artificialisés	Habitat	Habitat continu (Centre ancien, centre ville)		0.5
		Habitat discontinu	Habitat collectif	0.5
			Habitat mixte	0.5
			Habitat individuel	0.5
	Espaces urbains spécialisés	Emprises scolaires et universitaires		0.5
		Emprises hospitalières		0.5
		Emprises culturelles et patrimoine		0.5
		Cimetières		0.5
		Autres espaces urbains spécialisés		0.5
		Grandes emprises	Emprises industrielles	
	Emprises commerciales et artisanales			0.5
	Zones d'activités tertiaires			0.5
	Emprises militaires			0.5
	Gravières et sablières		Bâtiments	0.5
			Zones d'exploitation	0.5
	Carrières		Bâtiments	0.5
			Zones d'exploitation	0.5
	Friches minières		Terrils	0.5
			Bâtiments industriels et espaces associés	0.5
	Chantiers et remblais			1
	Emprise réseau ferré			12 m
	Emprise réseau routier			12 m
	Emprises aéroporutaires (aéroport & aérodrome)		Pistes	0.5
			Bâtiments	0.5
			Autres espaces	0.5
			Emprises portuaires	
	Exploitations agricoles			0.3
	Espaces verts artificialisés	Espaces verts urbains	Pelouses et zones arborées	0.75
			Jardins ouvriers	0.75
		Golfs		0.5
	Equipements sportifs et de loisirs		0.5	
	Espaces libres	Friches industrielles		1
Autres espaces libres			0.5	
Territoires agricoles	Cultures annuelles		1	
	Cultures permanentes	Vignes		1
		Houblon		1
		Vergers	Vergers traditionnels	1
			Vergers intensifs	1
		Prairies		1
		Bosquets et haies		0.25
Cultures spécifiques		1		
Espaces forestiers et semi-naturels	Forêts	Forêts de feuillus		1
		Forêts de résineux		1
		Forêts mixtes		1
		Coupes à blanc et jeunes plantations		0.5
		Ripsisylves		0.03
	Formations pré-forestières	Pelouses et pâturages de montagne		0.7
		Tourbières et marais		0.5
		Landes		0.75
Fourrés et fructicées		0.5		
Roches nues		1		
Milieux hydrographiques	Surfaces en eau	Cours d'eau principaux		0.03
		Canaux principaux		1
		Etangs et lacs		0.03
		Bassins artificiels		0.03



## **Annexe C**

Bases de données communales pour décrire la  
végétation

**Annexe C.1** – Exemple (a) de bases de données pour décrire la végétation de Strasbourg et (b) image Pléiades associée (source : Base de données CARTO 10 000 – Base de données Arbre point espèce; Eurométropole de Strasbourg)



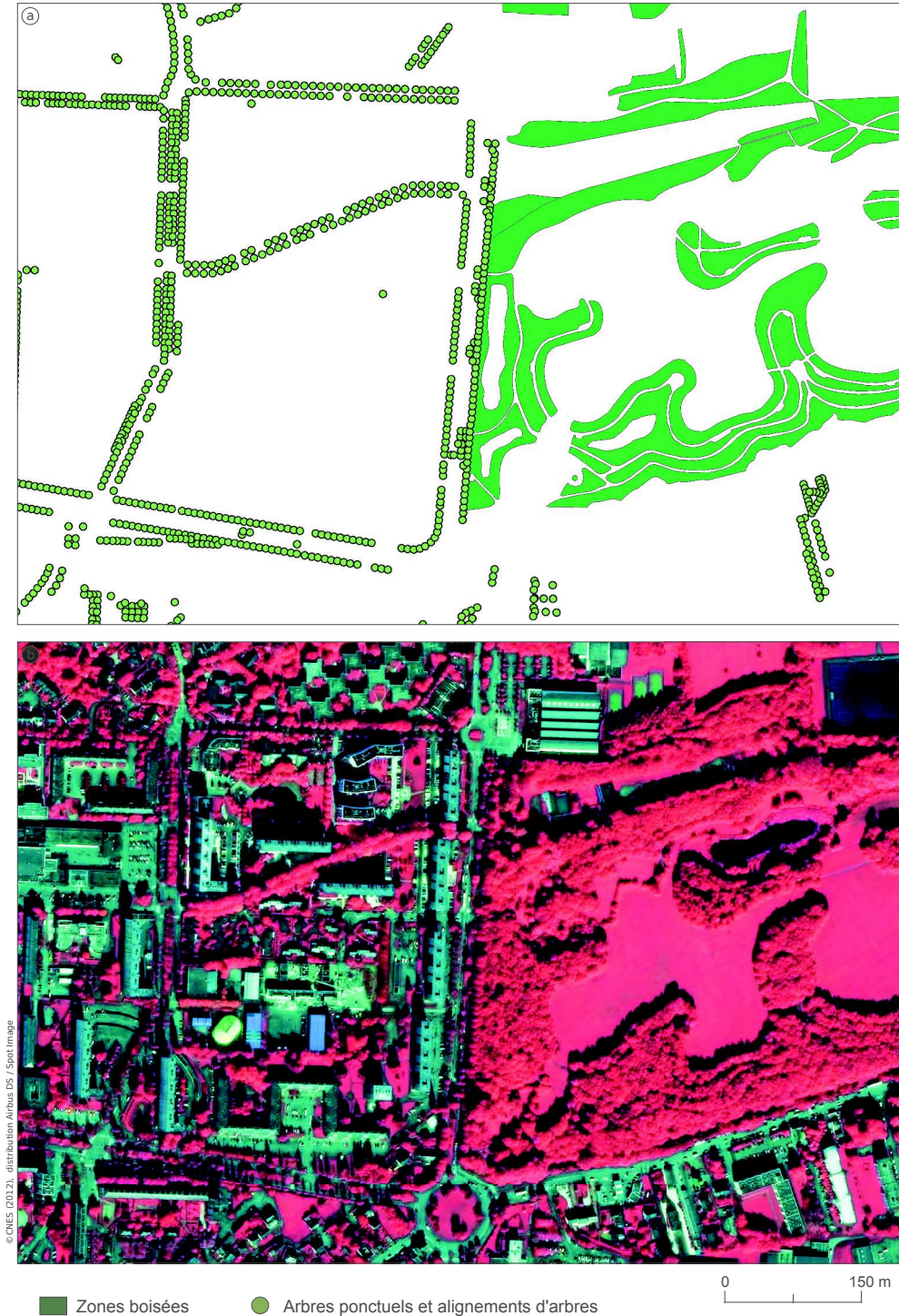
© CNES (2012). distribution Airbus DS / Spot Image

■ Espaces verts

● Arbres ponctuels gérés par la commune

0 150 m

Annexe C.2 – Exemple (a) de base de données pour décrire la végétation de Rennes et (b) image Pléiades associée (source : Cartographie générale au 1/2 000 de la ville de Rennes ; Rennes Métropole)





## **Annexe D**

Résultats par classes d'occupation du sol des critères  
d'évaluation des unités d'analyse de la structure  
urbaine

**Annexe D.1** – Rang moyen de la variance intra-classe par unité d’analyse et par classe d’occupation du sol de la BDOCS 2012 (CIGAL)  
(Le symbole • signifie que le critère d’évaluation n’a pas pu être calculé pour cette classe et cette unité)

Classe d’occupation du sol	Parcelle	Îlot	Grille carrée			Grille hexagonale		
			125 m	250 m	500 m	125 m	250 m	500 m
Habitat continu	6,9	5,1	4,4	4,5	3,4	4,7	3,7	3,2
Habitat collectif	7,0	5,2	5,0	3,9	2,7	4,9	3,7	3,4
Habitat individuel	6,8	4,4	4,3	3,9	4,3	4,7	3,6	4,0
Habitat mixte	6,8	5,2	4,9	3,9	3,4	5,0	3,2	3,2
Emprises scolaires et universitaires	6,5	5,1	4,6	3,7	3,9	4,6	3,8	3,6
Emprises hospitalières	6,3	4,9	5,0	3,8	3,7	4,8	3,6	2,6
Emprises culturelles et patrimoine	5,5	3,9	4,0	3,4	2,4	4,7	3,8	7,8
Espaces urbains spécialisés peu ou pas bâtis	6,3	4,9	4,9	4,1	2,7	5,3	4,0	3,6
Emprises commerciales et artisanales	6,2	4,7	5,0	3,9	3,5	5,0	4,0	3,6
Emprises industrielles	6,2	5,0	4,7	3,9	3,9	4,8	3,8	3,7
Zones d’activités tertiaires	6,8	4,9	5,1	4,4	2,9	5,3	3,7	2,9
Emprises militaires	5,8	4,5	5,0	3,5	4,4	4,6	3,9	3,5
Emprises réseau routier	5,8	4,4	4,7	5,1	3,4	4,9	3,9	3,8
Grandes emprises peu ou pas bâties	6,4	4,6	4,9	4,2	3,3	5,1	4,3	3,2
Emprises réseau ferré	5,8	4,5	5,0	4,7	3,6	4,9	4,3	2,9
Emprises aéroportuaires	5,6	4,7	4,3	4,6	3,8	4,1	4,1	4,2
Emprises portuaires	5,5	4,1	4,6	4,0	3,1	4,4	3,9	3,6
Pelouses et zones arborées	5,8	4,9	4,9	4,0	3,1	5,1	3,8	4,2
Jardins ouvriers	6,3	4,3	4,9	4,4	3,6	5,2	4,2	2,9
Équipements sportifs et de loisirs	6,6	4,2	5,0	3,8	4,1	4,7	4,0	3,4
Cultures annuelles	5,3	4,5	3,7	4,8	5,3	3,4	4,0	5,0
Cultures permanentes	5,7	4,8	4,6	4,4	4,1	5,0	3,9	3,4
Forêts	5,7	4,9	3,8	4,3	4,5	3,5	4,5	4,8
Formations préforestières	5,8	4,8	5,1	4,7	3,5	4,7	4,2	2,9
Ripisylves	3,2	2,2	2,3	•	•	2,3	4,9	•
Cours d’eau et canaux	5,2	4,3	4,6	4,6	4,2	4,5	4,0	4,4
Surface en eau	5,3	3,9	4,3	5,0	4,4	3,8	4,6	4,5



**Annexe D.2** – Rang de la trace de Pillai-Bartlett pour la MANOVA par unité d'analyse et par classe d'occupation du sol de la BDOCS 2012 (CIGAL) (Le symbole • signifie que le critère d'évaluation n'a pas pu être calculé pour cette classe et cette unité)

Classe d'occupation du sol	Parcelle	Îlot	Grille carrée			Grille hexagonale		
			125 m	250 m	500 m	125 m	250 m	500 m
Habitat continu	8	1	6	5	3	7	4	2
Habitat collectif	7	3	1	5	6	2	4	8
Habitat individuel	8	1	5	3	6	2	4	7
Habitat mixte	8	5	4	2	3	7	6	1
Emprises scolaires et universitaires	8	5	2	4	6	1	3	7
Emprises hospitalières	8	1	2	5	7	3	6	4
Emprises culturelles et patrimoine	7	2	6	4	3	5	8	1
Espaces urbains spécialisés peu ou pas bâtis	8	7	5	4	3	6	2	1
Emprises commerciales et artisanales	8	6	3	5	1	2	4	7
Emprises industrielles	8	5	1	4	6	2	7	3
Zones d'activités tertiaires	8	1	6	2	4	5	7	3
Emprises militaires	3	1	7	8	6	5	4	2
Grandes emprises peu ou pas bâties	8	1	4	5	3	6	2	7
Emprises réseau routier	8	1	3	4	6	2	5	7
Emprises réseau ferré	6	1	4	2	3	5	7	8
Emprises aéroportuaires	8	7	5	4	1	6	3	2
Emprises portuaires	7	8	5	4	3	6	1	2
Pelouses et zones arborées	8	5	4	2	6	3	1	7
Jardins ouvriers	6	8	4	5	1	2	3	7
Équipements sportifs et de loisirs	7	8	6	3	2	5	1	4
Cultures annuelles	8	7	2	4	5	1	3	6
Cultures permanentes	2	1	5	7	4	6	8	3
Forêts	8	7	2	4	3	1	5	6
Formations préforestières	8	7	1	5	6	4	3	2
Ripisylves	2	1	4	•	•	3	5	•
Cours d'eau et canaux	8	7	1	2	6	3	4	5
Surface en eau	6	8	2	5	7	1	4	3

**Annexe D.3** – Rang de de l'incertitude de l'attribution de la classe d'occupation du sol à une unité par unité d'analyse par classe d'occupation du sol de la BDOCS 2012 (CIGAL) (Le symbole • signifie que le critère d'évaluation n'a pas pu être calculé pour cette classe et cette unité)

Classe d'occupation du sol	Parcelle	Îlot	Grille carrée			Grille hexagonale		
			125 m	250 m	500 m	125 m	250 m	500 m
Habitat continu	2	1	4	6	7	3	5	8
Habitat collectif	2	1	3	6	7	4	5	8
Habitat individuel	1	2	3	6	8	4	5	7
Habitat mixte	1	2	4	6	7	3	5	8
Emprises scolaires et universitaires	1	2	3	8	7	4	5	6
Emprises hospitalières	2	1	3	5	8	4	6	7
Emprises culturelles et patrimoine	2	1	6	4	8	5	7	3
Espaces urbains spécialisés peu ou pas bâtis	2	1	4	5	8	3	6	7
Emprises commerciales et artisanales	2	1	4	5	7	3	6	8
Emprises industrielles	2	1	3	6	8	4	5	7
Zones d'activités tertiaires	2	1	4	6	8	3	5	7
Emprises militaires	2	1	4	6	8	3	5	7
Grandes emprises peu ou pas bâties	2	1	4	5	8	3	6	7
Emprises réseau routier	7	1	2	4	8	3	5	6
Emprises réseau ferré	2	1	3	6	8	4	5	7
Emprises aéroportuaires	4	1	2	6	7	3	5	8
Emprises portuaires	2	1	3	5	8	4	6	7
Pelouses et zones arborées	4	1	2	6	8	3	5	7
Jardins ouvriers	1	2	3	6	8	4	5	7
Équipements sportifs et de loisirs	1	2	4	6	7	3	5	8
Cultures annuelles	8	7	2	3	5	1	4	6
Cultures permanentes	4	1	3	6	8	2	5	7
Forêts	8	5	1	4	6	2	3	7
Formations préforestières	3	8	2	4	6	1	5	7
Ripisylves	2	1	3	•	•	5	4	•
Cours d'eau et canaux	8	7	2	4	6	1	3	5
Surface en eau	7	8	1	4	5	2	3	6

## **Annexe E**

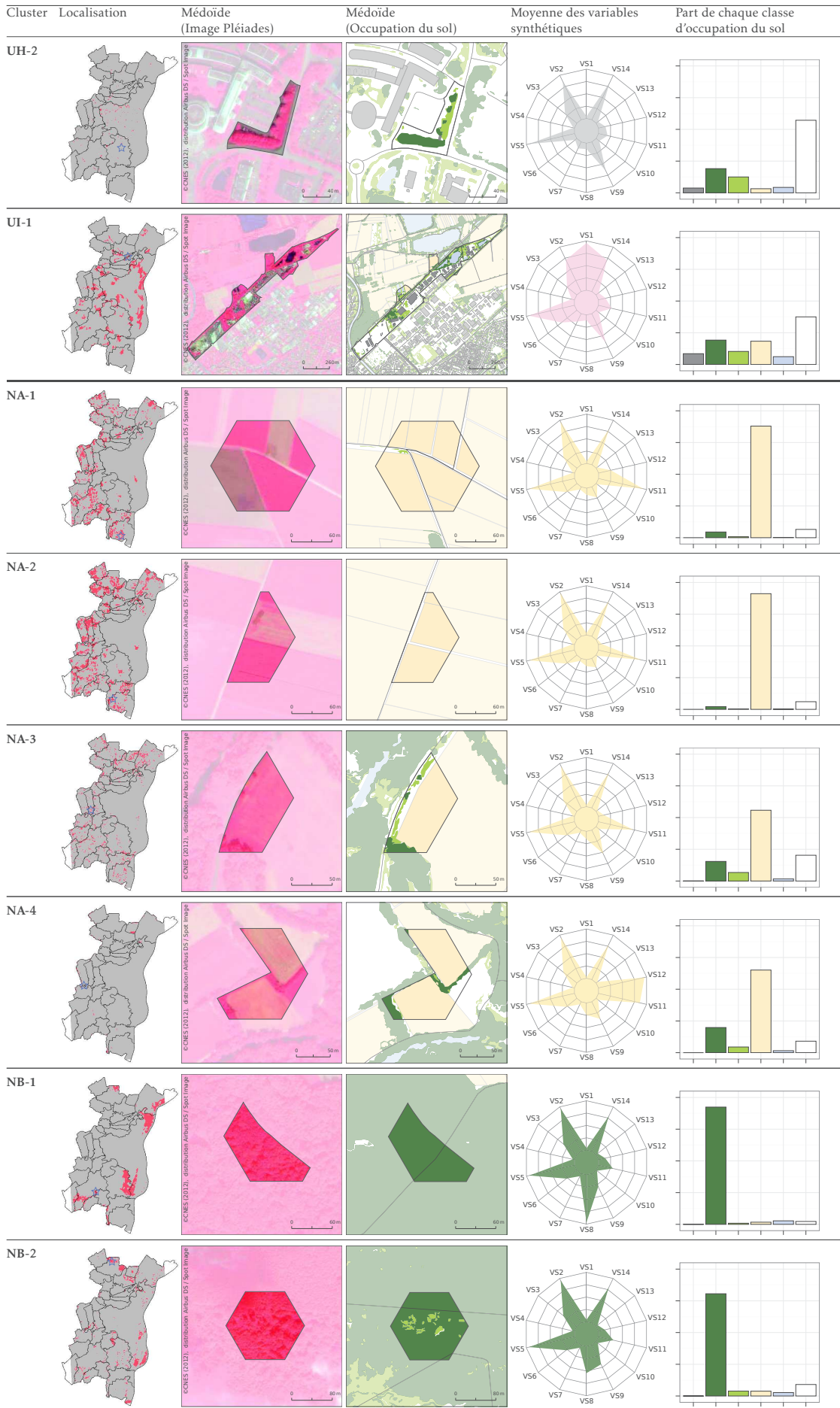
Description des types de niveau 2 issus de la  
classification des mailles de Strasbourg

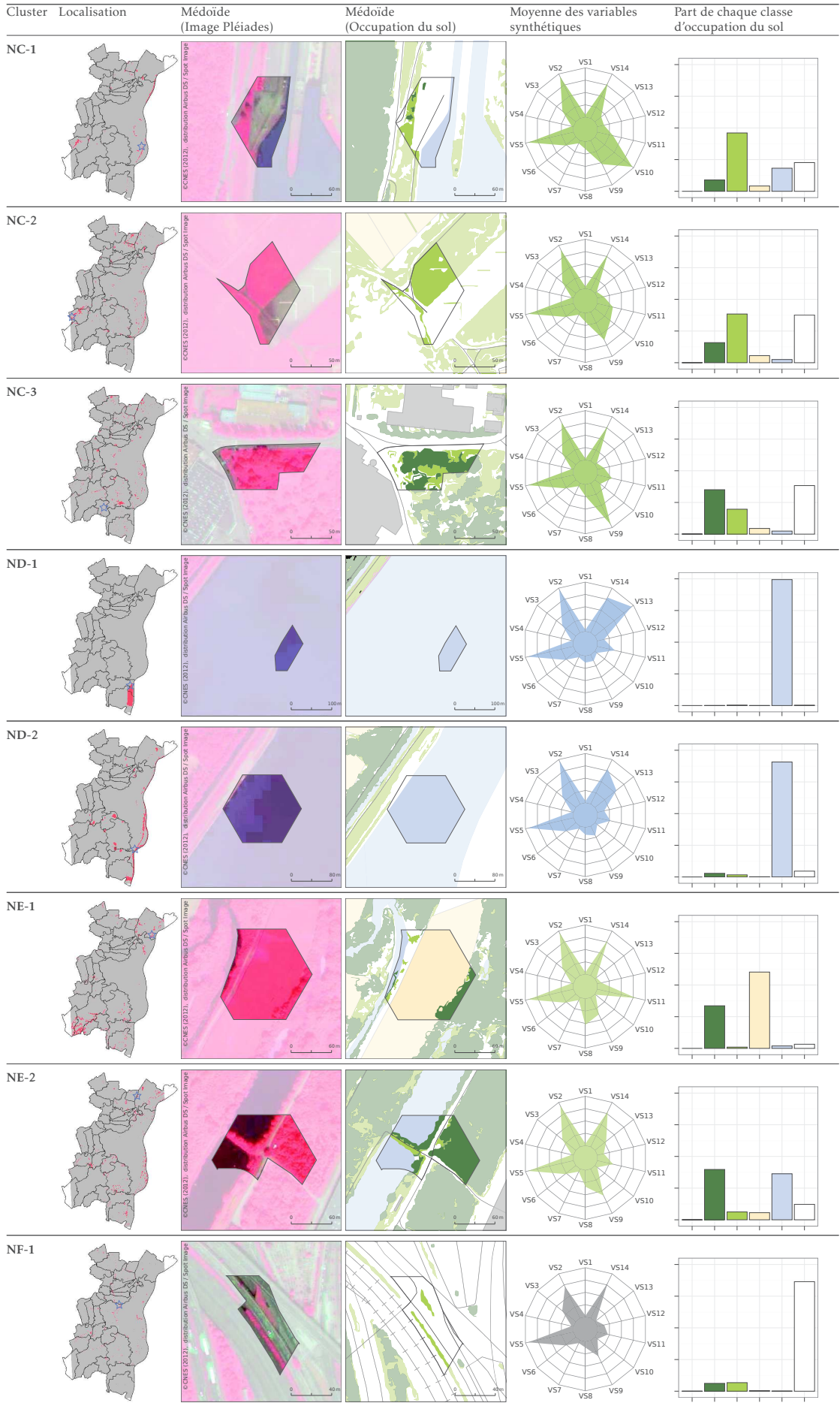
Annexe E.1 – Présentation des clusters obtenus au niveau 2 de la classification des mailles de Strasbourg



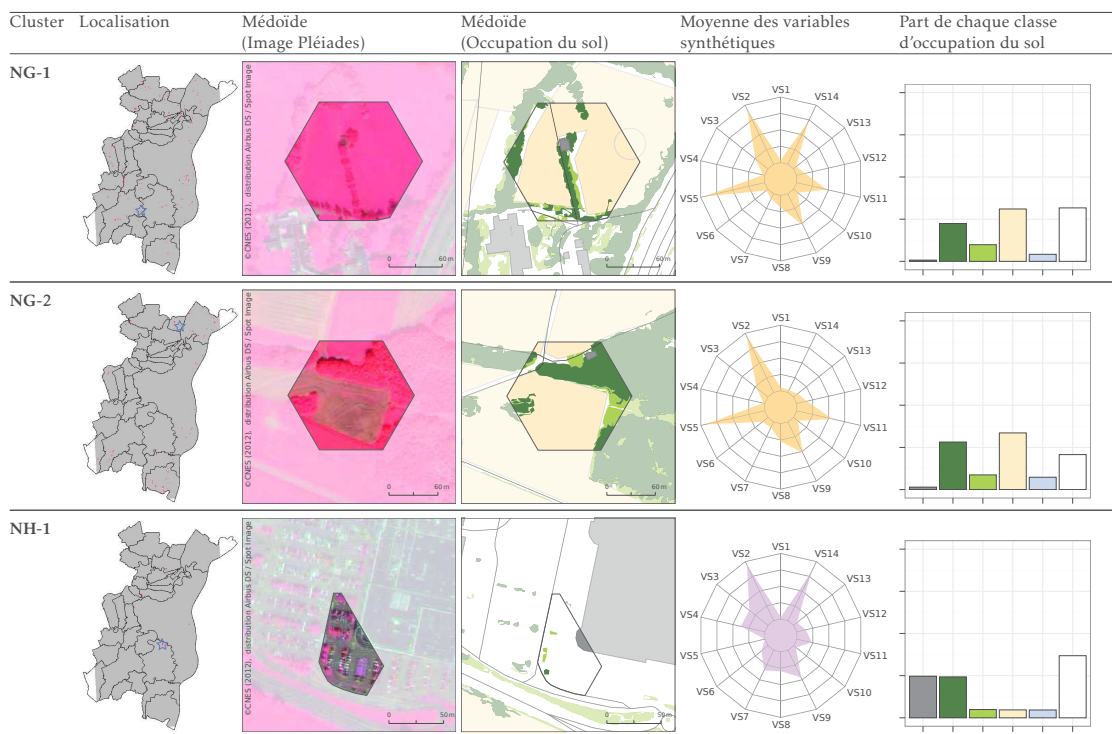


Cluster	Localisation	Médoïde (Image Pléiades)	Médoïde (Occupation du sol)	Moyenne des variables synthétiques	Part de chaque classe d'occupation du sol
UF-1					
UF-2					
UF-3					
UF-4					
UF-5					
UG-1					
UG-2					
UH-1					











## **Annexe F**

Matrices de confusions des classifications des mailles  
de Strasbourg et de Rennes

**Annexe F.1** – Matrice de confusion et précisions par groupe issus de la classification des mailles de Strasbourg évaluées par la BDOCS 2012 (CIGAL) (Le symbole • signifie que la précision n'a pas pu être calculée car le type n'as pas d'équivalent dans la vérité terrain)

	1	2	3	4	5	6	7	8	9	10	Précision
NA	0	0	2	0	142	5	30	7325	135	2	0,96
NB	0	0	2	0	12	14	53	102	2920	7	0,94
NC	0	4	10	2	375	302	221	606	369	74	•
ND	0	0	0	0	4	1	1	3	0	1081	0,99
NE	0	1	2	0	49	4	76	663	426	235	•
NF	0	1	0	0	132	377	22	70	10	1	0,62
NG	0	0	14	0	58	15	37	191	58	12	•
NH	0	0	0	0	13	0	5	1	4	0	0,57
UA	269	74	10	32	40	0	0	0	0	0	0,63
UB	54	281	36	29	164	1	6	2	0	4	0,49
UC	67	23	398	62	13	0	0	0	0	0	•
UD	30	18	987	20	128	6	28	126	18	3	0,72
UE	15	56	152	25	43	0	5	1	0	1	0,08
UF	5	36	21	3	299	11	58	19	7	2	0,65
UG	15	25	13	3	53	307	24	6	3	10	0,67
UH	13	44	49	6	267	68	95	38	13	24	•
UI	0	0	13	0	21	1	7	25	1	2	•

1–Habitat continu, 2–Habitat collectif, 3–Habitat individuel, 4–Habitat mixte, 5–Grandes emprises, 6–Emprises réseaux, 7–Espaces verts artificialisés, 8–Territoires agricoles, 9–Espaces forestiers et semi-naturels, 10–Milieux hydrographiques

**Annexe F.2 – Matrice de confusion et précisions par type issus de la classification des mailles de Strasbourg évaluées par la BDOCS 2012 (CIGAL) (Le symbole • signifie que la précision n'a pas pu être calculée car le type n'as pas d'équivalent dans la vérité terrain)**

	1	2	3	4	5	6	7	8	9	10	Précision
NA-1	0	0	0	0	23	0	0	2316	22	0	0,98
NA-2	0	0	0	0	33	0	2	3435	3	0	0,99
NA-3	0	0	2	0	86	4	26	1432	79	1	0,88
NA-4	0	0	0	0	0	1	2	142	31	1	0,80
NB-1	0	0	1	0	0	0	3	12	1799	4	0,99
NB-2	0	0	1	0	12	14	50	90	1121	3	0,87
NC-1	0	0	0	0	56	8	2	123	59	61	•
NC-2	0	0	0	0	190	148	40	206	114	10	•
NC-3	0	4	10	2	129	146	179	277	196	3	•
ND-1	0	0	0	0	0	0	0	0	0	291	1
ND-2	0	0	0	0	4	1	1	3	0	790	0,99
NE-1	0	0	0	0	0	0	1	543	186	0	•
NE-2	0	1	2	0	49	4	75	120	240	235	•
NF-1	0	1	0	0	132	377	22	70	10	1	0,62
NG-1	0	0	11	0	41	13	28	125	33	6	•
NG-2	0	0	3	0	17	2	9	66	25	6	•
NH-1	0	0	0	0	13	0	5	1	4	0	0,48
UA-1	109	17	2	8	3	0	0	0	0	0	0,78
UA-2	85	22	2	3	18	0	0	0	0	0	0,65
UA-3	75	35	6	21	19	0	0	0	0	0	0,48
UB-1	15	105	10	9	36	1	2	0	0	2	0,58
UB-2	14	42	1	3	45	0	0	1	0	1	0,39
UB-3	25	134	25	17	83	0	4	1	0	1	0,46
UC-1	55	15	138	54	10	0	0	0	0	0	•
UC-2	12	8	260	8	3	0	0	0	0	0	•
UD-1	17	8	264	13	19	1	3	6	0	0	0,80
UD-2	0	4	328	2	5	1	1	2	0	0	0,96
UD-3	1	1	199	0	11	1	1	11	0	0	0,88
UD-4	0	1	55	0	23	3	15	69	6	3	0,31
UD-5	9	4	110	2	48	0	5	14	8	0	0,55
UD-6	3	0	31	3	22	0	3	24	4	0	0,34
UE-1	15	56	152	25	43	0	5	1	0	1	0,08
UF-1	0	0	0	0	15	0	0	0	0	0	1

1–Habitat continu, 2–Habitat collectif, 3–Habitat individuel, 4–Habitat mixte, 5–Grandes emprises, 6–Emprises réseaux, 7–Espaces verts artificialisés, 8–Territoires agricoles, 9–Espaces forestiers et semi-naturels, 10–Milieux hydrographiques

	1	2	3	4	5	6	7	8	9	10	Précision
UF-2	2	0	0	0	50	5	6	0	1	0	0,78
UF-3	0	4	1	1	61	0	1	1	0	0	0,88
UF-4	3	18	17	1	91	0	17	6	2	0	0,59
UF-5	0	14	3	1	82	6	34	12	4	2	0,52
UG-1	3	4	2	1	11	178	4	3	0	0	0,86
UG-2	12	21	11	2	42	129	20	3	3	10	0,51
UH-1	9	33	28	2	199	51	68	25	10	19	•
UH-2	4	11	21	4	68	17	27	13	3	5	•
UI-1	0	0	13	0	21	1	7	25	1	2	•

1–Habitat continu, 2–Habitat collectif, 3–Habitat individuel, 4–Habitat mixte, 5–Grandes emprises, 6–Emprises réseaux, 7–Espaces verts artificialisés , 8–Territoires agricoles, 9–Espaces forestiers et semi-naturels, 10–Milieux hydrographiques

**Annexe E.3** – Matrice de confusion et précisions par groupe issus de la classification des mailles de Strasbourg évaluées par la base de données Urban Atlas 2012 (Le symbole • signifie que la précision n'a pas pu être calculée car le type n'as pas d'équivalent dans la vérité terrain)

	1	2	3	4	5	6	7	8	Précision
NA	0	75	42	75	7277	91	3	78	0,95
NB	0	18	43	75	174	2737	16	47	0,88
NC	17	275	398	256	602	245	94	76	•
ND	1	8	0	0	3	2	1074	2	0,99
NE	8	46	30	77	724	296	230	45	•
NF	1	113	385	32	65	8	3	6	0,63
NG	7	55	19	31	192	53	17	11	•
NH	0	14	0	4	2	3	0	0	0,61
UA	363	61	0	1	0	0	0	0	0,85
UB	373	188	5	7	2	0	2	0	0,65
UC	545	18	0	0	0	0	0	0	0,97
UD	1064	128	8	29	118	13	2	2	0,78
UE	250	40	0	5	1	1	0	1	0,84
UF	74	285	13	60	22	2	2	3	0,62
UG	29	39	346	30	1	0	14	0	0,75
UH	94	224	101	114	28	8	37	11	•
UI	15	23	1	7	21	0	2	1	•

1–Habitat, 2–Grandes emprises, 3–Emprises réseaux, 4–Espaces verts artificialisés, 5–Territoires agricoles, 6–Espaces forestiers et semi-naturels, 7–Milieux hydrographiques, 8–Autre

**Annexe F.4 – Matrice de confusion et précisions par type issus de la classification des mailles de Strasbourg évaluées par la base de données Urban Atlas 2012 (Le symbole • signifie que la précision n’a pas pu être calculée car le type n’as pas d’équivalent dans la vérité terrain)**

	1	2	3	4	5	6	7	8	Précision
NA-1	0	1	1	6	2309	15	0	29	0,98
NA-2	0	11	6	28	3403	4	0	21	0,98
NA-3	0	63	34	39	1423	42	3	26	0,87
NA-4	0	0	1	2	142	30	0	2	0,80
NB-1	0	1	2	12	55	1729	10	10	0,95
NB-2	0	17	41	63	119	1008	6	37	0,78
NC-1	0	47	19	5	111	33	73	21	•
NC-2	0	147	192	40	205	85	11	28	•
NC-3	17	81	187	211	286	127	10	27	•
ND-1	0	0	0	0	0	0	291	0	1
ND-2	1	8	0	0	3	2	783	2	0,98
NE-1	0	0	1	3	554	146	1	25	•
NE-2	8	46	29	74	170	150	229	20	•
NF-1	1	113	385	32	65	8	3	6	0,63
NG-1	7	38	19	22	125	30	11	5	•
NG-2	0	17	0	9	67	23	6	6	•
NH-1	0	14	0	4	2	3	0	0	0,61
UA-1	131	8	0	0	0	0	0	0	0,94
UA-2	106	23	0	1	0	0	0	0	0,82
UA-3	126	30	0	0	0	0	0	0	0,81
UB-1	131	44	2	3	0	0	0	0	0,73
UB-2	50	52	2	0	2	0	1	0	0,47
UB-3	192	92	1	4	0	0	1	0	0,66
UC-1	260	12	0	0	0	0	0	0	0,96
UC-2	285	6	0	0	0	0	0	0	0,98
UD-1	304	21	1	2	3	0	0	0	0,92
UD-2	333	5	2	1	2	0	0	0	0,97
UD-3	204	8	2	1	10	0	0	0	0,91
UD-4	62	22	3	15	65	6	2	0	0,35
UD-5	124	48	0	6	16	5	0	1	0,62
UD-6	37	24	0	4	22	2	0	1	0,41
UE-1	250	40	0	5	1	1	0	1	0,84

1–Habitat, 2–Grandes emprises, 3–Emprises réseaux, 4–Espaces verts artificialisés, 5–Territoires agricoles, 6–Espaces forestiers et semi-naturels, 7–Milieux hydrographiques, 8–Autre



	1	2	3	4	5	6	7	8	Précision
UF-1	0	15	0	0	0	0	0	0	1
UF-2	1	50	6	6	0	1	0	0	0,78
UF-3	6	57	2	0	2	0	0	2	0,83
UF-4	45	83	0	21	5	0	0	1	0,53
UF-5	22	80	5	33	15	1	2	0	0,51
UG-1	12	7	178	8	1	0	0	0	0,86
UG-2	17	32	168	22	0	0	14	0	0,66
UH-1	55	162	79	85	21	4	31	7	•
UH-2	39	62	22	29	7	4	6	4	•
UI-1	15	23	1	7	21	0	2	1	•

1–Habitat, 2–Grandes emprises, 3–Emprises réseaux, 4–Espaces verts artificialisés, 5–Territoires agricoles, 6–Espaces forestiers et semi-naturels, 7–Milieux hydrographiques, 8–Autre

**Annexe F.5** – Matrice de confusion et précisions par groupe issus de la classification des mailles de Rennes évaluées par la base de données Urban Atlas 2012 (Le symbole • signifie que la précision n'a pas pu être calculée car le type n'as pas d'équivalent dans la vérité terrain)

	1	2	3	4	5	6	7	8	Précision
NA	4	63	16	7	7324	29	0	8	0,98
NB	10	65	28	66	430	1258	0	3	0,68
NC	76	256	471	128	1507	272	34	22	•
ND	0	3	0	0	0	0	70	0	0,96
NE	5	23	10	47	1009	44	97	2	•
NF	12	97	330	1	71	1	0	0	0,64
NG	4	14	3	3	126	10	2	13	•
NH	0	19	0	2	1	0	0	0	0,86
UA	70	9	1	0	0	0	0	1	0,86
UB	370	119	4	1	1	1	1	0	0,74
UC	71	13	1	0	0	0	0	0	0,84
UD	570	137	6	16	383	6	4	22	0,50
UE	169	11	0	0	0	0	0	0	0,94
UF	150	244	3	22	29	4	4	3	0,53
UG	8	7	170	0	2	0	14	0	0,85
UH	90	158	106	30	23	1	13	5	•
UI	14	13	1	4	36	0	0	0	•

1–Habitat, 2–Grandes emprises, 3–Emprises réseaux, 4–Espaces verts artificialisés, 5–Territoires agricoles, 6–Espaces forestiers et semi-naturels, 7–Milieux hydrographiques, 8–Autre

**Annexe F.6** – Matrice de confusion et précisions par type issus de la classification des mailles de Rennes évaluées par la base de données Urban Atlas 2012 (Le symbole • signifie que la précision n'a pas pu être calculée car le type n'as pas d'équivalent dans la vérité terrain)

	1	2	3	4	5	6	7	8	Précision
NA-1	0	4	1	0	1812	5	0	0	0,99
NA-2	0	15	0	0	3077	4	0	1	0,99
NA-3	4	43	15	5	2417	20	0	7	0,96
NA-4	0	1	0	2	18	0	0	0	0,86
NB-1	1	1	1	6	89	1048	0	0	0,91
NB-2	9	64	27	60	341	210	0	3	0,29
NC-1	3	20	13	2	146	135	18	2	•
NC-2	23	143	269	52	696	85	10	7	•
NC-3	50	93	189	74	665	52	6	13	•
ND-2	0	3	0	0	0	0	70	0	0,96
NE-1	1	3	0	1	787	22	0	2	•
NE-2	4	20	10	46	222	22	97	0	•
NF-1	12	97	330	1	71	1	0	0	0,64
NG-1	3	9	3	3	65	6	0	8	•
NG-2	1	5	0	0	61	4	2	5	•
NH-1	0	19	0	2	1	0	0	0	0,86
UA-1	21	3	1	0	0	0	0	0	0,84
UA-2	21	2	0	0	0	0	0	1	0,88
UA-3	28	4	0	0	0	0	0	0	0,88
UB-1	42	15	1	0	0	0	1	0	0,71
UB-2	105	52	1	1	1	1	0	0	0,65
UB-3	223	52	2	0	0	0	0	0	0,81
UC-1	53	2	0	0	0	0	0	0	0,96
UC-2	18	11	1	0	0	0	0	0	0,60
UD-1	239	24	0	0	3	0	0	1	0,90
UD-2	106	8	1	1	2	0	0	0	0,90
UD-3	122	22	3	3	44	1	0	10	0,60
UD-4	59	26	2	11	143	3	3	0	0,24
UD-5	32	36	0	0	57	2	1	2	0,25
UD-6	12	21	0	1	134	0	0	9	0,07
UE-1	169	11	0	0	0	0	0	0	0,94
UF-1	0	4	0	0	0	0	0	0	1
UF-2	2	19	1	0	0	0	0	0	0,86

1–Habitat, 2–Grandes emprises, 3–Emprises réseaux, 4–Espaces verts artificialisés, 5–Territoires agricoles, 6–Espaces forestiers et semi-naturels, 7–Milieux hydrographiques, 8–Autre

	1	2	3	4	5	6	7	8	Précision
UF-3	23	40	0	1	4	0	0	1	0,58
UF-4	57	62	0	7	4	2	2	2	0,46
UF-5	68	119	2	14	21	2	2	0	0,52
UG-1	0	1	56	0	0	0	0	0	0,98
UG-2	8	6	114	0	2	0	14	0	0,79
UH-1	78	133	99	22	21	1	10	2	•
UH-2	12	25	7	8	2	0	3	3	•
UI-1	14	13	1	4	36	0	0	0	•

1–Habitat, 2–Grandes emprises, 3–Emprises réseaux, 4–Espaces verts artificialisés, 5–Territoires agricoles, 6–Espaces forestiers et semi-naturels, 7–Milieux hydrographiques, 8–Autre

**Annexe E.7** – Matrice de confusion et précisions par groupe issus de la classification des mailles de Rennes évaluées par la base de données des îlots photo-interprétée (Le symbole • signifie que la précision n’a pas pu être calculée car le type n’as pas d’équivalent dans la vérité terrain)

	1	2	3	4	5	6	7	Précision
NA	0	0	15	0	3	7	30	•
NB	0	0	0	1	11	3	54	•
NC	0	7	10	1	48	16	152	•
ND	0	0	0	0	8	0	3	•
NE	0	0	0	0	6	3	20	•
NF	0	3	0	0	32	2	104	•
NG	0	0	0	0	3	1	1	•
NH	0	0	0	0	0	0	2	•
UA	37	31	0	7	6	0	0	0,46
UB	39	172	35	117	124	1	3	0,35
UC	2	8	30	28	12	0	1	•
UD	0	6	201	50	36	6	29	0,61
UE	1	16	60	85	15	0	1	0,48
UF	1	44	27	36	194	6	27	0,58
UG	0	1	2	0	8	1	165	•
UH	0	41	10	6	98	4	138	•
UI	0	0	10	4	15	0	1	•

1–Habitat continu, 2–Habitat collectif, 3–Habitat individuel, 4–Habitat mixte, 5–Grandes emprises, 6–Zones spécialisées peu ou pas bâties, 7–Autre

**Annexe F.8** – Matrice de confusion et précisions par type issus de la classification des mailles de Rennes évaluées par la base de données des îlots photo-interprétée (Le symbole • signifie que la précision n’a pas pu être calculée car le type n’as pas d’équivalent dans la vérité terrain)

	1	2	3	4	5	6	7	Précision
NA-1	0	0	3	0	0	0	8	•
NA-2	0	0	2	0	1	3	6	•
NA-3	0	0	10	0	2	4	16	•
NB-1	0	0	0	0	0	0	6	•
NB-2	0	0	0	1	11	3	48	•
NC-1	0	0	0	0	8	3	6	•
NC-2	0	1	3	0	15	1	90	•
NC-3	0	6	7	1	25	12	56	•
ND-2	0	0	0	0	8	0	3	•
NE-2	0	0	0	0	6	3	20	•
NF-1	0	3	0	0	32	2	104	•
NG-1	0	0	0	0	3	1	1	•
NH-1	0	0	0	0	0	0	2	•
UA-1	18	6	0	0	1	0	0	0,72
UA-2	13	9	0	0	2	0	0	0,54
UA-3	6	16	0	7	3	0	0	0,19
UB-1	6	26	6	3	15	1	1	0,45
UB-2	30	68	0	8	53	0	0	0,43
UB-3	3	78	29	106	56	0	2	0,28
UC-1	2	4	17	28	3	0	0	•
UC-2	0	4	13	0	9	0	1	•
UD-1	0	4	110	48	11	0	4	0,62
UD-2	0	1	54	2	6	0	1	0,84
UD-3	0	0	24	0	1	1	7	0,73
UD-4	0	1	12	0	5	5	16	0,31
UD-5	0	0	1	0	13	0	0	0,07
UD-6	0	0	0	0	0	0	1	0
UE-1	1	16	60	85	15	0	1	0,48
UF-1	0	0	0	0	2	0	0	1
UF-2	0	1	0	0	17	0	0	0,94
UF-3	1	14	1	0	29	0	4	0,59
UF-4	0	8	12	20	50	2	7	0,51

1–Habitat continu, 2–Habitat collectif, 3–Habitat individuel, 4–Habitat mixte, 5–Grandes entreprises, 6–Zones spécialisées peu ou pas bâties, 7–Autre

	1	2	3	4	5	6	7	Précision
UF-5	0	21	14	16	96	4	16	0,57
UG-1	0	0	1	0	2	0	43	•
UG-2	0	1	1	0	6	1	122	•
UH-1	0	37	8	5	78	4	119	•
UH-2	0	4	2	1	20	0	19	•
UI-1	0	0	10	4	15	0	1	•

1–Habitat continu, 2–Habitat collectif, 3–Habitat individuel, 4–Habitat mixte, 5–Grandes emprises, 6–Zones spécialisées peu ou pas bâties, 7–Autre





## Table des matières

Remerciements .....	i
Liste des abréviations .....	iii
Sommaire .....	vii
Introduction générale .....	1
<b>PARTIE I</b>	
<b>Structures et formes urbaines : concepts, typologies, données disponibles et état de l'art .....</b>	<b>5</b>
Introduction de la partie I .....	7
<b>CHAPITRE 1 : Structure spatiale urbaine et morphologie urbaine : définitions et concepts .....</b>	<b>9</b>
Introduction . . . . .	10
1.1 Formes urbaines et morphologie urbaine . . . . .	10
1.2 La trame grise . . . . .	11
1.2.1 La notion de tissu urbain . . . . .	12
1.2.1.1 Éléments constitutifs des tissus . . . . .	13
1.2.1.2 Formes des tissus, évolutions et tailles . . . . .	14
1.2.2 Intérêts de l'étude des tissus urbains . . . . .	16
1.2.3 Typologies grises . . . . .	20
1.3 La trame verte . . . . .	25
1.3.1 La prise en compte de la végétation en urbanisme . . . . .	25
1.3.2 La notion de trame verte . . . . .	27
1.3.3 Multifonctionnalité de la végétation urbaine . . . . .	30
1.3.4 Typologies vertes . . . . .	32
Conclusion. . . . .	34
<b>CHAPITRE 2 : Données géographiques et méthodes existantes pour la cartographie de la structure urbaine .....</b>	<b>37</b>
Introduction . . . . .	38
2.1 Les données géographiques . . . . .	38
2.1.1 Types de données géographiques . . . . .	38
2.1.2 Évolution de l'offre en imagerie optique . . . . .	39
2.1.3 Notion d'échelle et lien avec la résolution des images . . . . .	41
2.2 Sélection des sites d'études et données disponibles . . . . .	43
2.2.1 Sites d'études . . . . .	43

2.2.2 Bases de données géographiques disponibles . . . . .	45
2.2.2.1 Bases de données géographiques spatialement continues . . . . .	46
2.2.2.2 Bases de données géographiques non continues spatialement . . . . .	50
2.3 Cartographie de la structure urbaine : aspects méthodologiques . . . . .	54
2.3.1 Évolution des approches pour l'analyse d'image : du pixel à l'objet . . . . .	55
2.3.2 Cartographie des éléments verts . . . . .	57
2.3.3 Cartographie des tissus urbains . . . . .	61
2.3.3.1 Unités d'analyse pour la cartographie des tissus urbains . . . . .	63
2.3.3.2 La prise en compte de la végétation pour la cartographie des tissus . . . . .	68
Conclusion. . . . .	70
Conclusion de la partie I et positionnement de recherche.....	71
<b>PARTIE II</b>	
<b>Extraction de la végétation urbaine à partir d'images satellites à Très Haute</b>	
<b>Résolution Spatiale</b> .....	77
Introduction de la partie II.....	79
<b>CHAPITRE 3 : Sélection de la segmentation optimale pour l'extraction des éléments</b>	
<b>verts urbains</b> .....	83
Introduction . . . . .	84
3.1 Segmentation de l'image et évaluation : choix des méthodes. . . . .	84
3.1.1 Choix des méthodes de segmentation . . . . .	85
3.1.1.1 La segmentation Multi-Resolution . . . . .	86
3.1.1.2 La segmentation Mean Shift. . . . .	86
3.1.2 Évaluation de la qualité des segmentations . . . . .	87
3.1.2.1 Méthodes d'évaluation non supervisées des segmentations . . . . .	88
3.1.2.2 Méthodes d'évaluation supervisées des segmentations . . . . .	89
3.1.2.3 Méthodes d'évaluation des segmentations par classification . . . . .	90
3.2 Sélection des zones tests et données . . . . .	92
3.2.1 Les images Pléiades disponibles . . . . .	92
3.2.2 Choix des sites tests. . . . .	93
3.2.3 Prétraitements des images Pléiades et présentation des attributs calculés . . . . .	94
3.2.4 Créations des données d'entraînement et de validation . . . . .	95
3.3 Sélection des segmentations adaptées à l'extraction de la végétation urbaine . . . . .	97
3.3.1 Présentation des segmentations réalisées . . . . .	97
3.3.2 Évaluation des segmentations pour la cartographie des éléments verts. . . . .	99
3.3.2.1 Sélection de la segmentation optimale par une approche non super-	
visée . . . . .	99
3.3.2.2 Sélection de la segmentation optimale par une approche supervisée . . . . .	101
3.3.2.3 Sélection de la segmentation optimale par classification . . . . .	103
3.3.3 Comparaison des méthodes d'évaluation de segmentation . . . . .	105
Conclusion. . . . .	108
<b>CHAPITRE 4 : Comparaison de stratégies d'échantillonnage pour la classification</b>	
<b>d'images THRS</b> .....	111
Introduction . . . . .	112

4.1	La sélection d'échantillons pour la classification d'images . . . . .	112
4.2	Méthodes proposées pour la sélection d'échantillons d'apprentissage . . . . .	116
4.2.1	Apprentissage actif spatial multi-classes (MCSAL) . . . . .	118
4.2.2	Apprentissage actif spatial multi-classes prenant en compte la diversité (MCSAL <sub>D</sub> ) . . . . .	118
4.2.3	Apprentissage actif spatial multi-classes stratifié (MCSAL <sub>S</sub> ) . . . . .	119
4.2.4	Échantillonnage aléatoire stratifié (SRGS) . . . . .	120
4.3	Préparation des données. . . . .	120
4.4	Protocole de tests . . . . .	121
4.5	Évaluation des méthodes d'apprentissage actif proposées. . . . .	123
4.5.1	Effet de la normalisation de l'entropie . . . . .	123
4.5.2	Sensibilité des paramètres du critère de diversité . . . . .	124
4.5.3	Choix de la taille des régions et comparaison des méthodes d'apprentis- sage actif avec l'aléatoire stratifié . . . . .	126
4.5.4	Apport des données auxiliaires pour l'amélioration des performances . . .	130
4.5.5	Évaluation des erreurs selon l'aire et la pureté de la classe . . . . .	131
	Conclusion. . . . .	135
<b>CHAPITRE 5 : Généralisation de l'approche orientée-objet par apprentissage actif pour la classification de la végétation.....</b>		
	Introduction . . . . .	137
5.1	Problèmes liés à la taille des images et solutions utilisées. . . . .	138
5.1.1	Tuilage des zones d'étude . . . . .	139
5.1.2	Sélection de nouveaux exemples d'entraînement par apprentissage actif .	140
5.2	Sélection des méthodes de classification et de leurs paramètres . . . . .	141
5.2.1	Choix des algorithmes de classification . . . . .	141
5.2.2	Sélection des paramètres des méthodes de classification . . . . .	143
5.3	Classifications finales de la végétation arborée et herbacée . . . . .	144
5.3.1	Sélection des mailles pour la validation des classifications . . . . .	144
5.3.2	Évaluation des méthodes de classification par les mailles de validation .	145
5.3.3	Protocole de post-traitements et première évaluation des résultats . . .	148
5.4	Évaluation des bases de données végétation . . . . .	151
5.4.1	Évaluation des classifications selon les strates d'urbanisation . . . . .	152
5.4.2	Évaluation des classifications selon l'utilisation du sol . . . . .	152
5.4.3	Évaluation à l'échelle des objets . . . . .	156
	Conclusion. . . . .	162
	Conclusion de la partie II .....	163
<b>PARTIE III</b>		
<b>Cartographie et analyse quantitative de la structure urbaine à partir de don- nées multi-sources .....</b>		
	Introduction de la partie III .....	167
<b>CHAPITRE 6 : Création d'une base de données d'éléments urbains multi-sources ..</b>		
	Introduction . . . . .	171
	Introduction . . . . .	172

6.1	Vers une base de données d'occupation du sol à grande échelle . . . . .	172
6.1.1	Choix des données pour cartographier la trame grise . . . . .	172
6.1.2	Choix des données pour cartographier les trames verte et bleue . . . . .	174
6.1.3	Les données INSEE sur la population . . . . .	175
6.2	La caractérisation de la structure urbaine . . . . .	176
6.2.1	Indicateurs issus de l'urbanisme . . . . .	176
6.2.2	Indicateurs issus de l'écologie du paysage . . . . .	180
6.2.3	Autres indicateurs proposés dans cette recherche . . . . .	182
6.2.3.1	Indicateurs d'accessibilité et de visibilité. . . . .	182
6.2.3.2	Indicateurs sociodémographiques . . . . .	183
6.2.3.3	Indicateurs de centralité . . . . .	183
6.3	Unités testées pour caractériser la structure urbaine. . . . .	188
6.3.1	Découpage par carroyage . . . . .	188
6.3.2	Découpage à la parcelle morphologique. . . . .	189
6.3.3	Découpage à l'îlot . . . . .	190
6.4	Sélection de l'unité d'analyse pour la cartographie de la structure urbaine . .	193
6.4.1	Sélection des critères d'évaluation . . . . .	193
6.4.1.1	Critères d'évaluation existants pour le choix d'une unité d'analyse .	193
6.4.1.2	Critères d'évaluation retenus . . . . .	194
6.4.2	Évaluation des mailles pour l'analyse de la structure urbaine . . . . .	195
6.5	Proposition d'un nouveau maillage pour la cartographie de la structure ur- baine . . . . .	198
	Conclusion. . . . .	202
	<b>CHAPITRE 7 : Application de méthodes d'extraction de connaissances.....</b>	<b>203</b>
	Introduction . . . . .	204
7.1	Analyse des relations entre les variables . . . . .	205
7.2	Méthodes de regroupement des données . . . . .	208
7.2.1	Sélection d'une méthode de clustering . . . . .	209
7.2.2	Le problème du nombre de clusters . . . . .	215
7.3	Classification de la structure urbaine de Strasbourg . . . . .	216
7.3.1	Classification des indicateurs et interprétation des variables synthétiques	216
7.3.2	Classification non-supervisée des mailles fondée sur les variables synthé- tiques . . . . .	225
7.3.2.1	Choix du nombre de classes. . . . .	225
7.3.2.2	Analyse des types urbains et non-urbains en deux niveaux de classi- fication . . . . .	227
7.3.3	Évaluation de la classification de Strasbourg . . . . .	238
7.3.3.1	Évaluation à partir de la BDOCS . . . . .	238
7.3.3.2	Évaluation à partir de la base Urban Atlas . . . . .	242
7.4	Classification de la structure urbaine de Rennes . . . . .	244
7.4.1	Application des paramètres des classifications de Strasbourg sur Rennes.	244
7.4.2	Évaluation de la classification des mailles de Rennes. . . . .	245
	Conclusion. . . . .	250

Conclusion de la partie III .....	253
Conclusion générale et perspectives de recherche .....	257
Références bibliographiques .....	263
Liste des figures .....	298
Liste des tableaux .....	300
Annexes .....	301
Table des matières .....	343



# Apport des images satellites à Très Haute Résolution Spatiale couplées à des données géographiques multi-sources pour l'analyse des espaces urbains

## Résumé

Les villes sont confrontées à de nombreuses problématiques environnementales. Leurs gestionnaires ont besoin d'outils et d'une bonne connaissance de leur territoire. Un objectif est de mieux comprendre comment s'articulent les trames grise et verte pour les analyser et les représenter. Il s'agit aussi de proposer une méthodologie pour cartographier la structure urbaine à l'échelle des tissus en tenant compte de ces trames. Les bases de données existantes ne cartographient pas la végétation de manière exhaustive. Ainsi la première étape est d'extraire la végétation arborée et herbacée à partir d'images satellites Pléiades par une analyse orientée-objet et une classification par apprentissage actif. Sur la base de ces classifications et de données multi-sources, la cartographie des tissus se base sur une démarche d'extraction de connaissances à partir d'indicateurs issus de l'urbanisme et de l'écologie du paysage. Cette méthodologie est construite sur Strasbourg puis appliquée à Rennes.

**Mots-clés** : cartographie de la structure urbaine, trames grise et verte, image Pléiades THRS, analyse d'image orientée-objet, apprentissage actif, extraction de connaissances dans les bases de données, données multi-sources

## Abstract

Climate change presents cities with significant environmental challenges. Urban planners need decision-making tools and a better knowledge of their territory. One objective is to better understand the link between the grey and the green infrastructures in order to analyse and represent them. The second objective is to propose a methodology to map the urban structure at urban fabric scale taking into account the grey and green infrastructures. In current databases, vegetation is not mapped in an exhaustive way. Therefore the first step is to extract tree and grass vegetation using Pléiades satellite images using an object-based image analysis and an active learning classification. Based on those classifications and multi-sources data, an approach based on knowledge discovery in databases is proposed. It is focused on set of indicators mostly coming from urbanism and landscape ecology. The methodology is built on Strasbourg and applied on Rennes to validate and check its reproducibility.

**Keywords:** urban structure mapping, grey and green infrastructures, VHSR Pléiades image, object based image analysis, active learning, knowledge discovery in databases, multi-sources data