Université catholique de Louvain
Faculté des Sciences, École de Géographie
Center for Operations Research and Econometrics

Morphological delineation and characterisation of urban spaces

---

Doctoral Dissertation presented by

Gaëtan Montero Redondo

in fulfillment of the requirements
for the degree of Doctor in Sciences

Thesis committee:

Prof.	Isabelle Thomas (Supervisor)	Université catholique de Louvain
Prof.	Marie-Laurence De Keersmaecker (Chair)	Université catholique de Louvain
Dr.	Justin Delloye (Secretary)	Université catholique de Louvain
Dr.	Cécile Tannier	Université Bourgogne Franche-Comté
Prof.	Jacques Teller	Université de Liège
Dr.	Giovanni Fusco	Université Côte d'Azur

Louvain-la-Neuve, Belgium
March 2023

Remerciements

Ecrire une thèse, c'est prendre part à une expérience incroyable. Chaque jour est fait de découvertes, d'émerveillements, de joies, de défis mais aussi, parfois, de peines et de découragements. Toutes ces émotions, j'ai eu la grande chance de les vivre sous le regard bienveillant de ma promotrice, la Professeure Isabelle Thomas. Merci, Isabelle, de m'avoir offert la chance d'intégrer votre équipe de recherche. Merci de m'avoir embarqué dans l'aventure. Plus qu'écrire une thèse, vous m'avez appris à me questionner, à puiser l'essentiel, à me battre, à faire preuve de rigueur, à communiquer, et tant d'autres choses encore. Votre enthousiasme pour la géographie (#papillon) et votre combativité (#lion) résonneront en moi tout le reste de ma carrière, c'est certain !

Un travail de recherche ne s'effectue heureusement pas seul. Outre ma promotrice, j'ai eu la chance de participer à de nombreuses collaborations. Un merci tout particulier à Cécile Tannier qui a su m'encadrer avec beaucoup de patiente. Merci Cécile pour votre créativité et votre bienveillance. Merci également à Geoffrey Caruso et Mohamed Hilal avec qui le travail était très stimulant. Un doctorat est aussi fait de projets non aboutis. Merci à Laura Merla d'avoir rêvé avec moi et d'avoir éveillé ma curiosité vers des questions plus sociologiques. Merci aussi aux membres qui ont rejoint mon jury, Marie-Laurence De Keersmaecker, Jacques Teller, Giovanni Fusco et Justin Delloye. Merci pour votre temps et vos conseils avisés.

Impossible de rédiger des remerciements sans saluer les formidables collègues que j'ai eu la chance de côtoyer durant mon parcours. L'équipe Géo-en-Core est pour beaucoup dans l'accomplissement de cette thèse. Cela a été un réel plaisir de travailler au sein de cette équipe. Encouragements, échanges, bonne humeur et challenges ont été notre quotidien. Jonathan, Adeline, Olivier, Justin, Arnaud, Sonia et Madeleine, de tout cœur, mille

mercis !

Pas de thèse sans un minimum de confort financier et matériel. Merci à Innoviris d'avoir supporté mon entrée dans le monde de la recherche. Merci au CORE pour l'accueil et le support logistique. Et grand merci à l'Ecole de Géographie de m'avoir permis d'exercer le magnifique métier d'assistant.

Bien que mes travaux de recherche et d'enseignement ne soient pas directement liés, je tiens ici à remercier les différentes générations d'étudiants que j'ai eu la chance d'encadrer. Que ce soit de l'initiation à la géographie avec des bacs 1, des questions très méthodologiques et statistiques avec des bacs 2 et 3 ou des questions plus approfondies de recherche avec des masters, j'ai pris énormément de plaisir à les encadrer. Merci pour leur enthousiasme durant mes TPs.

Un tout grand merci enfin à ma famille et mes amis pour leur amour. Pas toujours facile d'expliquer à quoi j'occupais mes journées. Merci à vous d'avoir essayé de comprendre. Merci pour vos questions, merci d'y avoir cru avec moi.

Contents

1 Introduction **1**
- 1.1 Measuring cities . . . 2
- 1.2 Morphological delineation . . . 5
- 1.3 Morphological characterisation . . . 9
- 1.4 Objectives and structure of the PhD thesis . . . 11

2 Complexity of the spatial structure of cities: the example of Brussels-Capital Region **13**
- 2.1 Partitioning an urban area . . . 14
- 2.2 Functions and interactions of/between places . . . 16
  - 2.2.1 Urban factorial ecology: a benchmark . . . 16
  - 2.2.2 Community detection in ICT data: phone basins . . . 18
- 2.3 Morphometrics of built-up footprint . . . 21
  - 2.3.1 Built-up density . . . 21
  - 2.3.2 From “natural cities” to “natural urban clusters” . . . 22
- 2.4 Conclusion and comments . . . 24

3 Fractal geometry in geography: a general background **27**
- 3.1 Fractal geometry and fractal dimension . . . 28
- 3.2 Measuring the fractal dimension . . . 31

4 Analysis of box-counting and correlation dimensions: calculation parameters and morphological meaning **33**
- 4.1 Introduction . . . 34
- 4.2 Les tissus bâtis étudiés . . . 35
  - 4.2.1 Les tissus bâtis réels . . . 35
  - 4.2.2 Les tissus bâtis théoriques . . . 38
  - 4.2.3 Étendue des zones d’étude . . . 42
- 4.3 Choix des valeurs des paramètres . . . 43
  - 4.3.1 Le format de l’image et sa résolution . . . 44

4.3.2 Les caractéristiques des fenêtres de comptage 45
4.3.3 L'ajustement de la courbe théorique à la courbe empirique 53
4.3.4 Récapitulatif des valeurs de paramètres choisies pour les analyses 54
4.4 Comparaison des dimensions de boîte et de corrélation des cinquante tissus bâtis 55
4.4.1 Des classifications pour observer les hiérarchies 56
4.4.2 Dimensions et caractéristiques morphologiques 58
4.5 Conclusion 64

5 Delineation of cities based on scaling properties of urban patterns: a comparison of three methods 65
5.1 Introduction 66
5.2 Description and comparison of the three methods with theoretical urban patterns 67
5.2.1 *Natural cities* 67
5.2.2 *MorphoLim* 70
5.2.3 *Hierarchical Percolation* 74
5.2.4 *Comparing the methods* 75
5.3 Delineation of the morphological agglomeration of Brussels . . 78
5.3.1 Data and study area 78
5.3.2 Natural Cities method applied to Brussels 80
5.3.3 *MorphoLim* method applied to Brussels 83
5.3.4 *Hierarchical Percolation* method applied to Brussels . . 85
5.3.5 Discussion 88
5.4 Conclusion 91

6 Characterisation of urban agglomerations delineated by *MorphoLim* 95
6.1 Introduction 96
6.2 Les 82 aires urbains étudiées 97
6.3 Méthodologie 99
6.3.1 Délimitation des agglomérations morphologiques 99
6.3.2 Caractéristiques des tissus bâtis des 82 villes étudiées . 100
6.3.3 Caractéristiques morphologiques du périmètre des 82 villes étudiées 101
6.4 Résultats 103
6.4.1 Délimitation des agglomérations morphologiques 103
6.4.2 Morphologie des agglomérations et de leurs périphéries . 106
6.4.3 Caractéristiques morphologiques des périmètres des villes 110
6.4.4 Typologie à partir des indices morphologiques caractérisant les tissu bâti et leur périmètre 117
6.5 Conclusion 121

7 Delineation of a « diffuse city » at ground level (2D) and according to the number of floors (3D) 125
7.1 Introduction . . . 126
7.2 Delineation issues and proposed alternatives . . . 128
7.3 Materials and Methods . . . 130
7.3.1 Study area and data input . . . 130
7.3.2 Methods . . . 132
7.4 Results and discussion . . . 133
7.4.1 Delineation at ground level (2D) . . . 133
7.4.2 Delineation according to the number of floors (3D): the city of the upper districts . . . 136
7.5 Conclusion . . . 140

8 A partition-free spatial clustering that preserves topology: application to built-up density 141
8.1 Introduction . . . 142
8.2 Methodological advances . . . 143
8.3 Materials and Methods . . . 147
8.3.1 Data input and study area . . . 147
8.3.2 Methods . . . 147
8.4 Results . . . 152
8.4.1 The Minimum Spanning Tree . . . 152
8.4.2 Moran scatterplots and their outliers . . . 153
8.4.3 Clusters of buildings . . . 155
8.4.4 Density of buildings . . . 157
8.5 Discussion . . . 161
8.5.1 Thematic contribution . . . 161
8.5.2 Methodological limits . . . 162
8.6 Conclusion . . . 164

9 Conclusion 165
9.1 Summary of our findings . . . 166
9.1.1 General objectives of this thesis . . . 166
9.1.2 Delineation and Characterisation . . . 166
9.2 Critical assessment and avenues for further research . . . 168
9.2.1 A vast field of research . . . 168
9.2.2 Form to function? . . . 170
9.2.3 Understanding is not predicting . . . 172
9.3 Concluding remarks . . . 173

A Appendices of chapter 2 175
A.1 48 variables used in the factorial ecology . . . 175

B Appendices of chapter 5 **179**
B.1 Natural cities: head/tail breaks tables 179

C Appendices of chapter 8 **181**
C.1 Chunking of database 181
C.2 Calculation of the Spatial Lag 182
C.3 Sensitivity to condition (1) 183
C.4 The 26 462 clusters for Belgium according to the SDHC 186
C.5 The topology-based built density index computed for Belgium . 187

Bibliography **188**

List of Figures **210**

List of Tables **216**

List of publications

The chapters of this thesis are based on the following publications:

- **Chapter 2**: Adam, Arnaud, Gaetan Montero, Olivier Finance, Ann Verhetsel, and Isabelle Thomas (2019) "Revisiting urban models with information and communication technology data? Some examples from Brussels," in *Digital Social Networks and Travel Behaviour in Urban Environments*: Routledge.

→ Gaëtan Montero is responsible for three of the four characterisations in the chapter. The order of the authors follows the alphabetical order.

- **Chapter 4**: Tannier, Cécile, Gaetan Montero, François Sémécurbe, and Isabelle Thomas (à paraître) "Méthodes d'analyse fractale pour caractériser la distribution spatiale des implantations humaines," in *Analyse et modélisation fractale de la distribution spatiale des implantations humaines, Cécile Tannier (dir.), Encyclopédie Sciences, domaine Géographie et démographie, ISTE éditions*.

→ The book chapter is in two parts. Only the second part, written by Gaëtan Montero, is published in this thesis. This chapter of the thesis is published in French (following the request of the publisher of the book from which it is taken). A translation of the encyclopaedia into English will be published.

- **Chapter 5**: Montero, Gaëtan, Cécile Tannier, and Isabelle Thomas (2021) "Delineation of cities based on scaling properties of urban patterns: a comparison of three methods," *International Journal of Geographical Information Science*, Vol. 35, pp. 919–947, Number: 5.

- **Chapter 6**: Montero, Gaetan, Cécile Tannier, and Isabelle Thomas (à paraître) "Délimitation morphologique des agglomérations urbaines : comparaison de méthodes fractales et non fractales (chap. 5)," in *Analyse et modélisation fractale de la distribution spatiale des implantations humaines, Cécile Tannier (dir.), Encyclopédie Sciences, domaine Géographie et démographie, ISTE éditions*.

→ This chapter of the thesis is published in French (following the request of the publisher of the book from which it is taken). A translation of the encyclopaedia into English will be published.

- **Chapter 7**: Salom-Carrasco, Julia, Isabelle Thomas, and Gaëtan Montero (2021) "Delimitación y caracterización morfométrica del área metropolitana de Valencia," *Cybergeo*.

→ Gaëtan Montero is responsible for the methodological ideas and development.

- **Chapter 8**: Montero, Gaëtan, Geoffrey Caruso, Mohamed Hilal, and Isabelle Thomas (2022) "A partition-free spatial clustering that preserves topology: application to built-up density," *Journal of Geographical Systems*.

Main acronyms used in this thesis

BCR	Brussels-Capital Region
BSU	Basic spatial unit
D_B	Box-counting dimension
D_C	Correlation dimension
D_S	Similarity dimension
D_s	Fractal dimension for built-up surfaces
D_b	Fractal dimension for a boundary
AM	Agglomération morphologique
MAUP	Modifiable areal unit problem
TIN	Triangulated Irregular Networks
AU	Aire urbaine

"La terre nous en apprend plus long sur nous que les livres. Parce qu'elle nous résiste. L'homme se découvre quand il se mesure avec l'obstacle. Mais, pour l'atteindre, il lui faut un outil. Il lui faut un rabot, ou une charrue. Le paysan, dans son labour, arrache peu à peu quelques secrets à la nature, et la vérité qu'il dégage est universelle. De même l'avion, l'outil des lignes aériennes, mêle l'homme à tous les vieux problèmes"

Antoine de Saint-Exupéry, *Terre des hommes*, Gallimard, 1939, Paris, France.

1

Introduction

The main objective of this thesis is to contribute to the understanding of urban spaces. In particular, the aim is to critically analyse and develop measurements used to morphologically characterise cities. This first chapter introduces the context and main objective of the present thesis. Section 1.1 starts by explaining the motivation behind our work. It sets the scene and allows us to situate our research in the existing scholarly literature relating to the field of urban form. Section 1.2 introduces our work in relation to questions concerning the delineation of urban spaces. Section 1.3 addresses the possible strategies we have employed to characterise urban spaces morphologically. Finally, we return to our objectives and situate them in the plan of our thesis in Section 1.4.

1.1 Measuring cities

> *"We commit ourselves to promoting the development of urban spatial frameworks, including urban planning and design instruments that support sustainable management and use of natural resources and land, appropriate compactness and density, polycentrism and mixed uses, through infill or planned urban extension strategies, as applicable, to trigger economies of scale and agglomeration, strengthen food system planning and enhance resource efficiency, urban resilience and environmental sustainability".*

This excerpt is point 52 of the New Urban Agenda (*Habitat III*) endorsed by member countries of the United Nations at the *UN Conference on Housing and Sustainable Development* in Quito (Ecuador) in 2016 (Nations, 2016). It is a challenge for a geographer to develop instruments that support the sustainable development of cities in order to obtain all possible advantages without reducing the benefits for future generations.

But how can such instruments be developed? *"Logically, one needs to understand what cities are made of, how they come about and function, what they mean to people, and so on, in order to design "good" cities"* (Moudon, 1992). To develop relevant urban planning instruments, it is therefore important to understand what a city is. It is widely recognised that this task is really not trivial because a city is a complex system that can be seen as the superposition and interaction of social, physical and virtual networks (Rybski and González, 2022). That is a lot to measure and understand.

Our understanding of this multiplayer spatial organisation therefore depends, among other things, on the data and methods used to measure it. Which layer is measured and how does interact with others ? This is illustrated in the second chapter of this thesis. Five years ago, our research team sought to partition the Brussels-Capital Region (Belgium) in order to understand its spatial organisation in relation to the major models of urban geography (see Chapter 2). We used four different methods and obtained almost as many urban models. This work provided real motivation for those who followed because it raised the question of the meaning of the measurements used. If the conclusion of a study depends on the measurements used, it seems essential to us to understand the meaning of these measurements. What are we really measuring? This question should be asked in every study in order to strengthen the conclusions. For example, if we consider the well-known study of Newman and Kenworthy (1989), in which the authors observe that population density is inversely proportional to

the annual transport gasoline consumption per capita, it follows from this analysis that in order to save energy for transport, it would be wise to make cities denser[1]. Without particularly questioning the results of this study, we can nevertheless ask ourselves the question of the meaning of the measurement. Would the results, and hence the conclusions, have been the same if a different definition of population density, or a different delimitation of cities, had been used?

When choosing data to measure the city, researchers generally opt methodologically for either a functional approach or a morphological approach. The functional approach is based on variables characterising either places and/or people living in these places (X_i). This is, for example, the case with Dujardin et al. (2008), who study the spatial organisation of Brussels on the basis of census data of its residents. The functional approach can also be based on variables characterising exchanges (w_{ij}) between places. One example is the study by Adam et al. (2017), which also analyses the spatial structure of Brussels but using residential migration data, data on commuting to and from the workplace and phone call data. The morphological approach is based on physical data (buildings, roads, etc.). As with the functional approach, it is the physical characteristics themselves that are studied (Berghauser Pont and Haupt (2005), for example, characterise urban forms based on building's footprints) or their interrelationship (Caruso et al. (2017), for example, use the Euclidean distance between buildings to measure urban forms). In the first approach, methods focus on the functions of the space. In the second approach, it is the physical urban forms that are measured.

The interplay between form and function is central to urban science and to understanding how cities are made up: *'from form we infer processes that create the structures we see in cities, thus enabling us to build models of these processes, that in turn will simulate forms'* (Batty, 2013, p.79). The physical part of the city, essentially buildings, roads and plots, reveals the presence of inhabitants, activities and hence movements (flows), which themselves relate to planning and normative issues, such as what is the "good" city form (Lynch, 1984). In this thesis, we choose to focus on the second approach: the morphological approach. Like many other researchers from various disciplines (geographers, architects, urban planners, etc.) working on urban morphology, we assume that the city can be understood through its physical form (Moudon, 1997). The physical city (built city) serves, indeed, as a basis for other cities such as the consumption city, the employment city, etc. (Parr,

[1]Other studies have since shown the advantages and disadvantages of urban densification on transport but also on the environment, the economy, the well-being, etc (for a complete review see Berghauser Pont et al. (2021)).

2007). This does not mean that the functional approach is useless. On the contrary, it is also necessary to understand the interactions between the functions and the forms of an environment. Nevertheless, we choose to focus our attention on some morphological aspects.

As we are focusing on the morphological features of cities, our work is part of the *"urban form"* research field. Indeed, this area of research can be defined as the study of *"morphological attributes of an urban area at all scales"* (Williams et al., 2000). The study of urban form is then the study of features such as building materials at a fine scale or the study of housing type, street type and their spatial arrangement at a broader scale, or even the study of urban settlement type such as a market town, central business district and suburbs at a regional scale (Dempsey et al., 2009). Our work is in line with this approach. In particular, Moudon (1997) wrote that morphological analyses are based on three principles:

- Buildings, streets and plots are the fundamental physical elements defining an urban form
- Urban form can be understood at different levels of resolution (buildings/lot, street/block, the city, the region)
- Urban form should be understood historically since the physical elements are continuously being transformed and replaced

In our work, we follow the first two principles listed. We measure urban forms using data such as buildings and road intersections. And we perform different analyses at different scales, such as regional scale and more local intra-urban scale. We will not exploit the third principle listed by Moudon (1997). We are aware that urban form is the result of processes that require a diachronic approach. However, we use a synchronic approach (study of the urban form at a given moment) because, like Lévy (2005), we are convinced that a better knowledge of the urban form as a object is as indispensable as the understanding of the processes that lead to the object. We work on two principles linked to the term 'urban form': (1) the form of the urban layout, which refers to the shape of the city, and (2) the form of the urban fabric through its spatial structures (Lévy, 2005).

In order to measure and understand an urban space morphologically, it is useful to work on two aspects. The first is the delineation of the urban space; the second is the actual characterisation of the delimited space. Let us note immediately that the two aspects are inseparable - *"a chicken and egg problem"* to quote Isard (1975). The delineation of a space impacts the characterisation of the space itself (Goodchild, 2016). Moreover the definition of a region (i.e.

its delimitation) depends on the elements we use to characterise it (Isard, 1956). For the sake of clarity, we dissociate delineation from characterisation in sections 1.2 and 1.3 while throughout the thesis both aspects are tightly related.

1.2 Morphological delineation

Defining city boundaries is invariably an enthralling challenge for planners and geographers. Planners need to know where cities start and end in order to govern and manage them efficiently (count populations, levy taxes, organise transport systems, etc.); geographers need to know the city limits in order to study the growth of cities, characterise land-use changes, compare cities (their area and population), measure the impacts of different boundaries on social, economic and ecological indicator values, etc (see for example Tayyebi et al., 2011; Chakraborti et al., 2018; Wang et al., 2021; de Olde and Oosterlynck, 2021; Haisch and Müller, 2013; Uchiyama and Mori, 2017; Thomas et al., 2018; Han et al., 2009; Ding et al., 1999). Sotomayor-Gómez and Samaniego (2020) quote a United Nations World Cities report, which indicates that countries' urban proportions face multiple non-comparable definitions. As an example, the city of Toronto in Canada may be considered to have between 2.6 and 5.1 million inhabitants, depending on which definition of city is used to calculate its size. According to Sotomayor-Gómez and Samaniego (2020), such variable figures reflect annual growth rates ranging from 0.9% to 1.8%, which may clearly lead to significant complications for efficient planning efforts. Moreover, in most disciplines, assessing the effect of system boundaries is important to provide certainty and robustness in model findings. This has long been an issue for geographers, from either a theoretical or an empirical perspective, as explained in Thomas et al. (2018).

All these challenges encourage us in our search for the delimitation of urban spaces. Our approach will be exclusively dual. Indeed, we will dissociate the urban space from the rest of space. We are aware that the boundary between urban and rural spaces is becoming fuzzier and that some authors/planners seek to understand these fuzzy spaces somewhere between town and countryside (see for example Steel et al., 2013). However, we will not enter into this field of research and remain focused on the delimitation of an urban space as a closed entity. We believe that this approach remains useful for many aspects of the management of urban spaces (limiting road traffic within a closed area, study the growth of cities, etc.).

Even though the "bounding problem" was clearly defined by Haggett (1972) early on, nowadays, there is no consensus on the most appropriate way to delineate urban agglomerations either in terms of methods or in terms of criteria or thresholds, and hence results do not converge (see for examples and reviews Cottineau et al., 2019; Ferreira et al., 2010; Rozenfeld et al., 2011; Thomas et al., 2018).

Many studies delineate urban spaces using a functional approach (see for example Rybnikova et al., 2021; Dong et al., 2022; Kauffmann, 2012; Guérois et al., 2014; Chen and Gar-On Yeh, 2022; Fujishima et al., 2020; Blondel et al., 2010). As mentioned in Section 1.1, we will focus, in this thesis, on the morphological approach (based on physical data) to delineate cities. We note that the criteria used in morphological delineation are often readily comparable internationally and more objective (Weber, 2001; Parr, 2007; Fang and Yu, 2017; Goerlich Gisbert et al., 2017). In comparison, functional approaches are less inclusive and capture only part of the urban functioning, depending on the variables considered (Thomas et al., 2018).

Methods used to delineate cities by morphological criteria can be grouped into four categories:

1. Density measures – a density threshold is set, above which a place is considered "urban". This threshold can be applied on discrete maps, where elementary spatial units are administrative units or cells of a grid, or on smoothed maps using a kernel density estimation. Different physical characteristics can be considered in the density calculation. For example, Borruso (2003) calculates the density of road intersections in the Trieste region of Italy through a grid. Zhou (2015) applies a grid directly to New Zealand's road networks and performs a smoothing (kernel function) to measure the density of the roads. Tan and Huang (2022) and Harig et al. (2021) calculate the density of buildings in China and Germany, respectively.

2. Classification of remote sensing images or aerial photographs, so that similar spatial units (pixels) can be grouped together. Usually a criterion of contiguity – generally involving a distance threshold – is added. Identical pixels can be clustered into groups using more or less complex methods related simply to distance (Nicolau and Cavaco, 2018) or according to statistical criteria such as maximum likelihood (Henderson et al., 2003). Clustering can also be done at different scales of analysis as in Zhou et al. (2018). Sometimes it is not the commonalities between pixels that are used to delineate urban areas but rather their differences. For example, Hu et al. (2015) calculate an entropy index to draw a boundary where changes in the

satellite image are strongest. Similarly, Ban et al. (2021) search for breaking points.

3. Clustering methods that group elements (usually buildings) according to their proximity or other predefined common characteristics. Researchers group elements directly according to different criteria and indices (Usui, 2019) or they group them via the prior use of graphs (Deng et al., 2011; Zhong and Duckham, 2016). For the former, we can cite Arribas-Bel et al. (2019), who use a density criterion. Chaudhry and Mackaness (2008) calculate the *citiness index*, which is a function of the distances between buildings and their surfaces. Another frequently used method is the City Clustering Algorithm (CCA) (Rozenfeld et al., 2008, 2011). Researchers using graphs seek to identify clusters based on the distances between measured physical elements. Caruso et al. (2017) use a distance criterion (an edge between two building centroids must not be less than 200 m) and combine it with the calculation of a LISA (Local Indicators of Spatial Association). Liu et al. (2013) use similarity criteria between physical elements in the graph. Liu et al. (2012) use the Gestalt principle to determine whether a point on the graph is similar to or different from its neighbours in terms of distance.

4. Analyses of a statistical distribution across scales – for example, distribution of distances between buildings, street blocks, road intersections or points of interest – and identification of a threshold in this scaling distribution. For example, Jiang and Liu (2012), Jiang et al. (2016), Jiang (2013, 2015) calculate the average of a rank/size distribution of the edges' length of a Triangulated Irregular Network (TIN) connecting road intersections. Tannier et al. (2011) and Tannier and Thomas (2013) search for the maximum curvature point of a curve showing the number of built-up aggregates with increasing dilation distances. Arcaute et al. (2014) and Masucci et al. (2015) search for the percolation threshold in a curve describing the number of road intersections in the cluster that contains the most road intersections at different dilation distances.

The methods of the first two categories are mainly based on aggregated data, i.e. the elements considered (pixels, buildings, etc.) are aggregated into basic spatial units (BSUs). These BSUs are generally administrative entities (districts, municipalities, etc.) or meshes of a regular grid. Because of the use of BSUs, the delimitations obtained with these methods are inevitably dependent on the level of aggregation adopted and on the shape and position of the elementary spatial units. These dependencies are well known to geographers as the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1983). To limit the impact of the MAUP, researchers sometimes perform

smoothing on the characteristics of BSUs via, for example, kernel-smoothed density estimates (Zhou, 2015; Ban et al., 2021; Bellefon et al., 2019). However, the smoothing functions are themselves dependent on parameters, namely the bandwidth and the kernel density. Zhou and Guo (2018) compared New-Zealand city boundaries obtained for different spatial resolutions of the input data (pixels of varying sizes) and different parameterisations of the smoothing function. They found that, most of the time, the most appropriate parameter value thresholds for the delineation of different cities are identical or very close to each other. In some cases, however, these threshold values differ significantly.

Morphological delineation methods based on clustering (third category of methods) or on the analysis of a statistical distribution across scales (fourth category) have the advantage of initially avoiding the MAUP related to the aggregation of elements (buildings, pixels, etc.) into BSUs. Another advantage of the methods in the fourth category over the third is that they avoid using any predefined quantified threshold (size, distance, built density, etc.) to detect crucial discontinuities in space. Indeed, clustering methods require making choices (aggregation mode, number of classes, distance threshold, etc.) a priori. These choices inevitably bias the results and complicate comparisons between different studies and different regions. On the contrary, with methods based on analysis of a statistical distribution across scales, the thresholds are identified a posteriori; they are specific to each area under study, and cities characterised, for example, by similar global densities may exhibit different thresholds. Avoiding ex ante thresholds is especially valuable, as it makes it possible to compare cities of varying sizes and shapes (Le Gléau et al., 1996).

Few papers broach the comparison of the morphological limits of cities resulting from different methods. For example, Cetinkaya et al. (2015) compare four delineation algorithms, all of which are applied to graphs (third category of methods). Borruso (2003) compares the results of a delineation obtained using road density via a grid or via the application of a kernel smoothing function (first category). The methods from the first, second and third categories all ultimately depend on the researchers' choices (level of aggregation, threshold). Is the same true for the methods in the fourth category? Finding an answer to this question is the research objective of Chapter 5 of this thesis. We seek to determine the impact of the MAUP and the extent of the study area on the delineation of morphological agglomerations for different methods based on the analysis of a statistical distribution across scales.

The shape of a boundary can be the result of internal city forces (economic, social, political, etc. forces that interact in a finite space) (Medda et al., 1998; Batty and Longley, 1994). In order to understand an urban space, we therefore want to morphologically characterise the shape of its boundaries. Are there any similarities or contradictions between the morphological characteristics of the boundaries drawn and the morphological characteristics of the (intra-urban) spaces they delimit? These questions are addressed in Chapter 6.

The exploration of the third dimension is our final contribution to the understanding of cities through their boundaries. In Chapters 5 and 6, delineations are made at ground level (building footprint - 2D). Chapter 7 is an application of the method used in Chapters 5 and 6 but where the height of the buildings is considered in the measurements.

1.3 Morphological characterisation

Two main strategies are generally used to measure and describe urban forms (Berghauser Pont, 2021): (1) the discrete areas strategy and (2) the network strategy.

In the discrete strategy, researchers measure morphometric characters such as size, shape, density, etc. (see Fleischmann et al., 2021b, for a review) of discrete elements (e.g. buildings) in more or less complex discrete areas. Indeed, leaving aside the problem of flows (e.g. Andrienko et al., 2010; Hurvitz et al., 2014), switching from physical structures to functions often leads researchers and planners to switch to areal objects, and count population or activities over a given area (buildings, parcels, grid cells, etc.). An a priori selected reference (basic) spatial unit (BSU) is then used to measure a density index or a more complex metric.

The BSUs used can be fixed beforehand (administrative areas, cell grid). A good example is the set of urban metrics computed over a grid proposed by Galster et al. (2001) to outcompete density while measuring sprawl. Each of their eight indices, not just the average density, requires a count of population or land use category over a set of exogenous grid cells, before being further aggregated over an urban region. In the same vein, Berghauser Pont and Haupt (2005) propose computing four indices (Floor Space Index, Ground Space Index, Open Space Ratio and Height) on buildings grouped by plots to detect typologies of plots in a Spacemate. Godoy-Shimizu et al. (2021) compute Floor Space Index and Ground Space Index on different sites (groups

of buildings in plots). In these examples, there is an aggregation process, and hence information loss, depending on the resolution and placement of the grid and the definition of the plots or depending on the original recording units (e.g. census tracts). These zonal and scale effects (*MAUP* (Openshaw, 1983)), inevitably bias measurements of urban form such as density (Zhang and Kukadia, 2005). Some authors then seek to define discrete zones (BSUs) specific to their study area and the characteristics of the objects measured. This is the case, for example, with Fleischmann et al. (2021a); Schirmer and Axhausen (2015); Araldi and Fusco (2019) who perform clustering, local spatial statistics and spatial smoothing within the topological constraints of building-based tessellation adjacencies or street network topology. In these three papers, BSUs are adapted in size and shape to the urban intensity; morphometrics characters are then computed on these BSUs.

The morphometric characteristics computed on discrete areas are then often further associated with the evaluation of the urban form in terms of, for example, liveability (Martino et al., 2021), waste production, traffic volume, water and energy consumption (Pauleit and Duhme, 2000), densification (Teller, 2021), heat island and the flow of air (Boyko and Cooper, 2011), the urban vitality (Bobkova et al., 2017). It is not merely the density of these structures that matters, but also their geometry at specific scales (Schirmer and Axhausen, 2015). Batty (2013, p.180) suggests that spatial interactions and the functioning of connections within cities *"need to be physically rooted in the detailed geometry of buildings"*. Similarly, for Gehl (1987, p.83), building density *"says nothing conclusive about whether human activities are adequately concentrated. The design of buildings in relation to relevant human dimensions is crucial"*. This is where the network strategy comes in.

To measure and describe urban form, the network strategy does not focus on discrete objects but on systems of objects. Researchers study the street network in particular, which involves the study of network structure, connectivity, centrality, hierarchy, etc. (Marshall et al., 2018). The Space Syntax movement, initiated by Hillier's seminal work (Hillier, 1996), is one of the precursors of this approach, with attention devoted to the relative position of lines. Study of the spatial configuration of street elements allows the urban form and its impacts to be measured. For example, Berghauser Pont et al. (2019) investigate the centrality of the road network and its impact on pedestrian movement in three cities. The studies of street networks predominantly use graph-based methods (Marshall et al., 2018). In this case, the graph may be a primal graph (for example street intersections are the nodes of the graph and streets are the edges) or, as is the case for Space Syntax, it may be a dual graph (streets becomes nodes and streets

intersections are edges) (Porta et al., 2006).

The two strategies (discrete area and network) are both complementary (Berghauser Pont, 2021). The first can be used to obtain information on the objects studied independently, while the second helps understand the objects in their interactions with each other (location of one object in relation to others). Chapter 8 pursues this idea. We will present a method to group buildings according to their spatial distribution in order to measure a topological density index.

Chapter 4 is based on the strategy of discrete areas. For theoretical and real urban areas, we will measure fractal dimensions and question their meaning. A fractal dimension is an index that gives information on the way in which an object being studied (in our case, buildings) fills a space. Chapter 3 goes back to the basic notions of fractal geometry and fractal dimensions to enable a full understanding of this thesis (especially Chapter 4)

1.4 Objectives and structure of the PhD thesis

The main objective of this thesis is to contribute to the understanding of urban spaces. In particular, the aim is to critically analyse and develop measurements that researchers and urban planners use to morphologically characterise cities. *"What are we measuring?"* is the thread that runs through our research and consequently through this thesis. In our work, we try to understand what the different methods and indices measure in order to better understand what they tell us about the morphological organisation of urban spaces.

Chapter 2 is the starting point of our work. In this chapter, we characterise the Brussels-Capital Region using four methods (two functional and two morphological). We note that each method leads to a specific model of Brussels. Unfortunately, we will not investigate the relationship between functional processes and morphological characteristics further in the thesis, as we want to focus on the understanding of the latter.

Fractal geometry, and more specifically the fractal dimension, allows the characterisation of urban spaces across scales. Chapter 3 and Chapter 4 deal with fractal dimension. Chapter 3 goes back to some basic notions. In chapter 4, we create a protocol that limits the biases when measuring two fractal dimensions (box-counting and correlation). In addition to establishing the measurement protocol, we try to understand the morphological features revealed by each dimension.

Chapters 5, 6 and 7 focus on the delineations of urban spaces and the characteristics of these delimitations. Chapter 5 presents a comparison of three methods based on analyses of a statistical distribution across scales to delineate urban agglomeration. Are the results of these methods different? What are the characteristics of each method that influence the results? These are the questions that drive this chapter. We will see that each method leads to specific results that can be used thematically.

Since there is no such thing as a single delimitation, it is necessary to understand the characteristics of the delimited spaces (boundaries and built-up fabrics). In Chapter 6 (Characterisation of urban agglomerations delineated by *MorphoLim*), we delimit morphological agglomerations using *MorphoLim*, which is a delineation method based on the fractal properties of urban spaces. We then measure the morphological characteristics of these boundaries and the spaces they delimit. Are these characteristics complementary in understanding the morphology of an urban space? We will see that they complement each other by highlighting specific characteristics. Together they allow the characterisation and comparison of different agglomerations.

The contribution of the third dimension (building height) in the delineation of cities is discussed in Chapter 7, focusing specifically on Valencia (Spain). Still using *MorphoLim*, we note that the agglomeration of Valencia is both very dispersed and very fragmented. We carry out different delimitations of the agglomeration according to different building heights.

In Chapter 8, we present a new method to group buildings according to their spatial distribution in order to measure the topological density index. Using this method, a density index is computed without pre-defining a reference area. The density index is dependent on the distance between buildings.

Finally, we return to the main results and the limits of our work in the concluding chapter (Chapter 9). It provides an opportunity to outline the work that remains to be done by looking to the future.

2

Complexity of the spatial structure of cities: the example of Brussels-Capital Region

In this chapter, we present a case study: the partitioning of the Brussels-Capital Region (BCR) using several methods and data. Results show that different urban models are rediscovered (the concentric models of Alonso-Muth and Burgess, the sectoral model of Hoyt and the multiple nuclei model of Harris and Ullman) depending on the data and methods used. This case study served as a starting point for our reflections and work. Indeed, whether the data (census, telephony, buildings) and methods (urban factorial ecology, community detection, density index, natural cities, etc.) are classic or more recent, they lead to different interpretations of the BCR urban model. Our wish to critically analyse the measurements commonly used to characterise urban spaces stems from this observation.

2.1 Partitioning an urban area

How measuring and understanding a city as a complex system? What variables and indices should we rely on in order to understand the spatial organisation of a city? Since the seminal work of the Chicago school (1930s), a common practice to understand a city is partitioning it into groups of places sharing similar properties. It is a way to classify elementary spatial units to better understand their similarities and discrepancies and, above all, to better understand intra-urban spatial structures and dynamics within the city. This is usually done in terms of the socio-economic composition of each place but can also be done in terms of interrelationship between georeferenced populations or that of built-up morphologies: each of these aspects are relevant to figure out the intraurban homo- or heterogeneity.

Quantitative revolution (1960s) has, methodologically, facilitated the partitioning with the development of factorial and cluster analysis, and faster and faster computing systems, even if available data to describe elementary spatial units were rather scarce (urban factorial ecology). Digital revolution has recently exponentially increased the amount of available data (Floridi, 2012; Kitchin, 2013) and, among them, a lot of localized information. This does not mean that these data did not exist before, yet the detection, the registration, the share and the use of these data reached an unprecedented level since the end of the 20th century. The risk with these new emerging data sciences is that some scientists dive into data before elaborating on the specific goals of the research and could lead to data crunching rather than modelling.

Two general kinds of information are commonly used to measure an urban area. First, specific attributes characterizing distinct elementary spatial unit (X_i) of the study area can be used. These attributes can either describe the population living in the units, the economic activities, or the characteristics of the built environment. Second, geographers use information about the interrelationship (w_{ij}) between couples of elementary spatial units i and j. In that case, graph theory can help geographers to analyse the network of interrelations between places, as well as community detection algorithms. For urban geographers interested since a long time in describing and understanding intra-urban structures and dynamics, the emergence of Information and Communication Technology (ICT) data (Global Positioning System, sensors, etc.) offers new ways to get information about places (X_i) and interrelationships between them (w_{ij}).

Limiting urban analyses to ICT data is, for sure, not appropriate nowadays: there are no time series, and the spatial representativeness is often questionable. Some parts of the population can be totally missing in the data coverage (see e.g. Longley et al., 2015). Therefore, censuses and surveys are both still relevant for measuring urban complexity. Censuses have the advantage of being well defined, covering almost the entire population and comparable through time. ICT data measure other aspects, often in a very short time lag and at the individual level (see e.g. Longley et al., 2015; Miller and Goodchild, 2015).

In this chapter, we mobilize classical data and methods but also new data and recent innovative methods to partitioning the Brussels-Capital Region (noted as BCR, which is an administrative and political entity on its own), that constitutes the dense urban core of the capital city of Belgium. The objective is to understand the spatial organisation of the BCR through a historical model of urban geography. Do the results of different methods converge? To answers to this question, we partition the region according to four different methods (by order of appearance):

1. an "urban factorial ecology" to characterise the space with socio-economic, ethnic, demographic, etc. variables (characterisation of multiple x_i)
2. a community detection method partitioning phone call data (characterisation of w_{ij})
3. a density measured on built-up footprint (characterisation of a morphological x_i)
4. a *"natural cities"* method (characterisation of a morphological x_i)

The first two methods use a functional approach while the next two have a morphological approach. We analyse the results of each method in relation to the historical models of urban geography. We then compare the methods with each other.

2.2 Functions and interactions of/between places

2.2.1 Urban factorial ecology: a benchmark

Urban factorial ecology was scientifically very fashionable in early quantitative spatial analysis (1960s) (see e.g. Berry and Ress, 1969; Hunter, 1972; Johnston, 1978) and had already been applied to Brussels before (see e.g. Dujardin et al., 2008). The aim of these methods is to identify groups of places for which inhabitants have similar characteristics (X_i; socio-economic, ethnic, demographic, etc.) by means of a factorial analysis followed by a clustering method (Pruvot and Weber-Klein, 1984). We here use a principal component analysis and a hierarchical cluster analysis in order to identify spatial clusters and further interpret them in terms of classical "urban models".

In the frame of our case study, several analyses were performed with different sets of variables, with and without a preliminary component analysis. The results in terms of urban spatial structure are a little sensitive to the used method; here, we develop one example. The data used are provided at the scale of the neighbourhood by the *Institut Bruxellois des Statistiques et d'Analyses* [1]. The selected attributes (see Appendix A.1) are available online, as well as a clear definition of each variable, making the data collection simple and reproducible. Among the 145 neighbourhoods of the BCR, 27 are excluded from the analysis because they are not inhabited (mostly correspond to green spaces or industrial zones). The selection of numerous variables (48) is inspired by the literature on factorial ecology, in order to catch the socio-economic conditions, as well as the demographic and ethnic conditions of each spatial unit.

As expected, variables are highly correlated and can easily be summarized in four components (see Fig.2.1 for the composition of each component). The score of the two first components are rather organized in concentric structures: the first one around the centre (mainly an age gradient), and the second around the European quarter (mainly a "metropolitan activities" gradient). The third component is organized in sectors and is largely based on income and characteristics of the housing stock: large income and large dwellings in the south-east and low income and small dwellings in the north-west. The fourth

[1] The "Monitoring des Quartiers" is an interactive tool available online. It has the objective to make available a selection of indicators that characterize disparities and dynamics within the BCR. Not all data are from the same date. Its availability makes the analysis easily reproducible. Retrieved from http://ibsa.brussels/chiffres/chiffres-par-quartier.

component reveals a structure in “donut”, highlighting places specialized in residential functions at a certain distance from the city centre.

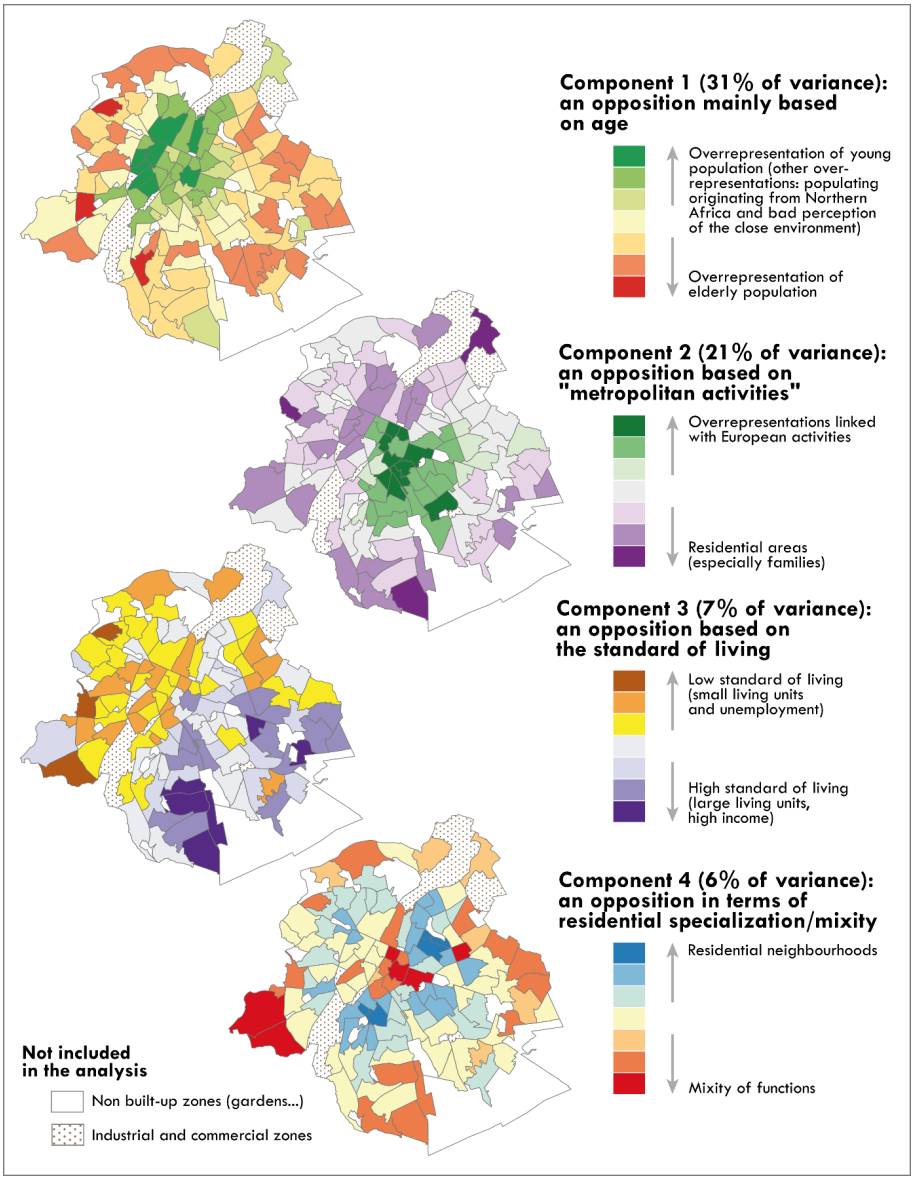


Figure 2.1. The four principal components of the PCA based of variables listed in Appendix A.1

A ward hierarchical cluster analysis based on the scores of the *Principal Component Analysis (PCA)* is then conducted for grouping the neighbourhoods that share similar profiles. Indicators for finding the best number of clusters (Cubic Clustering Criterion, entropy, Dunn, Calinski-Harabasz, shape of dendrogram) show that the optimal number of clusters turns around five to seven. The solution with six clusters is here illustrated (Fig.2.2a), six being also a number comparable to the number of mobile phone communities that are extracted and presented in the next subsection.

Results (Fig.2.2a) confirm former analyses conducted (see e.g. Vandermotten and Vermoesen, 1995; Thomas and Zenou, 1999): Brussels combines two structures strongly shaping the city: a concentric and a radial one (Hoyt, 1939). The factorial ecology analysis confirms the existence of two crowns around the Central Business District (CBD). Cluster 2 corresponds to the CBD with high values of the scores of the second component, little residential. The first crown is mainly residential, characterized by a high population density (positive values on Component 4) and divided into two sectors: Cluster 1 (pink) in the south-east which is mainly residential and better-off (high income, large dwellings) and Cluster 3 (green) in the north and west (low income, small housings). The second crown is also divided into two parts, even if Cluster 6 (orange) and Cluster 4 (purple) are less marked than those of the first crown. Last, Cluster 5 (blue) corresponds to specific urban structures such as large hospitals, sports halls or cultural centres.

2.2.2 Community detection in ICT data: phone basins

Nowadays, there is no need to wait for official census data to measure the city: the emergence of new sources of data now allows sensing its "pulses" in real-time. Contrarily to census data describing attributes of places in terms of socioeconomic, demographic and ethnic characteristics of their inhabitants, ICT data enable to characterize and follow people in space and time. ICT data are particularly useful to monitor daily life and short-term processes by detecting and measuring changes within space (Lee and Lee, 2014). They undoubtedly open new perspectives in interaction analyses and urban geography (Batty et al., 2012). This also allows the capture of intra-urban interrelationships between places (w_{ij}). The availability of such data increases every day due to the multiplication of apps on smartphones that enable following people and their geolocations.

We here limit ourselves to one example: mobile phone calls. They already have been proved to be very useful in monitoring and mapping the de facto population, as well as people's spatial mobility and social networks (Blondel et al., 2010; Griffiths et al., 2010; Batty et al., 2012). It enables one to better grasp the interactions between people and their environment (see e.g. Ahas et al., 2015; Deville et al., 2014). With this type of data, it is common to extract groups of people or locations with remarkably high interactions, leading to what is called communities.

If this kind of data are nowadays very attractive for approximating "social networks", they also appear to have major limitations (see e.g. Calabrese et al., 2014) including a time-consuming "cleaning" phase. The data are not publicly available, and several operators share the market with no information on market shares. It is quite common that the data are aggregated due to confidentiality issues (no distinction between professional and private use; calls are, in our case, located to the closest antenna and not at the exact location of the call).

We here use one month of phone communications of one of the three major service providers operating in Belgium (April/May 2015). The database includes all mobile phone calls, ignoring their duration, between two phone numbers from the provider, geocoded by antenna at the hour of the call. The data set is here limited to the calls for which both antennas are located in the BCR, that is more than 4.9 million calls. To approximate the coverage of each antenna and to improve the readability of the maps, a Voronoi diagram is designed around each antenna (see Adam et al., 2017). To partition the BCR we use the *Louvain Method*. It is commonly applied in community detection because it quickly finds partitions of nodes that maximize the modularity without defining a priori central places and/or thresholds (Adam et al., 2018a; Thomas et al., 2017).

Applied to the BCR mobile phone data set, the Louvain Method endogenously detects six communities that can be called "phone basins". A phone basin corresponds to antennas that have a higher propensity to call other antennas than any antenna classified in other communities. In order to avoid a suboptimal solution, the algorithm is run 1,000 times (Adam et al., 2018b). Hence, for each partition, each node can be characterized by the percentage of runs for which it is associated to the same community and so measuring the stability of the partition.

A striking first feature is that most contiguous Voronoi diagrams belong to the same community, this means that calls are more numerous between antennas that are closely located and that people call more often their closest neighbours than those farther located (Tobler, 1970): geography still matters! Tab.2.1 represents the matrix of calls emitted and received between the six communities. For each community, the highest number of calls is observed between antennas belonging to the same community; they correspond to the diagonal of the matrix. The intra-community calls represent more than 50% of the calls emitted or received by each community in the exception to Community 6 (45%). In terms of calls made between communities, Tab. 2.1 and Fig. 2.2c show different realities if the direction of the calls is taken into account. For instance, on one hand, a high number of calls is made from Community 1 to Community 2 (116,986 calls), and on the other, the antennas belonging to Community 2 were many times in communications with antennas from Community 4 (155,949 calls) as well as of Community 1 (123,096 calls).

Fig.2.2b clearly reveals the spatial organization of calls in the BCR. Two of them are located in the middle of the study area: a community centred on the Pentagon [2] with an extension to the North Railway Station (Community 4) and another one around the EU offices (Community 1). These two communities concentrate 43% of the calls made within the BCR; the other communities are further organized in sectors around these two.

	1	2	3	4	5	6
1	**684,066**	116,986	63,914	67,722	75,337	38,475
2	123,096	**741,082**	67,416	155,949	54,281	72,337
3	70,716	71,856	**263,209**	49,667	40,955	35,133
4	71,806	154,892	45,944	**629,775**	81,803	83,723
5	75,663	52,409	36,826	78,132	**316,415**	38,7
6	41,577	72,538	33,734	85,011	40,846	**217,392**

Table 2.1. Number of calls between communities

[2] Usual name of the area located inside the former 14th-century walls of the city.

2.3 Morphometrics of built-up footprint

Cities are complex by definition, and the analysis of their complexity includes, among the already mentioned analyses made on people and interactions, the morphology of their built-up components (see e.g. Thomas and Frankhauser, 2013). Having a specific focus on morphology allow the comparison of characteristics of people and their interrelationships through locations with the morphological features of places. The objective here is to give an image of the built-up disparities and further partition the city in terms of built-up urban similar morphologies. The literature mentions many indices, mainly issued from landscape ecology, for measuring and characterizing urban morphologies (see e.g. Medda et al., 1998; Schoning et al., 1999; Caruso et al., 2017). We here select two methods to quantitatively grasp the morphological reality of Brussels. (1) a built-up density index (2) the *Natural cities*.

We use the 2009 Cadastral footprint of the buildings as well as the centroids of each building. Every isolated small (less than $20m^2$) building was erased from the database. Analyses are performed on an area larger than the BCR (an area encompassing the entire former province of Brabant that surrounds the BCR) in order to avoid border effects. From this large set, we here isolated and present the results of the BCR.

2.3.1 Built-up density

Built-up density can be computed in different ways. The most common indices are the Floor Space Index (FSI) and the Ground Space Index (Berghauser Pont and Haupt, 2005; Teller, 2021). FSI is the ratio of the total area of the floors of buildings and the total ground area. GSI is the ratio of the area of the built-up footprint and the total ground area. As we do not have the number of floors per buildings, we here simply compute the GSI.The studied area is covered by a grid of squared cells (250m. x 250m.). We express the GSI as the ratio of the surface occupied by the buildings divided by the total surface of a basic spatial unit (62500 m2).

Results are reported in Fig.2.2d: a strong and clear centre–periphery structure appears. The density increases from the outskirts to the centre. This structure reminds clearly the classical Alonso-Muth model as already discussed for Belgium by (Verhetsel et al., 2010).

2.3.2 From "natural cities" to "natural urban clusters"

A set of papers has recently been initiated by Jiang around the concept of *"natural cities"* (Jia and Jiang, 2010; Jiang, 2013; Jiang and Miao, 2015, see e.g.). The methodology relies on a head/tail division rule to derive *"natural cities"*, based on the assumption that there are far more smalls things than larger ones. A triangulated irregular network (TIN) is used, made up of individual locations that are considered as nodes (initially street nodes and later any location-based social media users such as Twitter or Brightkite). If several authors have already supported the use of streets networks for delineating cities (Jia and Jiang, 2010; Arcaute et al., 2014; Masucci et al., 2015), Thomas et al. (2012) showed that the spatial organisation of buildings within a city is rather different from that of street networks. Hence, in order to compare results with the previous morphological partition, we here consider the centroid of each building as the nodes used in the TIN. The method identifies the edges between the nodes (on the TIN) smaller than the average length of the edges as limits of *"natural cities"*. The method relies on the fact that the length of the edges follows a hierarchical distribution where the mean length is used as a threshold for delineating the "cities" (large number of small edges corresponding to cities and small number of large edges corresponding to rural areas). We will see later (Chapter 5) that in order to obtain *"naturals cities"* we have to select some areas on the basis of a *head/tail breaks* classification (Jiang, 2013, 2015, 2019). We do not perform this step and therefore do not consider *"natural cities"* but *"natural urban clusters"*.

The method delineates zones where the spatial proximity of the centroids is higher than the mean length in the study area (48 m). Patches are clearly identified and mapped (Fig 11.1-e). The white space between the patches corresponds to large boulevards, wide infrastructures as railways, green spaces, squares, etc., or simply built-up wards with larger distances between their centroids (larger buildings or larger gardens). Some coloured patches are large and reflect the history and geography of the city (see e.g. Vandermotten and Vermoesen, 1995). Some others are very small, corresponding to specific allotments or pinpoint urban projects. Figure 2.2e clearly tells another story even if some resemblance exist; there is some resemblance between socio-economic realities (Fig. 2.2a) and built-up landscapes (Fig. 2.2e).

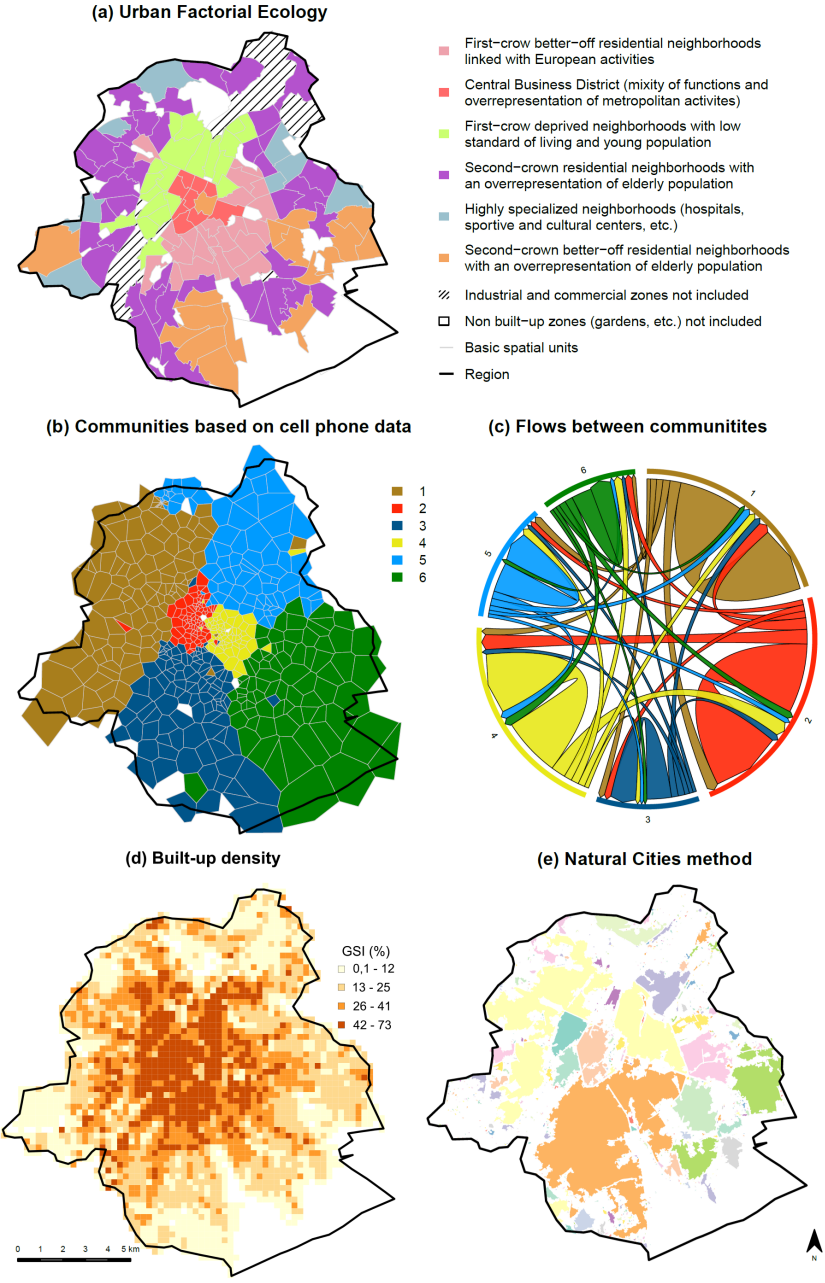


Figure 2.2. Partitioning of the Brussels-Capital Region according to the different methods and data

2.4 Conclusion and comments

The four partitions presented in Fig.2.2 result from a unique objective (understanding the city), but the type of data and the methodology, as well as the basic spatial units used, differ. Yet some commonalities are observed between the partitions, with some parts of the city being for example encompassed both in a given cluster in the factorial ecology and a specific community into the group of antennas. A synthesis is proposed in Tab.2.2. The spatial structure of each partition is clearly not similar, at the exception to core area that we found in each of them. Communities centred on the Pentagon and on EU offices are grouped in a socio-economic cluster characterized by a concentration of offices and urban facilities (CBD). The CBD and the first-crown in the factorial ecology are characterized by a dense organization of the built-up environment according to the morphological partition. On the opposite, it is obvious that the comparison of a pure centre–periphery gradient (built-up density) and a sectoral structure (communities based on cell phone data) will not lead to a strong resemblance between the two partitions.

With the example of Brussels, we confirm the complexity of the urban structure and the difficulties in fitting data, methods and objectives. There is no unique way of measuring this complexity: different databases and methodologies can lead to different spatial structures. The classical urban models are partly rediscovered. Concentric model (Alonso-Muth/Model of Burgess), sectoral model (Hoyt) and multiple nuclei model (Harris and Ullman) are clearly appearing in Fig.2.2 confirming former studies: Brussels is a superposition of different spatial patterns. Different forces structure the city and therefore make it complex. To understand the different aspects of the city, we had to understand the meaning of each result. The definition of the data needs to be controlled, as well as the underlying processes and the research objectives. We cannot let the data speak by themselves. We need to capture the meaning of the data, method and indices. In other words, we need to understand what we are measuring in order to draw the right conclusions.

The main goal of this thesis is to contribute to the field of research of urban geography by giving critical analysis to the measurements used to understand cities. We only focus on a morphological approach of the urban spaces. There will therefore be very little reference to functional spaces.

	Type of data	Basic spatial units	Approximative spatial structure of the clustering
Factorial ecology (PCA + clustering)	Socio-economic, conventional (census)	Statistical neighbourhoods	Combined centre-periphery and sectoral structures
Communities based in cell phone	Relational, unconventional	Voronoi diagrams around antennas	Sectoral structure
Built-up density	Built-up footprints	Cells grid	Centre-peryphery structure
"Natural urban clusters"	Built-up footprints (centroid)	/	Multiple nuclei

Table 2.2. Comparison of the three partitions of Brussels

3

Fractal geometry in geography: a general background

The aim of this chapter is to provide a quick reminder of the concepts necessary to understand the rest of the thesis (in particular Chapter 4)[1]. Section 3.1 explains what a fractal and a fractal dimension are and why we use them in urban geography. Section 3.2 explains how a fractal dimension is measured by presenting the box-counting dimension and the correlation dimension.

[1]For a much more complete analysis of the use of fractal analysis in geography, we recommend reading the book *"Analyse et modélisation fractale de la distribution spatiale des implantations humaines"* (Tannier, 2199).

3.1 Fractal geometry and fractal dimension

Since more than four decades, researchers have shown that spatial distributions of geographical elements taken at different scales have common morphological features (Goodchild, 1980; Arlinghaus, 1985, 1987; Batty and Longley, 1986, 1987; Goodchild and Mark, 1987; Frankhauser, 1988). This means that for some representations of the city, there is no specific scale with which to characterise them; there is *scale invariance*. Among other things, scale invariance can be observed and measured in fractal figures. This is why many studies have demonstrated the benefit of using fractal analyses to characterise urban fabrics (Longley and Mesev, 1997, 2000; De Keersmaecker et al., 2003; Tannier and Pumain, 2005; Thomas et al., 2007, 2008; Feng and Chen, 2010; Chen, 2011). Fractal analyses, being scale invariant, do not only apply to particular proprieties at a particular scale but show principles that appear at different scales [2] (Frankhauser, 1994). Indeed, fractal figures show self-similarity (the shape of the fractal figure is the same whatever the scale of observation). Fig.3.1, which represents a *Sierpinski carpet* (a well-known fractal figure), helps to understand this.

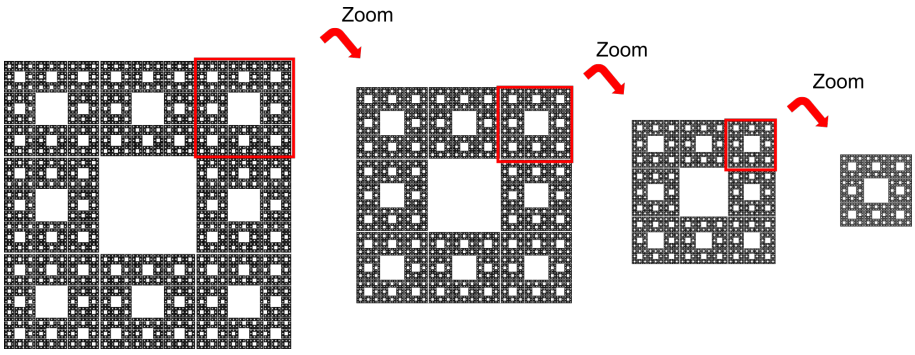


Figure 3.1. An example of self-similarity in fractal figure (Sierpinski carpet)

When they measure a city, one of the objectives of researchers is to measure how the space is filled (for example by built infrastructures). To do this, they use the concept of dimension. To understand what is a dimension, let us assume simple shapes such as a line, a square and a cube. To measure these shapes, we will use a smaller and smaller standard object. Logically, if we halve the size of the standard measuring the line, the line will measure the

[2] "We will see that the fractal concept does not only apply to particular properties of some levels of description, but will help us to discover principles that appear in comparable form at different levels of observation ». (Frankhauser, 1994, p.98)

equivalent of two lengths of the standard (Fig.3.2). For the square, if we divide the length of the side of the square standard by two, it will take four standard surfaces to cover the square. By the same logic, it will take eight standard volumes to fill the cube if we divide the length of the edges of the standard cube by two. The Tab.3.1 shows, for different reduction sizes of the standards, the number needed to recover their respective figure. There is a direct link between the size and the number of standards needed to recover the shapes. This link is the dimension of the objects. This dimension is 1.0 for the line, 2.0 for the square and 3.0 for the cube. These dimensions therefore characterise the way in which the objects fill the space (the number of possible direction). The line fills the space in one dimension, the square fills the space in two dimensions, and the cube fills the space in three dimensions.

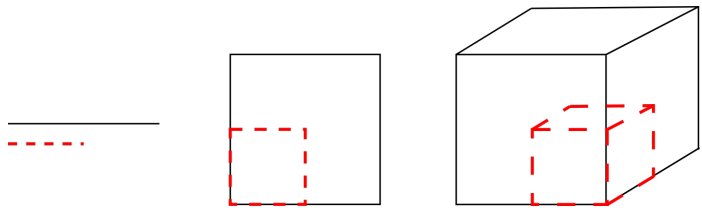


Figure 3.2. A line, a square and a cube with standard measuring (in red)

r	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	r
N for a line	$2 = 2^1$	$4 = 4^1$	$8 = 8^1$	$16 = 16^1$	$N = (\frac{1}{r})^1$
N for a square	$4 = 2^2$	$16 = 4^2$	$64 = 8^2$	$256 = 16^2$	$N = (\frac{1}{r})^2$
N for a cube	82^3	$64 = 4^3$	$512 = 8^3$	$4096 = 16^3$	$N = (\frac{1}{r})^3$

r is the reduction factor of the side length of the standard measuring (basis = 1) ; N is the number of standard measuring to fill the forms

Table 3.1. The dimension for simple shapes

We see for simple (Euclidean) forms that the dimensions are integers. These are called topological dimensions. They give the number of coordinates needed to locate a point on an object and thus measure its size (its length for a one-dimensional shape, its area for a two-dimensional shape and its volume for a three-dimensional shape (Tannier, 2017).

The dimension of fractal objects follows the same logic except that due to the complexity of fractal forms, the dimensions will no longer be integers. It is therefore no longer a topological dimension but a fractal dimension that will characterise the shape of the fractal object. The fractal dimension (D) is therefore a measurement of the amount of space occupied as the spatial resolution adopted becomes finer and finer (Falconer, 1990; Frankhauser, 2003). It is an indicator of the complexity of the shape of an object. To understand it, we need to return to the example of the Sierpinski carpet (Fig.3.1) and see how this figure is created. Fig.3.3 shows how it works. There is an initiator, a simple square of side L. There is then a generator consisting of a number (N) of copies of the iterator (in our case, eighth copies placed around the perimeter of the initiator) to which a reduction factor (r) has been applied (here, r = 1/3). The generator is then applied to each element of the figure then created in the first step according to the same reduction principles. After many repetitions, the figure is a Sierpinski carpet.

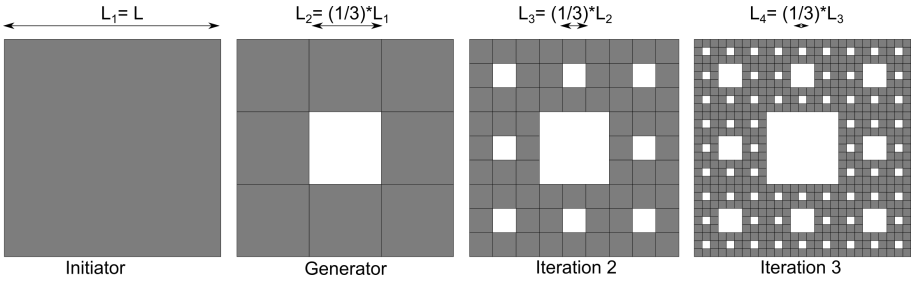


Figure 3.3. The initiator and the first three stages of construction of a Sierpinski carpet (adaptation from Mandelbrot (1982); Tannier (2017))

In the same way as for calculating the dimension of a Euclidean object, the fractal dimension (D) is the link between the number of elements (N) necessary to fill the fractal object and the size (r) of these elements (equation 3.1).

$$N = (\frac{1}{r})^D \tag{3.1}$$

The fractal dimension (D) called *"dimension d'homothétie" (ou de similarité) (D_S)* is,

$$D_S = \frac{\log N}{\log \frac{1}{r}} \tag{3.2}$$

In our example, $N = 8$ and $r = \frac{1}{3}$, then the fractal dimension D_S of our Sierpinski carpet is

$$D_S = \frac{\log 8}{\log 3} = 1.892 \tag{3.3}$$

The spatial distribution of human settlements can be studied as sets of points. The fractal dimension will then be close to 2 when the spatial distribution of the point mass is homogeneous (weakly hierarchical). Conversely, for a highly hierarchical space with large concentrations of mass in some places, the fractal dimension will be close to 0 (Frankhauser, 2003).

3.2 Measuring the fractal dimension

Urban patterns do not exhibit pure self-similarity like the Serpinski carpet. For real urban fabrics, it is then clearly impossible to determine the values of N and r. But at some scales, a statistical self-similarity appears which will allow us to use the properties of fractals as the fractal dimension to measure urban spaces (Frankhauser, 1994; Batty and Longley, 1994). To calculate the fractal dimension of an irregular object, Minkowski (1903) thought that only the local quantification of the occupied space was relevant (Tannier, 2017). Bouligand (1929) develops this thought by developing the principle of covering step by step. The idea is to cover the set of points by closed boxes and to count the number of boxes needed to recover the set of points. For an increasingly small box size, we can calculate the fractal dimension which is in fact the asymptotic behaviour of the number of boxes (N) when the size of each box (ϵ) tends towards 0 (equation 3.4.

$$D = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}} \tag{3.4}$$

For a real wold pattern such as exhibited by buildings, it is unrealistic to compute the N for ϵ close to 0. D is estimated considering sufficiently wide intervals of values ϵ, for which the curve representing N as a function of ϵ corresponds to a power law (equation 3.5) with N the number of boxes, ϵ the size of boxes, D the fractal dimension, a a form-factor and c a constant (De Keersmaecker et al., 2003). In geography, two ways of calculating the fractal dimension of built-up areas are mainly used and are based on this principle of recovering at different scales and fitting a power law. These are called the *box-counting dimension* and the *correlation dimension*.

$$N(\epsilon)\ a\epsilon^{D} + c \tag{3.5}$$

Box-counting dimension (D_B) - The set of points is covered by a grid. The number of mesh with a minimum of one point are counted. The

operation is then repeated several times with a smaller and smaller (or bigger and bigger) mesh size according to a specific factor. For a limited range of scale, the number of filled mesh (N) and the size of mesh (ϵ) follow a power law. A linear regression can fit this power law after logarithmic transformation. D_B is then the coefficient of slope of this regression [3].

Correlation dimension (D_C) - For each point in the pattern, the number of neighbouring points within a radius that increases or decreases by a specific factor is computed. For each radius size, the average number of neighbours is calculated for the entire point pattern. For a limited scale rank, the average number of neighbours and the size of the radius are related by a power law whose correlation dimension (D_C) is again the coefficient of the slope calculated directly or via a linear transformation.

The fractal dimension is therefore an index that measures the way in which urban elements such as buildings fill a space. It is therefore a key index for measuring urban form. However, we will see in Chapter 4 that numerous parameters influence the estimation of fractal dimensions, either by box-counting (D_B) or by correlation (D_C). Moreover, we will see that the values taken by D_B and D_C are not always identical. We will therefore question the meaning of these values. The objective is always to understand what we are measuring when we measure the urban form.

[3] D_B can also be estimated directly on the power law

4

Analysis of box-counting and correlation dimensions: calculation parameters and morphological meaning

In this chapter, we study how box-counting and correlation dimensions are calculated. We also question the meaning of the different values taken by these dimensions. The first section of the chapter presents fifty examples of built urban fabric. The second section is devoted to establishing a measurement protocol that should enable consistent measurement of the dimensions in order to ensure comparability. The search for morphological features revealed by each of the dimensions is discussed in the third section.

4.1 Introduction

Nous avons vu dans le chapitre 3 que la dimension fractale est un indice qui mesure la manière dont un objet (par exemple les bâtiments d'une ville) rempli un espace. Cet indice est donc utilisé dans diverse étude pour mesurer la forme urbaine (voir par exemple De Keersmaecker et al., 2004; Shen, 2002; Thomas et al., 2010). Il existe plusieurs méthodes pour mesurer la dimension fractale d'un objet. Comme mentionné à plusieurs reprises dans cette thèse, l'utilisation de méthode différentes conduit bien souvent à l'obtention de résultats différents. Les méthodes permettant le calcul des dimensions fractales ne font pas exception à ce principe.

Les dimensions de boîte et de corrélation sont les dimensions les plus souvent utilisées en géographie pour caractériser des tissus bâtis. L'objectif de ce chapitre est d'étudier plus en détail la manière dont se calcul ces deux dimensions. Quel sont les paramètres à prendre en compte qui pourraient influencer le calcul de la dimension de boite (D_B) ou de corrélation (D_C) (appelées "box-counting" et "correlation" dans le chapitre 3) ? Puisque les valeurs prises par D_B et D_C peuvent être différentes, un second objectif du chapitre est d'identifier les caractéristiques morphologiques des tissus bâtis qui influencent plus ou moins la valeur de chaque dimension. Dans son ensemble, ce chapitre vise à répondre au besoin, mis en avant par Lam (2009), d'une évaluation minutieuse de différents algorithmes d'analyse fractale pour qu'ils puissent être utilisés de manière fiable pour caractériser des objets géographiques.

Notre étude est basée sur un échantillon de cinquante tissus bâtis parmi lesquels vingt-cinq sont issus du monde réel et ont été téléchargés depuis la plateforme geofabrik d'Open Street Map (http://download.geofabrik.de/). Les vingt-cinq autres tissus sont qualifiés de théoriques car ils ont été créés de toutes pièces via un logiciel dédié à la génération de figures fractales. Ces tissus bâtis théoriques sont strictement auto-similaires et on connaît leur dimension de similarité D_S (cf. chapitre 3) avec exactitude. Ils servent de références auxquelles confronter les résultats d'analyse obtenus pour les tissus bâtis réels. Un certain nombre de paramètres d'analyse ont une influence sur la mesure et l'estimation de la dimension fractale. Il est donc nécessaire de fixer l'ensemble des valeurs de paramètres susceptibles d'influencer les dimensions de boîte et de corrélation des cinquante tissus bâtis. A cet effet, une analyse de variabilité est effectuée sur un sous-échantillon de dix tissus bâtis choisis parmi les cinquante tissus étudiés. L'idée est d'observer l'impact des variations de chaque paramètre pris séparément sur les dimensions fractales estimées afin de choisir le mieux

possible le jeu de valeurs de paramètres qui sera utilisé pour l'étude de l'échantillon dans son ensemble. Une fois ces paramètres d'analyse fixés et communs à tous les tissus, il est possible de comparer entre elles les valeurs de dimension fractale obtenues.

Ce chapitre se décompose en trois sections : présentation des tissus bâtis réels et théoriques étudiés (Section 4.2) ; Mise en évidence de la variabilité des dimensions estimées en fonction des valeurs des paramètres d'analyse et choix de ces valeurs (dix tissus bâtis) (Section 4.3); Mise en évidence des similitudes et des différences entre les dimensions de boîte et de corrélation (cinquante tissus bâtis) (Section 4.4).

4.2 Les tissus bâtis étudiés

4.2.1 Les tissus bâtis réels

Les vingt-cinq tissus réels (Fig. 4.1) sont issus de la plateforme *geofabrik* qui donne accès gratuitement à des bases de données géographiques à contributions libres. Les tissus bâtis sélectionnés ne sont pas le reflet de la réalité bâtie des villes considérées mais le résultat de ce qui a été cartographié par les contributeurs volontaires au projet OSM. Ceci n'est cependant pas gênant ici puisque notre objectif n'est pas de caractériser les villes à proprement parler mais d'étudier différentes formes d'organisations bâties. Les vingt-cinq tissus bâtis réels ont été choisis afin de constituer un échantillon le plus diversifié possible en termes de distribution spatiale des bâtiments, du nombre et de la surface de ceux-ci, ainsi que de la densité bâtie[1]. Les villes côtières et les villes frontalières ont été exclues de la sélection afin que les tissus aient un développement bâti dans toutes les directions du plan. Les villes aux éléments morphologiques caractéristiques comme les grands cours d'eau ou des reliefs marqués ne sont, elles, pas exclues de la sélection étant donné que ces éléments morphologiques contribuent à la diversité et à l'originalité des distributions spatiales de bâtiments dans chaque ville.

[1]La densité bâtie est mesurée à l'aide du coefficient d'emprise au sol (CES), qui est le rapport entre la surface totale couverte par les bâtiments et la surface de la zone d'étude. Cet indice est exprimé en pourcent (%).

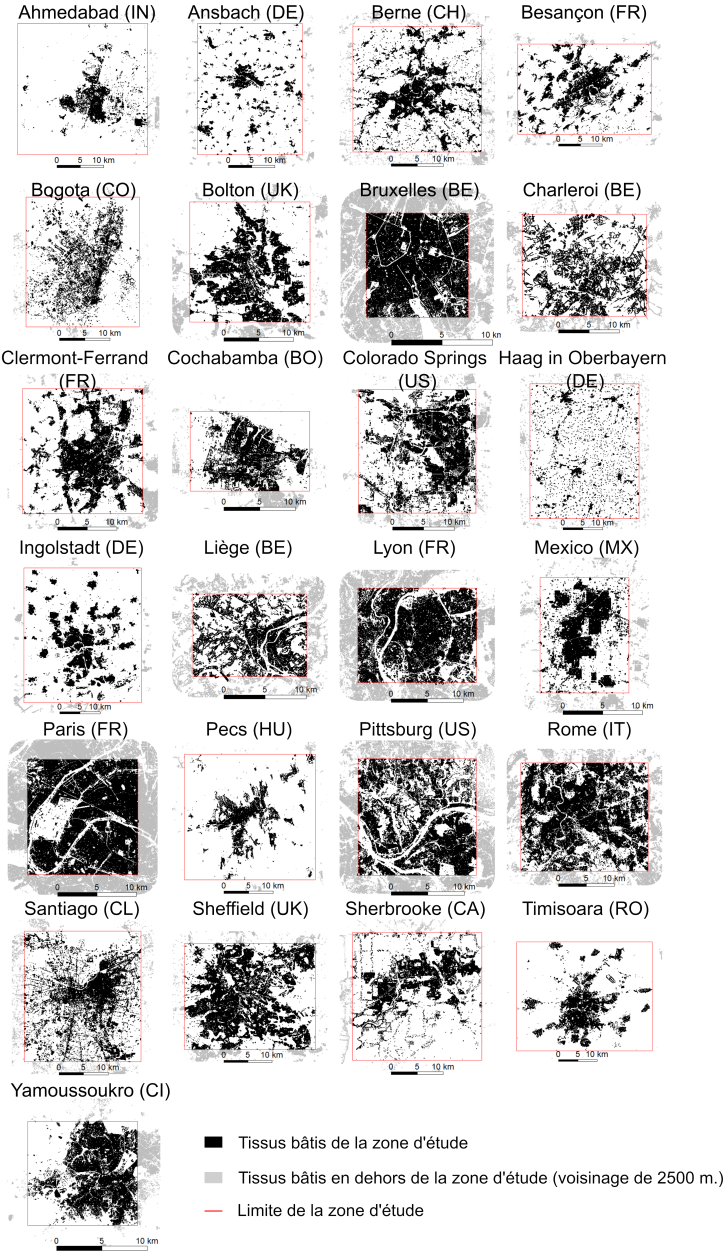


Figure 4.1. Les vingt-cinq tissus bâtis réels

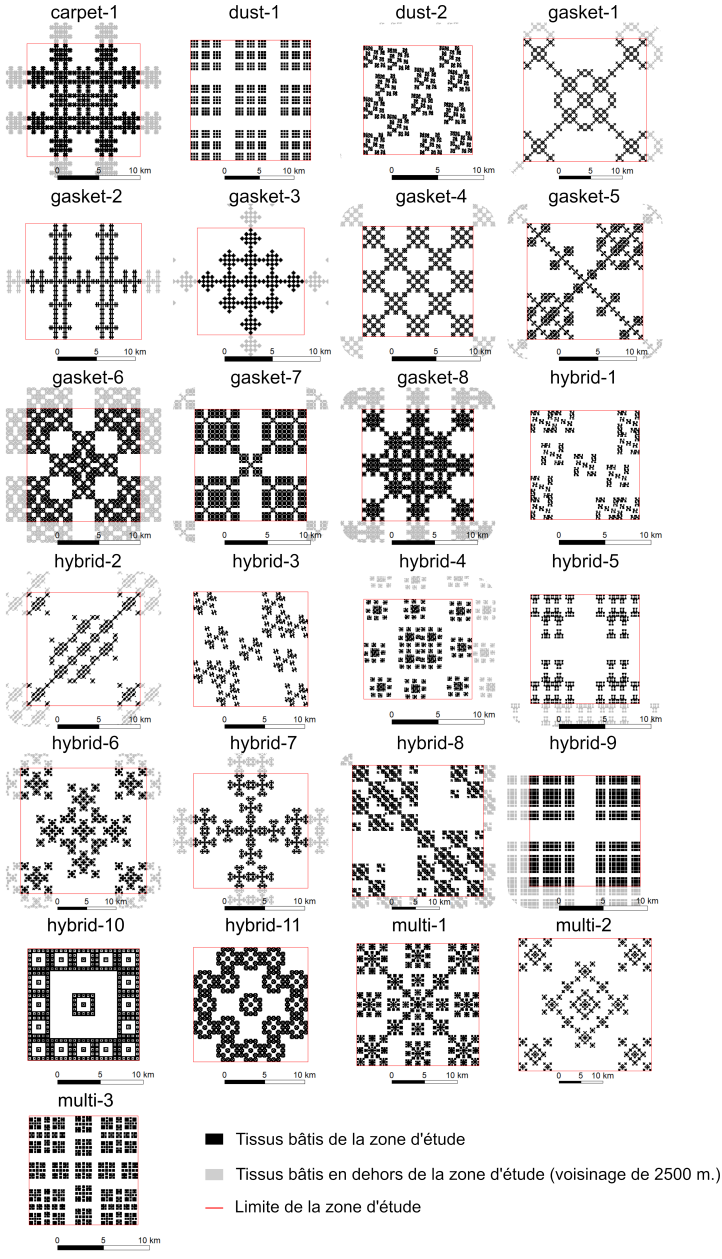


Figure 4.2. Les vingt-cinq tissus bâtis théoriques

4.2.2 Les tissus bâtis théoriques

Les tissus bâtis théoriques (Fig 4.2) ont été créés au moyen de systèmes de fonctions itérées (Iterated Function Systems – IFS) (Barnsley, 1988) via l'application *GenFrac* (https://sourcesup.renater.fr/www/genfrac/) selon un nombre d'itérations rapporté dans le Tableau 4.1. Les figures créées avec *GenFrac* ont ensuite été exportées en format shapefile (.shp). Afin que la surface moyenne des bâtiments corresponde à celle des tissus réels, une transformation affine a été effectuée a posteriori à l'aide d'un Système d'Information Géographique (SIG). A l'issue de cette transformation, l'étendue spatiale des tissus théoriques et leur nombre de bâtiments étaient souvent trop importants. Les tissus ont alors été tronqués pour n'en conserver qu'une partie. La troncature a été effectuée suivant la logique de construction d'un objet fractal, c'est-à-dire en conservant une partie de l'objet qui est identique à l'objet dans son ensemble.

Pour pouvoir comparer les résultats obtenus pour les tissus bâtis théoriques et ceux obtenus pour les tissus réels, il est nécessaire que ces tissus aient un certain nombre de caractéristiques communes : nombre de bâtiments, surface moyenne et médiane des bâtiments, densité bâtie et étendue spatiale des tissus. La diversité des caractéristiques de chaque tissu mais aussi les points communs entre les tissus réels, d'une part, et les tissus théoriques, d'autre part, sont observables sur la Fig. 4.3. Certains tissus théoriques sont des poussières fractales, d'autres des tapis, d'autre encore des figures hybrides ; 22 sont monofractals et 3 sont multifractals (Tab. 4.1) . Une bonne part des tissus théoriques ont une surface médiane des bâtiments légèrement supérieure à celle des tissus réels. Par définition, seuls les tissus multifractals présentent une diversité de taille des entités spatiales (carrés) élémentaires. Toutefois, les carrés élémentaires connexes de chaque figure monofractale n'ont pas été fusionnés. Or ils forment des bâtiments plus larges que l'entité spatiale (carré) élémentaire. La diversité de taille des bâtiments des tissus monofractals est donc en réalité plus importante que celle mesurée par la taille des carrés élémentaires.

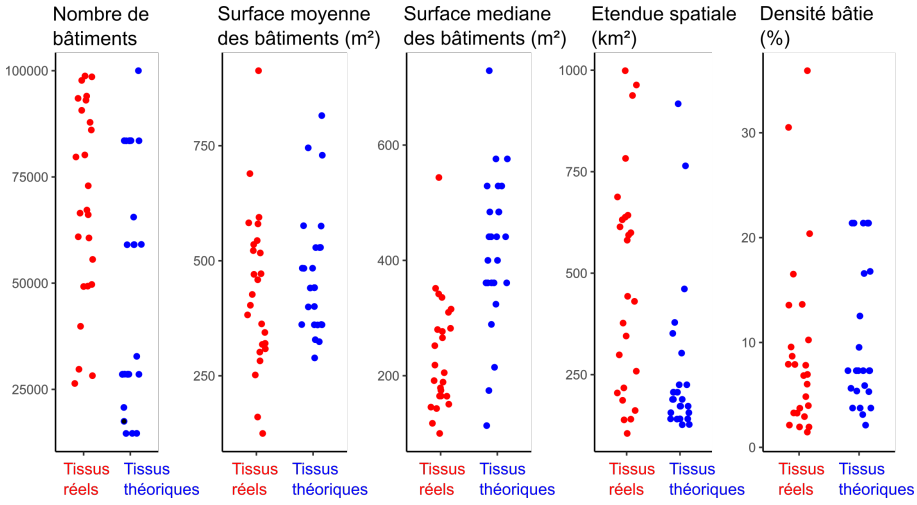


Figure 4.3. Caractéristiques des tissus étudiés

Dans le cas de tissus fractals théoriques, on pourrait penser qu'il est possible d'identifier les valeurs de paramètres qui permettent d'estimer au mieux leurs dimensions fractales de boîte D_B et de corrélation D_C, sachant que la dimension d'homothétie (ou de similarité) D_S doit coïncider avec celles-ci dans le cas d'objets fractals strictement auto-similaires. Cependant, cette égalité des trois dimensions est purement théorique ; elle suppose en effet de réaliser une couverture parfaite des figures, ce qui est rarement possible en pratique. Pour illustrer ceci, prenons l'exemple de la *fractale box* à l'itération 5 présentée en Figure 4.4. Chaque élément de la figure a une longueur égale à 1 mètre. On procède à une analyse fractale de corrélation en effectuant le comptage uniquement à partir des cinq points notés A sur la figure. La fenêtre de comptage est carrée. Sachant que le facteur de réduction ϵ de l'IFS de la fractale box est égal à $\frac{1}{3}$, la taille de la fenêtre d'analyse est multipliée par 3 d'une étape de comptage à une autre. Le résultat de l'analyse, considérant cinq tailles de fenêtres de comptage, est présenté en Figure 4.5a ; la dimension fractale estimée est D_C $1,46$ égale à la valeur de la dimension d'homothétie $D_S = \frac{log5}{log\frac{1}{3}}$, les points étant parfaitement alignés sur une droite. Dans un deuxième temps, on effectue le comptage en partant de tous les points de la figure, comme cela se fait pour les analyses fractales réalisées automatiquement au moyen de logiciels tels que *Fractalyse*. Le résultat de l'analyse est présenté en Figure 4.5b. Les points de mesure ne sont cette fois pas parfaitement alignés le long d'une droite ; la dimension fractale estimée est D_C $1,345$, donc différente de D_S. En effet, de nombreux points n'ont pas un nombre de voisins égal à cinq ou à un multiple de cinq

considérant les cinq tailles de fenêtres de comptage définies précédemment. Tels est le cas, par exemple, du point B noté sur la Figure 4.4.

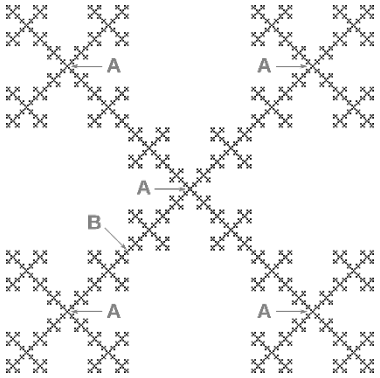


Figure 4.4. Une "Fractale box" à l'itération 5. IFS : $N = 5$ et $\epsilon = \frac{1}{3}$

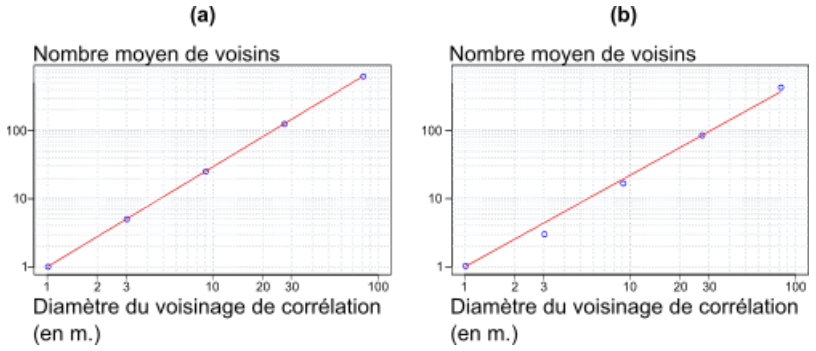


Figure 4.5. Analyses fractales de corrélation à partir des points A de la Figure 4.4 (a) et à partir de tous les points de la Figure 4.4 (b)

Intitulé	Type	D_S	Itération
hybrid-1	Fractale hybride plus proche d'une poussière que d'un tapis ou d'un tamis	1,4898	5
hybrid-2	Fractale hybride plus proche d'un tamis que d'une poussière	1,4898	5
hybrid-3	Fractale hybride plus proche d'une poussière que d'un tapis ou d'un tamis	1,4898	5
hybrid-4	Fractale hybride plus proche d'une poussière que d'un tapis ou d'un tamis	1,5439	5
hybrid-5	Fractale hybride plus proche d'une poussière que d'un tapis ou d'un tamis	1,5439	5
hybrid-6	Fractale hybride plus proche d'une poussière que d'un tapis ou d'un tamis	1,5849	6
hybrid-7	Fractale hybride plus proche d'un tamis que d'une poussière	1,5936	5
hybrid-8	Fractale hybride proche d'une poussière	1,6609	6
hybrid-9	Fractale hybride plus proche d'une poussière que d'un tapis ou d'un tamis	1,7227	5
hybrid-10	Fractale hybride plus proche d'un tapis que d'une poussière	1,7603	5
hybrid-11	Fractale hybride plus proche d'un tamis ou d'un tapis que d'une poussière	1,7603	5
gasket-1	Tamis fractal	1,4999	6
gasket-2	Tamis fractal	1,5936	5
gasket-3	Tamis fractal	1,5936	5
gasket-4	Tamis fractal	1,5936	5
gasket-5	Tamis fractal	1,5936	5
gasket-6	Tamis fractal	1,7603	5
gasket-7	Tamis fractal	1,7603	5
gasket-8	Tamis fractal	1,7603	5
dust-1	Poussière fractale	1,5849	6
dust-2	Poussière fractale	1,5936	5
carpet-1	Tapis fractal	1,7603	5
multi-1	Figure hybride multifractale	1,6547	6
multi-2	Figure hybride multifractale	1,4566	6
multi-3	Poussière multifractale	1,6747	5

Table 4.1. Présentation des tissus théoriques

4.2.3 Étendue des zones d'étude

Lors de la sélection des tissus bâtis réels et de la création des tissus bâtis théoriques, une attention particulière a été apportée à ce que tous aient des étendues géographiques proches (de même ordre de grandeur). En effet, l'étendue géographique d'un tissu bâti, autrement dit la taille de la zone d'étude, peut influencer notablement la dimension fractale estimée surtout lorsque des structures bâties différentes sont intégrées à mesure que la zone d'étude croît. A l'inverse, la dimension fractale variera très peu si des structures identiques sont intégrées. Comme le montre la Figure 4.6, les délimitations 1, 2 et 3 de la zone d'étude recouvrent une même structure spatiale ; les D_B estimées sont donc très proches voire égales pour ces trois zones d'étude. En revanche, la délimitation 4 de la zone d'étude intègre une structure spatiale différente, de laquelle résulte une D_B différente des dimensions obtenues pour les autres délimitations.

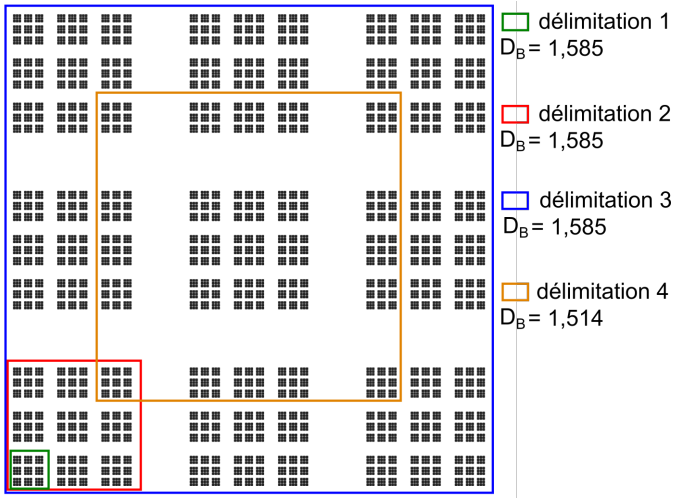


Figure 4.6. Variation de la dimension de boîte D_B estimée en fonction de la zone d'étude considérée

Il est assez facile de choisir la délimitation des zones d'étude des tissus théoriques de notre échantillon en contrôlant le type de structures spatiales qu'elles intègrent. En revanche, ceci est impossible dans le cas des tissus réels. C'est pourquoi, étant donné que la taille des fenêtres d'analyse influence directement le calcul des dimensions fractales, nous avons choisi de considérer des délimitations dont les étendues sont du même ordre de grandeur.

4.3 Choix des valeurs des paramètres

En vue d'assurer la comparabilité des résultats d'analyse des cinquante tissus bâtis, il nous faut choisir un seul et unique jeu de valeurs de paramètres qui sera systématiquement appliqué pour les analyses réalisées avec le logiciel *Fractalyse* (https://sourcesup.renater.fr/www/fractalyse/).

Pour aider au choix des valeurs de paramètres, des tests ont été effectués sur un sous-échantillon de dix tissus, cinq réels et cinq théoriques, parmi les cinquante contenus dans l'échantillon. Ces dix tissus ont été sélectionnés pour que ce sous-échantillon soit représentatif des différentes configurations spatiales présentes dans l'échantillon. Les dix tissus sélectionnés sont : Besançon, Charleroi, Paris, Pecs, Santiago, carpet1, gasket1, dust1, hybrid6, multi1. Le choix a été fait de retenir à la fois des tissus réels et des tissus théoriques. Ces derniers sont intéressants car leur dimension D_S est connue mais ils ne reflètent pas toute la complexité des tissus réels, pour lesquels les sources potentielles de variabilité des dimensions fractales estimées en fonction des valeurs de paramètres choisies peuvent être autres que dans le cas de tissus théoriques.

Trois grandes familles de paramètres peuvent avoir de l'influence sur le calcul de la dimension fractale de boîte ou de corrélation : (1) le format de l'image et sa résolution, (2) les caractéristiques des fenêtres de comptage, et (3) la méthode d'ajustement de la courbe théorique à la courbe empirique. Nous allons, dans cette partie du chapitre, passer en revue chacun des paramètres pouvant modifier la valeur de la dimension de boîte et de corrélation. Nous montrerons ainsi l'importance de choisir avec minutie les valeurs de ces paramètres. Cela nous permettra également de fixer les valeurs des paramètres pour l'analyse de nos cinquante tissus bâtis. Pour ce faire, nous démarrons l'analyse avec un jeu de valeurs de paramètres choisies par défaut. Nous faisons ensuite varier la valeur de chaque paramètre l'un après l'autre afin d'évaluer son impact sur le calcul de la dimension fractale et fixer la valeur qui nous semble la plus pertinente pour la suite de l'analyse. Dans le cas des tissus théoriques, nous choisissons les valeurs des paramètres pour que les D_B et D_C s'approchent de D_S. Nous sommes également attentifs à la significativité des dimensions obtenues à l'aide du R^2 et de la p-valeur de l'estimation.

4.3.1 Le format de l'image et sa résolution

Fractalyse permet d'opérer une analyse fractale de boîte sur format vectoriel et sur format raster. Il est donc utile de vérifier si l'utilisation d'un format plutôt qu'un autre influence la valeur de la dimension de boîte calculée. L'analyse fractale de corrélation ne peut, elle, se faire que sur format raster. Dans les deux cas, la résolution, c'est-à-dire la taille d'un pixel dans un format raster, influencera également la dimension fractale estimée. Pour la dimension de boîte, quatre tests ont été effectués : un test sur des fichiers vectoriels et trois sur des fichiers rasters avec des résolutions de 2 mètres, 4 mètres et 10 mètres. Les mêmes résolutions sont testées pour la dimension de corrélation (uniquement sur format raster).

Les résultats obtenus, présentés sur la Figure 4.7, montrent que pour une résolution assez grossière (pixels de 10 mètres de côté) la dimension de boîte est sensiblement différente (en général, plus faible) que pour des résolutions plus fines. C'est à résolution fine que la D_B des tissus théoriques se rapproche le plus de leur D_S. Ces deux tendances sont moins marquées pour la D_C pour laquelle une variation de résolution n'implique pas de variation particulièrement marquée de la dimension fractale. Les D_B obtenues sur format vectoriel sont légèrement supérieures à celles obtenues sur format raster à résolution fine et tendent globalement à s'éloigner un peu des D_S dans le cas des tissus théoriques. Ces résultats sont cohérents avec les observations de Ge and Lin (2009) qui notent que les analyses sur format raster sont meilleures que sur format vectoriel.

La suite des analyses se fera donc sur les fichiers en format raster à une résolution de 4 mètres. Avec ce format, les D_B et D_C sont effectivement proches des D_S pour les tissus théoriques. De plus, une résolution de 4 mètres, moins fine que celle de 2 mètres, permet de réduire le temps de calcul qui peut être important dans le cas des analyses de corrélation.

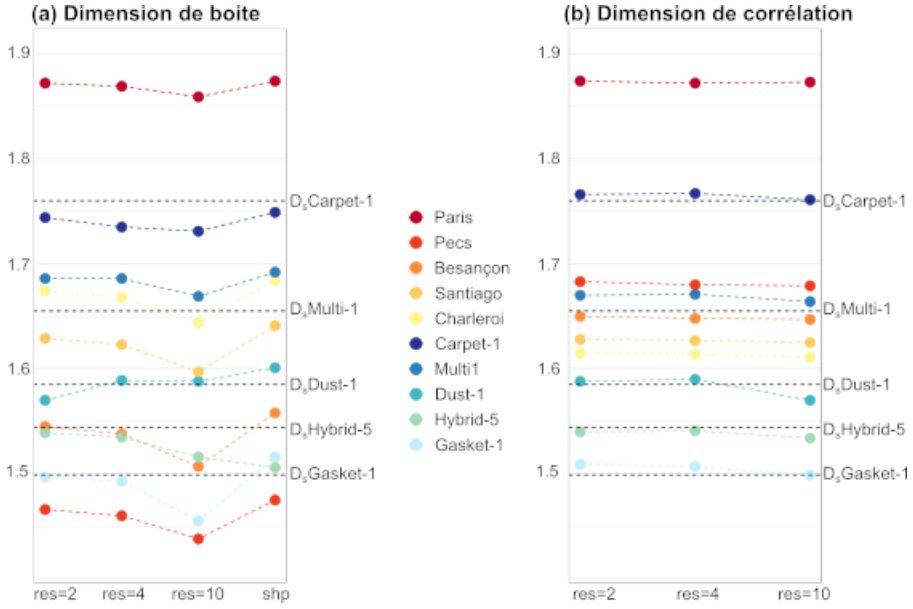


Figure 4.7. Variation des dimensions de boîte (a) et de corrélation (b) selon le format et la résolution de l'image source [2]

4.3.2 Les caractéristiques des fenêtres de comptage

Les fenêtres de comptage ont un rôle déterminant dans l'estimation des dimensions fractales. Pour un tissu bâti donné, il est clair que la taille, la position et la forme des boîtes qui le recouvrent vont directement impacter le calcul du nombre de boîtes occupées (cf. MAUP (Openshaw, 1983)) et, par conséquent, l'estimation de la dimension de boîte. Il en est de même pour la définition du voisinage des points de l'analyse de corrélation, qui impacte directement le nombre moyen de pixels occupés. Il est donc important de contrôler la définition des unités spatiales de base (BSUs), c'est-à-dire des fenêtres de comptage, pour une bonne estimation des dimensions fractales. Pour toutes les analyses, nous devons fixer une forme commune et des tailles identiques aux différentes fenêtres de comptage.

[2] Attention, les traits discontinus entre les points sur les graphiques ont uniquement vocation à faciliter la lecture du changement de valeur des dimensions en fonctions des valeurs de paramètre. Il ne s'agit en aucun cas des valeurs de dimensions prises pour des valeurs intermédiaires du paramètre.

Pour des raisons de facilité informatique, c'est le carré qui est la forme par défaut des fenêtres de comptage dans le programme *Fractalyse*. Les tissus sont recouverts d'une grille dont les éléments sont carrés pour la dimension de boîte. Pour la dimension de corrélation, le voisinage de chaque point est également défini dans un carré.

Concernant la taille des fenêtres de comptage, *Fractalyse* permet de fixer la taille minimale, la taille maximale et les tailles intermédiaires via un jeu de quatre paramètres : les paramètres « *min* » et « *max* » permettent de fixer respectivement la plus petite taille de fenêtre et la plus grande ; les paramètres « *coef* » et « *seq* » permettent ensemble de définir le nombre et la taille des fenêtres intermédiaires. Le paramètre « *coef* » est un coefficient qui détermine la taille de la fenêtre d'analyse à chaque étape en se basant sur la taille de la fenêtre à l'étape précédente. Ce coefficient est appliqué de manière additive dans le cadre d'une séquence arithmétique et de manière multiplicative dans le cas d'une séquence géométrique. La nature de la séquence est fonction du paramètre « *seq* ».

Le paramètre « *min* »

Le paramètre « *min* » fixe la taille minimale de ϵ dans les analyses de boîte et de corrélation (cf. section 3). Nous testons ici quatre tailles de fenêtre minimales : 4 m. (équivalent à la résolution de l'image), 10m., 20m. et 30m.

En fonction de la taille de la fenêtre minimale, de fortes variations de D_B pour des tissus réels sont visibles sur la Figure 4.8. Ces variations sont moins marquées mais bien présentes pour les tissus théoriques. Des variations des D_C sont également visibles. Il est clair, pour les cinq tissus théoriques, que D_C s'approche de D_S à mesure que la taille minimale s'approche des 30 mètres. Cela est également vrai pour trois tissus théoriques sur cinq avec la dimension de boîte mais un effet inverse est constaté pour le tissu multi-1.

Nous appliquerons donc, pour la suite des analyses, une taille minimale de fenêtre de comptage de 30 mètres. Ce choix est cohérent avec les préconisations trouvées dans la littérature. En effet, Benguigui et al. (2000) notent que la taille de la fenêtre minimale doit être au moins supérieure à celle du plus petit bâtiment. Dans notre base de données de 50 tissus bâtis, la plus grande taille minimale de bâtiment est de 576 m^2. Une fenêtre minimale d'analyse de 24 m ($24\times24 = 576$) serait donc la taille minimale de fenêtre acceptable. Shen (2002) confirme que la taille de la fenêtre minimale doit être supérieure à la résolution de l'image. Enfin, Thomas et al. (2010) précisent qu'une fenêtre de taille inférieure à 30 mètres capture les variations liées à la taille des bâtiments eux-mêmes et que cela peut provoquer des déviations

importantes des lois fractales estimées.

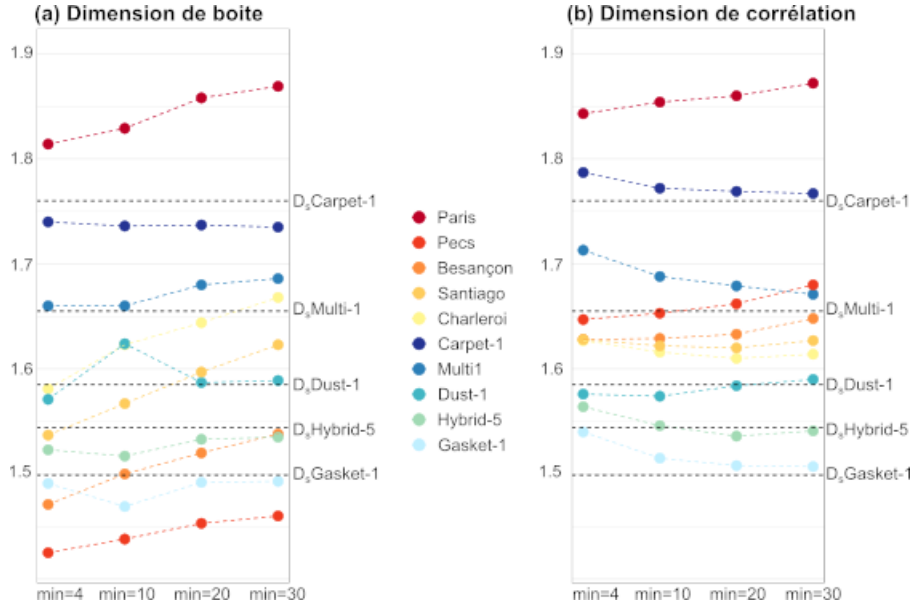


Figure 4.8. Variation des dimensions de boîte (a) et des dimensions de corrélation (b) selon la taille minimale de la fenêtre de comptage

Le paramètre « *max* »

Le paramètre « *max* » fixe la taille maximale de ϵ dans les analyses de boîte et de corrélation. Il est clair que la taille maximale de la fenêtre de comptage ne peut pas être supérieure à la taille de l'image elle-même. L'application *Fractalyse* utilise par défaut une fenêtre maximale de taille égale à la moitié du plus petit côté de l'image. Comme tous les tissus de la base de données n'ont pas tout à fait la même étendue spatiale (cf. Fig. 4.3) et, de surcroît, n'ont pas la même longueur pour leur côté le plus petit, il en découle que les maximums fixés par défaut ne sont pas les mêmes. Nous devons donc vérifier que l'adoption d'une taille maximale commune à tous les tissus ne modifie pas outre mesure la dimension fractale de ces tissus au regard de la dimension estimée avec la taille maximale par défaut.

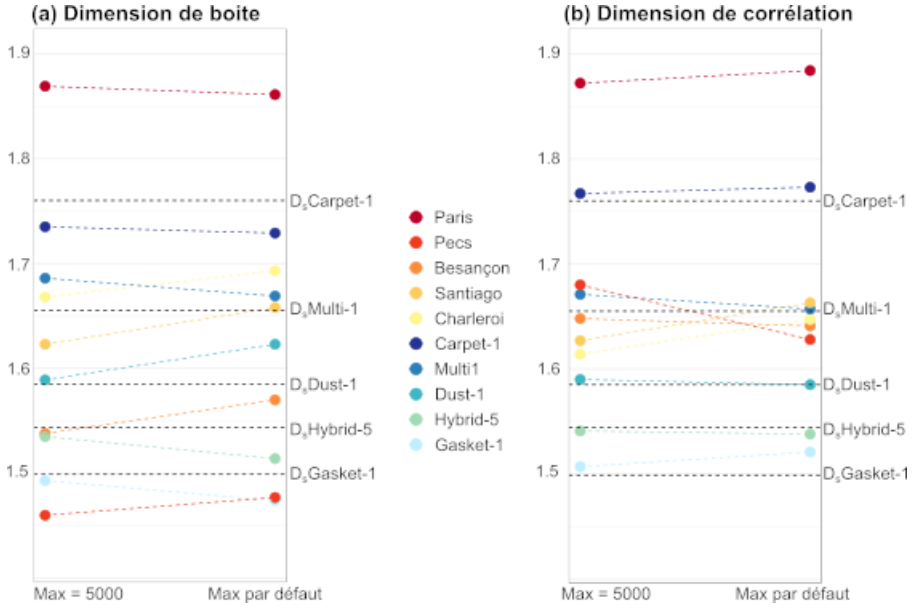


Figure 4.9. Variation des dimensions de boîte (a) et de corrélation (b) selon la taille maximale de la fenêtre de comptage

Sur les cinquante tissus bâtis, c'est le tissu de Liège qui a le plus petit côté avec une longueur légèrement supérieure à 10 000 mètres. Nous fixons donc le maximum commun à 5 000 mètres (soit 10 000 divisé par 2). Avec ce seuil, l'intervalle de taille des fenêtres de comptage couvre trois ordres de grandeur (ϵ va de 30 m à 5 000 m.)

Pour les tissus théoriques, la Figure 4.9 montre que les D_C sont moins influencées par la modification de la taille maximale de fenêtre que ne le sont les D_B. Notons qu'à l'exception du tissu multi-1, les D_B des tissus théoriques sont plus proches de leur D_S lorsque la taille limite est fixée à 5 000 m. Les dimensions fractales des tissus réels semblent, elles, davantage influencées par la modification du paramètre « *max* ». Les différences sont cependant assez faibles. En effet, la plus grande différence est constatée pour le tissu de Santiago, pour lequel la dimension mesurée avec le paramètre « *max* » par défaut est supérieur de 0,035 à la dimension mesurée avec le « *max* » égal à 5 000 m.

L'influence de la taille maximale de la fenêtre de comptage étant raisonnable, nous fixons ce paramètre à 5 000 mètres pour tous les tissus. Il est important de noter que les cinquante tissus bâtis de la présente étude ont

des étendues spatiales différentes mais du même ordre de grandeur. L'influence de la taille maximale de la fenêtre de comptage aurait peut-être été tout autre si un plus grand écart avait existé entre la plus petite étendue et la plus grande.

Le paramètre « *coef* »

Le paramètre « *coef* » fixe le nombre et la taille des fenêtres intermédiaires entre la valeur minimale (ici 30 m.) et la valeur maximale (ici 5 000 m.). Admettons une valeur de coefficient égale à 2. Si on décide d'appliquer une séquence géométrique (voir le paramètre « *seq* ») et puisque la première fenêtre de comptage à une taille de 30 m., la seconde aura une taille de 60 m. (30×2), la troisième une taille de 120 m. (60×2), etc. Il est illusoire de penser qu'un coefficient spécifique peut permettre de trouver exactement une valeur de D_B ou D_C égale ou proche de D_S pour tous les tissus, en particulier les tissus théoriques. En effet, même en faisant abstraction de la taille minimale de la fenêtre de comptage (qui influence inévitablement les tailles intermédiaires), il faudrait que le coefficient soit égal au facteur de réduction qui a permis de construire chaque tissu. Ces facteurs de réduction étant différents d'un tissu à l'autre, il est impossible qu'un coefficient unique soit totalement satisfaisant pour tous les tissus. Nous pouvons cependant chercher un coefficient qui permette d'approcher au mieux D_S pour un maximum de tissus théoriques.

Trois valeurs de coefficient (1,5 ; 2 et 2,5) sont testées. A nouveau, les D_C sont plus stables que les D_B (Fig 4.10). Des différences marquées apparaissent entre les D_B des tissus théoriques. C'est pour une valeur de coefficient de 1,5 que les D_B sont généralement plus proches de leur D_S (sauf pour le tissu carpet-1). C'est donc la valeur de 1,5 qui sera fixée pour le paramètre « *coef* » dans la suite des analyses.

Notons que la qualité des ajustements, mesurée par le R^2 et les intervalles de confiance, n'est pas fondamentalement impactée par le nombre de fenêtres intermédiaires (11 pour coef = 1,5 ; 8 pour coef = 2 ; 6 pour coef = 2,5). Seule la p-valeur est systématiquement plus faible avec un coefficient plus petit.

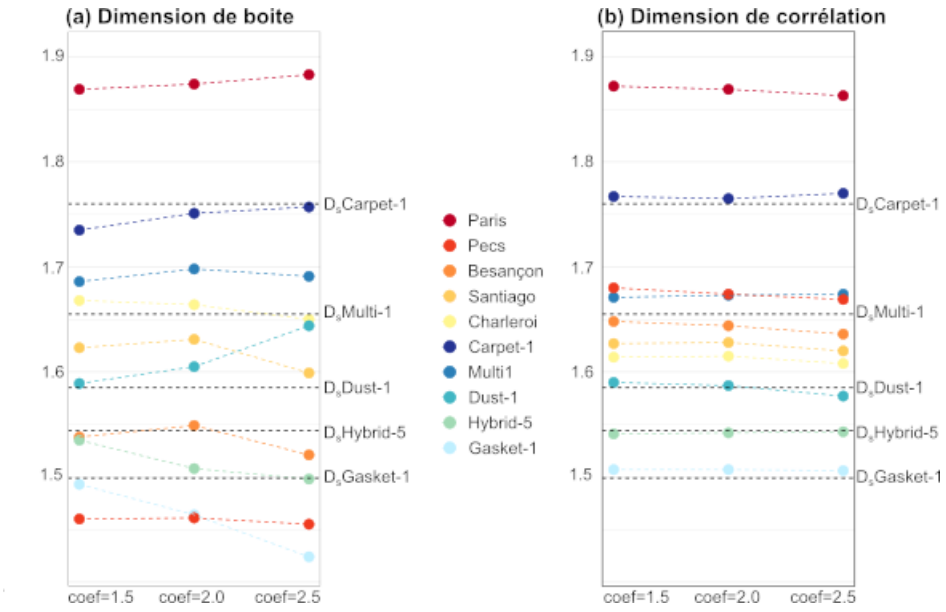


Figure 4.10. Variation des dimensions de boîte (a) et de corrélation (b) selon le coefficient déterminant la taille des fenêtres intermédiaires de comptage

Le paramètre « *seq* »

Le paramètre « *seq* » permet de fixer la séquence de tailles des fenêtres intermédiaires. Le paramètre « *coef* » peut en effet être appliqué suivant une progression arithmétique (i. e. additive) ou géométrique (i. e. multiplicative). La progression arithmétique n'est pas en accord avec la logique fractale, qui est par définition multiplicative. Il est donc plus adéquat d'utiliser une séquence géométrique qu'une séquence arithmétique. Cela est confirmé par les résultats des tests (Fig 4.11). Mise à part la D_C du tissu multi-1, toutes les D_B et D_C sont plus proches de leur D_S avec une séquence géométrique. Le nombre bien plus important de points de mesure avec la séquence arithmétique n'implique donc pas un meilleur ajustement.

La pertinence d'utiliser une séquence géométrique, plus propice à analyser un processus fractal que ne l'est une séquence arithmétique, peut encore être démontrée à l'aide de la Figure 4.12 qui présente les droites d'ajustement de D_B pour le tissu théorique dust-1. La progression arithmétique implique que pour des distances ϵ différentes mais proches l'une de l'autre, le nombre de pixels bâtis $N(\epsilon)$ soit identique. Cela engendre du bruit dans l'estimation de la droite déterminant la D_C. Ce type de bruit ne se retrouve pas lorsque la progression est géométrique. C'est donc bien cette progression géométrique

qui sera utilisée pour la suite des analyses.

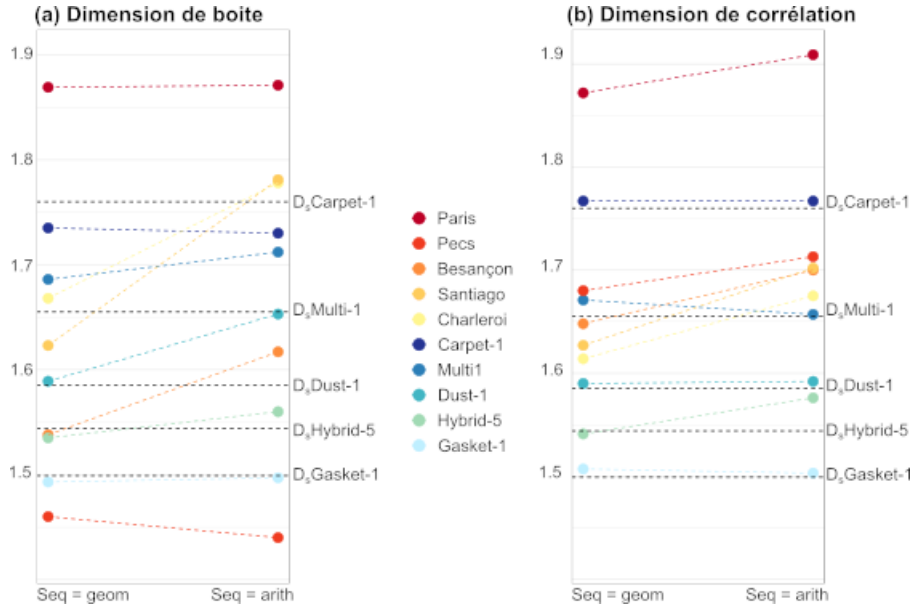


Figure 4.11. Variation des dimensions de boîte (a) et de corrélation (b) selon le type de séquence (géométrique ou arithmétique) déterminant la taille des fenêtres intermédiaires de comptage

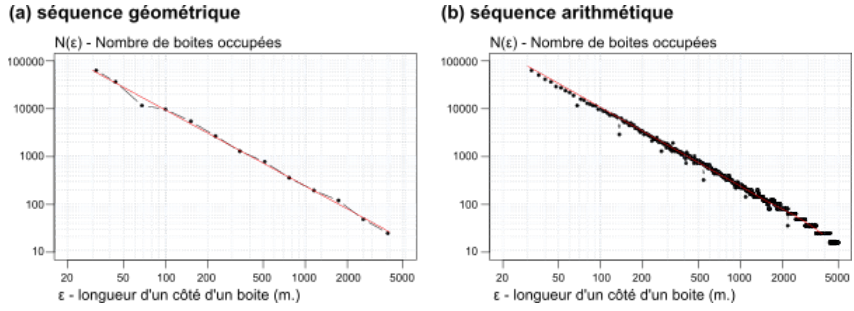


Figure 4.12. Analyse fractale de boîte des distributions spatiales des bâtiments du tissu théorique « Dust-1 » : séquence géométrique (a) et arithmétique (b)

Concernant la dimension de corrélation, il importe de prendre en compte les effets de bords (Agterberg, 2013). En effet, un pixel en bordure de tissu risque de se voir privé d'une partie de son voisinage si celui-ci est situé en dehors de l'image. Dès lors, à mesure que le voisinage ϵ augmente, le nombre

moyen de pixels N_ϵ est sous-estimé puisque de plus en plus de voisins situés hors de la zone d'étude ne sont pas pris en compte dans le comptage. Plus le voisinage sera grand, plus la dimension de corrélation sera biaisée vers des valeurs inférieures (Carlson, 1991). Pour contrer cet effet de bord, il est possible de définir une bande autour du tissu bâti étudié dont les pixels bâtis doivent être pris en compte dans les comptages de points dans le voisinage mais ne sont pas directement concernés par le calcul de la dimension. Admettons un pixel bâti (noir) en bordure d'une zone d'étude, soit le pixel A sur la Figure 4.13. Le pixel B, en dehors de la zone d'étude, est comptabilisé dans le voisinage du pixel A. En revanche, le nombre de points dans le voisinage du pixel B n'est pas compté ni pris en compte puisque le point B n'est pas dans la zone d'étude. Dans le programme *Fractalyse*, la fonction *border* permet de définir une bande autour de la zone d'étude, d'une largeur fixée par l'utilisateur, afin que les pixels de cette bande puissent être comptabilisés dans les voisinages des points situés à l'intérieur de la zone d'étude.

Etant donné que la fenêtre encadrant un pixel i dans le calcul d'une dimension de corrélation peut avoir un côté ϵ maximum de 5 000 m, cela signifie que tout pixel j à une distance de 2 500 m (5 000 $\div$ 2) doit pouvoir être considéré comme un voisin de i. C'est pourquoi il faut tenir compte pour chaque zone d'étude d'une bande d'au minimum 2 500 m. de large. La valeur du paramètre border est donc fixée à 2 500m (cf. Fig 4.1 et 4.2).

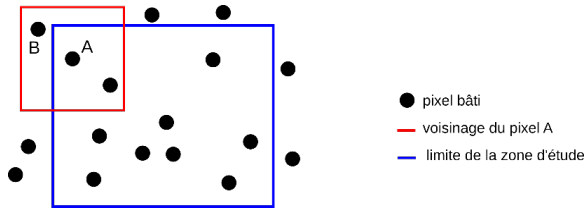


Figure 4.13. Illustration de la fonction border qui permet de prendre en compte les effets de bord de l'analyse de corrélation dans Fractalyse

Concernant le calcul des dimensions de boîtes, la position de la grille de comptage est susceptible d'influencer le résultat obtenu. Il est cependant difficile de contrôler ce paramètre. Avec l'option « *position optimale de la grille* », *Fractalyse* permet de mener des analyses sur un nombre prédéfini de grilles différentes. Le résultat retenu est celui pour lequel la position de la grille est jugée optimale, c'est-à-dire celle qui couvre l'image avec le moins de mailles. Cette option n'est disponible que pour l'analyse des fichiers en format vectoriel et non raster. Nous avons effectué des tests sur les formats vectoriels de nos

tissus échantillons et ceux-ci tendent à montrer que les dimensions varient très peu si la position de la grille est optimisée ou pas. Ce résultat contre intuitif (on s'attend à avoir de meilleurs résultats pour un grand nombre de grilles) s'explique par le fait que la position de la grille n'est pas identique entre deux étapes de comptage. Dès lors, une grille peut s'approcher de l'optimum à une étape et s'en éloigner à l'étape suivante. Etant donné l'absence de différences significatives, la méthode des boîtes utilisant une seule grille a été retenue pour les analyses qui suivent.

4.3.3 L'ajustement de la courbe théorique à la courbe empirique

Une opération de comptage aboutit à la création d'une courbe empirique présentant le nombre d'entités dénombrées en fonction de la taille des fenêtres d'analyse (nombre de boîtes comportant au minimum un pixel bâti pour la dimension de boîte ; nombre moyen de pixels bâtis dans un certain voisinage pour la dimension de corrélation). Pour estimer la dimension fractale, il est ensuite nécessaire d'ajuster une courbe théorique à cette courbe empirique. Dans le logiciel *Fractalyse*, il est possible d'effectuer cet ajustement de manière directe (forme exponentielle) ou en passant par l'intermédiaire d'une transformation logarithmique de la courbe empirique.

Le type d'ajustement utilisé influence grandement la valeur des dimensions de boîte et de corrélation (Fig. 4.14). Pour ces deux dimensions, c'est l'ajustement après transformation logarithmique qui produit les D_B et les D_C les plus proches des D_S des tissus théoriques. Cela s'explique notamment par le fait que, via un ajustement direct (fonction exponentielle), un poids trop important est donné aux observations faites à grandes échelles. C'est donc un ajustement logarithmique qui doit être privilégié.

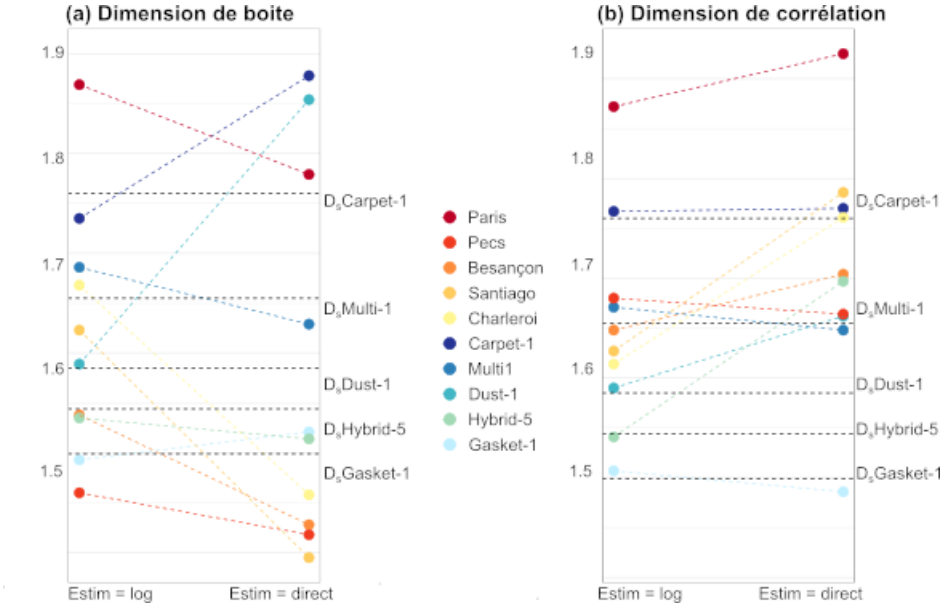


Figure 4.14. Variation des dimensions de boîte (a) et de corrélation (b) selon le type d'ajustement (logarithmique ou direct)

4.3.4 Récapitulatif des valeurs de paramètres choisies pour les analyses

Le Tableau 4.2 présente les valeurs de paramètres choisies pour le calcul des dimensions de boîte et de corrélation des cinquante tissus bâtis (vingt-cinq tissus théoriques et vingt-cinq tissus réels).

Fichier input (données d'entrée)	Format	Raster (.TIF)
	Résolution	4 m.
Fenêtre de comptage	Forme	Carrée
	Taille minimale	30 m.
	Taille maximale	5000 m.
	Coefficient de croissance	1,5
	Type de séquence de croissance	Géométrique
Ajustement	Estimation	Logarithmique

Table 4.2. Récapitulatif des valeurs de paramètres choisies pour les analyses de boîte et de corrélation

Les dimensions de boîtes sont calculées sur la base d'images reprenant uniquement les bâtiments de la zone d'étude. Pour les dimensions de corrélation, afin de tenir compte des effets de bord, les bâtiments situés en dehors de la zone d'étude sur une largeur de 2 500 mètres sont également pris en compte comme éléments du voisinage. Le paramètre *border*, spécifique au calcul de la dimension de corrélation dans *Fractalyse*, est fixé à 625 pixels.

4.4 Comparaison des dimensions de boîte et de corrélation des cinquante tissus bâtis

Le Tableau 4.3 présente les dimensions de boîte et de corrélation obtenues pour les cinquante tissus bâtis avec les valeurs de paramètres fixées précédemment.

Tissus	D_B	D_C	**Tissus**	D_B	D_C
Ahmedabad	1,414	1,668	carpet-1	1,735	1,767
Ansbach	1,385	1,548	dust-1	1,589	1,59
Berne	1,566	1,666	dust-2	1,622	1,597
Besançon	1,538	1,648	gasket-1	1,493	1,51
Bogota	1,531	1,512	gasket-2	1,581	1,606
Bolton	1,696	1,736	gasket-3	1,577	1,593
Bruxelles	1,922	1,885	gasket-4	1,566	1,594
Charleroi	1,668	1,614	gasket-5	1,594	1,581
Clermont-Ferrand	1,654	1,729	gasket-6	1,75	1,777
Cochabamba	1,702	1,764	gasket-7	1,721	1,734
Colorado	1,706	1,654	gasket-8	1,73	1,762
Haag in Oberbayern	1,292	1,369	hybrid-1	1,485	1,471
Ingolstadt	1,512	1,673	hybrid-2	1,465	1,509
Liège	1,759	1,726	hybrid-3	1,505	1,486
Lyon	1,839	1,799	hybrid-4	1,568	1,571
Mexico	1,704	1,788	hybrid-5	1,535	1,54
Paris	1,869	1,871	hybrid-6	1,588	1,592
Pecs	1,46	1,68	hybrid-7	1,612	1,607
Pittsburg	1,776	1,727	hybrid-8	1,704	1,667
Rome	1,83	1,779	hybrid-9	1,675	1,714
Santiago	1,623	1,627	hybrid-10	1,725	1,715
Sheffield	1,756	1,713	hybrid-11	1,758	1,749
Sherbrooke	1,549	1,651	multi-1	1,686	1,656
Timisoara	1,538	1,661	multi-2	1,464	1,444
Yamoussoukro	1,761	1,802	multi-3	1,697	1,646

Table 4.3. Dimension de boîte D_B et de corrélation D_C des cinquante tissus bâtis

4.4.1 Des classifications pour observer les hiérarchies

Il serait assez fastidieux d'analyser chaque tissu un à un. C'est pourquoi nous décidons de rassembler par groupes les tissus qui présentent des dimensions de boîtes identiques, d'une part, et des dimensions de corrélation identiques, d'autre part. Afin de regrouper entre eux des tissus théoriques et réels qui se ressemblent au regard de leurs dimensions fractales, deux classifications sont réalisées. La première se base sur la distance euclidienne standardisée entre les dimensions de boîte de tous les tissus bâtis. La deuxième classification fait de même pour les dimensions de corrélation. Dans les deux cas, la méthode de discrétisation choisie est celle de Ward afin de minimiser les écarts entre les dimensions d'un même groupe (intra-groupe) et de maximiser les écarts entre les groupes (inter-groupe).

Le dendrogramme de la Figure 4.15a associé aux valeurs de dimension de boîte du Tableau 4.3 permet de distinguer sept groupes de tissus. Le premier groupe est constitué de trois tissus bâtis réels pour lesquels les dimensions de boîte sont les plus faibles (entre 1,292 et 1,414). Ce groupe fusionne assez vite avec les groupes 2 et 3 qui ont des dimensions de boîte très similaires (respectivement de 1,460 à 1,535 et de 1,531 à 1,549). Les tissus réels de ces groupes sont Pecs, Ingolstadt (groupe 2) et Sherbrooke, Besançon, Timisoara et Bogota (groupe 3). Le groupe 2 est majoritairement composé de tissus théoriques (hybrid-1-2-3 ainsi que gasket-1 et multi-2). Le groupe 3 ne comporte qu'un seul tissu théorique (hybrid-5). Le groupe 4 est le dernier groupe qui peut être assimilé à des dimensions de boîte « faibles » (inférieures à 1,624). Il est constitué de neuf tissus théoriques (hybrid-4-6-7, gasket-2-3-4-5 et dust-1-2) et de deux tissus réels (Santiago et Berne). Au total, dans ces quatre groupes, il y a onze tissus réels et quinze tissus théoriques qui ont les dimensions de boîte les plus faibles de l'échantillon. Il y a donc, dans les groupes 5, 6 et 7, quatorze tissus réels et dix tissus théoriques ayant les dimensions de boîte les plus élevées. Le groupe 5 rassemble des tissus bâtis dont la dimension de boîte varie entre 1,654 et 1,706. Il s'agit des tissus de Bolton, Mexico, Cochabamba, Colorado, Charleroi et Clermont-Ferrand ainsi que des tissus multi-1, multi-3, hybrid-8 et hybrid-9. Le groupe 6 regroupe quatre tissus réels (Liège, Sheffield, Yamoussoukro, Pittsburg) et six tissus théoriques (hybrid-10-11, gasket-1-6-8 et carpet-1). Enfin, les tissus ayant les plus hautes dimensions de boîte (>1,830) sont dans le groupe 7 ; il s'agit de Rome, Lyon, Paris et Bruxelles. Au final, on constate que l'échantillon de tissus réels et de tissus théoriques couvre bien le spectre des dimensions de boîte (répartition homogène entre les faibles et les hautes valeurs de dimension).

Une analyse similaire peut être faite à partir du dendrogramme de la Figure 4.15b et des valeurs de la dimension corrélation du Tableau 4.3. Le dendrogramme suggère la présence de quatre groupes aux dimensions de corrélation plus faibles (groupes 1 à 4) et de trois groupes aux dimensions de corrélation plus élevées (groupes 5, 6 et 7). Le groupe 1 rassemble les tissus de Haag in Oberbayern, hybrid-1 et multi-2 avec des dimensions de corrélation inférieures à 1,499. Le groupe 2 a des dimensions de corrélation qui varient entre 1,5 et 1,548 (Ansbach, Bogota, Hybrid-2-3-5 et gasket-1). Avec une D_C entre 1,571 et 1,614, le groupe 3 est presque exclusivement composé de tissus théoriques (gasket-2-3-4-5, hybrid-4-6-7, dust-1-2) et seulement un tissu réel (Charleroi). Le groupe 4 est composé de neuf tissus réels (Santiago, Sherbrooke, Besançon, Colorado, Ahmedabad, Berne, Timisoara, Pecs et Ingolstadt) et de quatre tissus théoriques (multi-1-3, hybrid-8-10). Ainsi, treize tissus réels sont dans la catégorie des dimensions de corrélation plus faibles contre dix-neuf tissus théoriques. Il y a donc un léger déséquilibre dans l'échantillon de tissus théoriques dont la grande majorité ont des dimensions de corrélation relativement faibles. Le groupe 5 (Pittsburg, Liège, Clermont-Ferrand, Bolton, Sheffield, gasket-7 et hybrid-9) a des dimensions de corrélation comprises entre 1,713 et 1,736. Dans le groupe 6, Rome, Mexico, Yamoussoukro, Lyon, Cochabamba, gasket-6 et 8 ainsi que carpet-1 et hybrid-11 ont des dimensions de corrélation qui varient entre 1,751 et 1,802. Enfin, le groupe 7 ne contient que les tissus de Paris et Bruxelles avec des dimensions de corrélation respectivement de 1,871 et 1,885.

Des différences assez marquées apparaissent entre les deux classifications. Tout d'abord au niveau des valeurs de dimensions, la dimension de corrélation présente des limites de classes toujours légèrement supérieures à celles de la dimension de boîte. Ensuite, au niveau de la hiérarchie entre les tissus bâtis, des tissus renseignés comme ayant une dimension de boîte assez élevée (respectivement basse) peuvent être identifiés comme ayant une dimension de corrélation assez basse (respectivement élevée). C'est le cas, par exemple, du tissu de Charleroi qui est dans le groupe 5 des dimensions de boîte mais dans le groupe 3 des dimensions de corrélations. Admettons que les tissus soient classés de 1 (la plus faible valeur de dimension) à 50 (la valeur de dimension la plus élevée). Le tissu de Charleroi est à la 28ème position pour D_B mais seulement à la 19ème position pour D_C. A l'inverse, un tissu comme celui d'Ahmedabad passe de la 3ème place (groupe 1) pour D_B à la 30ème place (groupe 4) avec D_C. Ces bouleversements de hiérarchie sont également observables pour des tissus théoriques mais de façon moindre. Par exemple, dust-1 est à la 22ème place pour D_B (groupe 4) et à la 12ème place pour D_C (groupe 3). Dans l'autre sens (rang du tissu dans D_C supérieur à celui dans D_B), il n'y a presque pas de modification de hiérarchie. La modification la

plus marquée est celle du tissu hybrid-9 qui « perd » cinq places entre la hiérarchie de D_C et celle de D_B.

En résumé, des hiérarchies assez différentes apparaissent selon que la dimension fractale est calculée par boîte ou par corrélation. Ces différences sont surtout marquées pour les tissus réels. Cette observation devrait inviter les chercheurs à la prudence lorsqu'ils classent des entités géographiques (municipalités, district, région, etc.) selon leur dimension fractale ou qu'ils comparent la valeur d'une dimension fractale à d'autres indicateurs, cf. par exemple Shen (2002). En effet, l'impact de la méthode d'analyse fractale choisie sur les résultats obtenus est important.

4.4.2 Dimensions et caractéristiques morphologiques

Les dimensions de boîte et de corrélation ont en théorie des valeurs identiques pour des objets strictement auto-similaires. La Figure 4.16 montre en effet qu'il y a une corrélation linéaire claire entre les deux dimensions. Cependant, l'égalité entre les deux dimensions est loin d'être parfaite[3] surtout pour les tissus réels. Certains tissus bâtis ont une dimension de corrélation substantiellement supérieure à celle de boîte. Pour huit tissus réels, ces différences de dimension atteignent des valeurs supérieures ou égales à 0,1. La plus grande différence, d'une valeur de 0,254, est calculée sur le tissu bâti d'Ahmedabad. A l'inverse, certains tissus ont des dimensions de corrélation légèrement inférieures à leur dimension de boîte. Cette tendance est cependant moins forte. Les deux plus grandes différences sont ici calculées pour les tissus bâtis de Charleroi (différence = 0,054) et Colorado Springs (0,052). Ces différences entre la dimension de boîte et celle de corrélation sont à l'origine des hiérarchies différentes relevées précédemment.

[3]Notons ici qu'il est difficile d'évaluer la différence entre deux dimensions. Une différence de dimension de 0,05 représente-t-elle une vraie différence ou est-elle uniquement liée à des imprécisions ou des biais d'estimation ? Nous insistons donc sur les différences relativement les plus importantes.

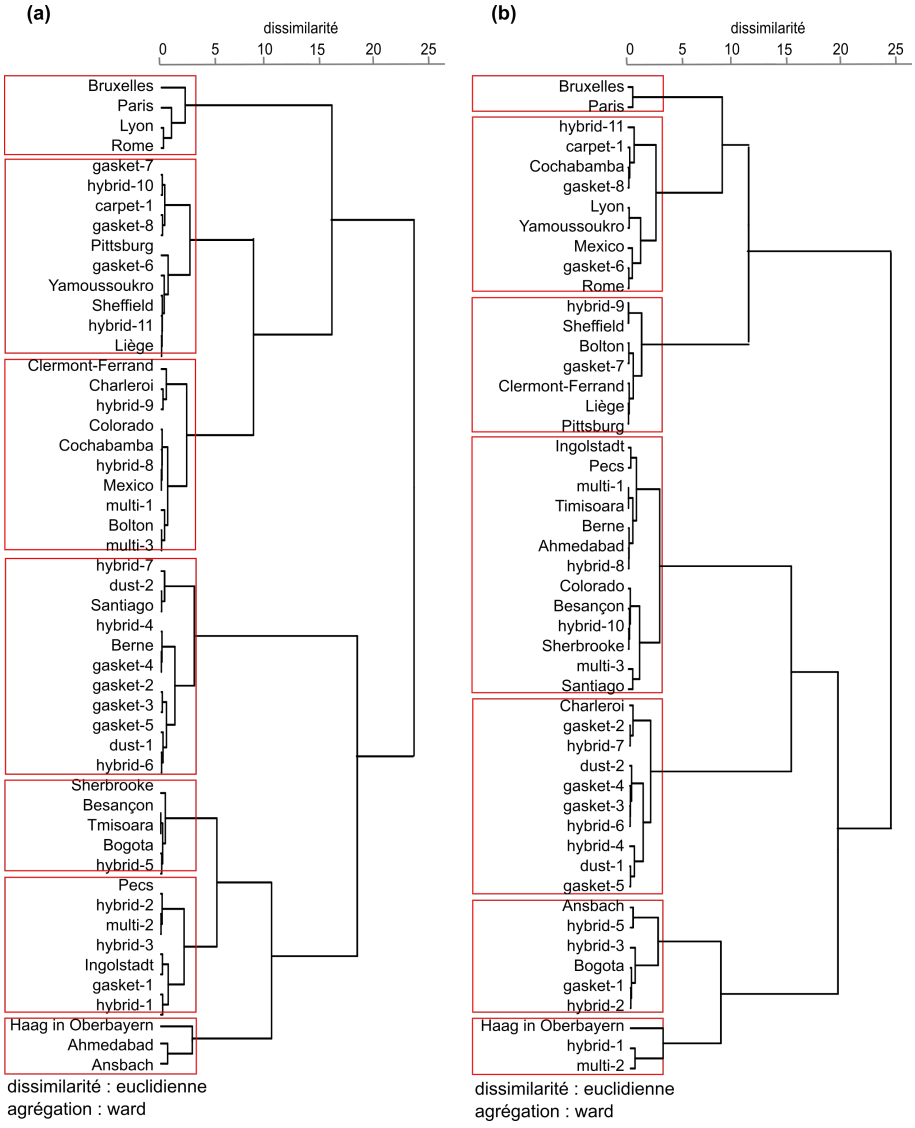


Figure 4.15. Classification des 25 tissus réels et des 25 tissus théoriques : dendrogramme basé sur la distance euclidienne entre les dimensions de boîte (à gauche) et les dimensions de corrélation (à droite)

Nous allons essayer de comprendre ce qui engendre les différences entre les deux dimensions. Pour ce faire, nous allons comparer les valeurs des dimensions avec celles d'autres indicateurs morphologiques caractérisant les cinquante tissus bâtis. Puisque les dimensions fractales décrivent en partie la quantité d'espace occupé, nous les comparons à l'indice de densité bâtie. La Figure 4.17 montre qu'il y a une relation positive entre les dimensions et le logarithme de la densité bâtie. Cette relation est davantage marquée entre la densité bâtie et la dimension de boîte (R^2 d'une régression simple = 0,84 avec tous les tissus et 0,94 avec uniquement les tissus réels) qu'entre la densité et la dimension de corrélation (R^2 d'une régression simple = 0,52). La dimension de boîte serait donc comparativement plus sensible à la densité bâtie que ne l'est la dimension de corrélation. La dimension de box-counting apporte donc peu d'information supplémentaire par rapport à la densité bâtie. Notons que cela n'est vrai qu'à l'echelle de l'agglomération (échelle de notre mesure) puisque la densité bâtie peut varier très fort selon l'échelle à laquelle elle est calculée. Cela n'est pas le cas des dimensions de boîte et de corrélation qui décrivent le tissu bâti à travers les échelles.

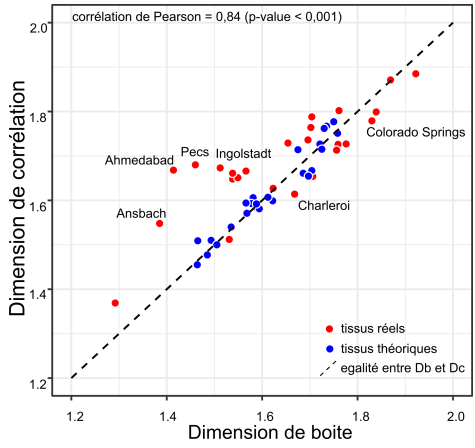


Figure 4.16. Relation entre les dimensions de boîte et les dimensions de corrélation des cinquante tissus bâtis

La Figure 4.17 montre également que pour une même densité bâtie, les tissus réels ont des dimensions de boîte et de corrélation systématiquement supérieures à celles des tissus théoriques. Ceci peut s'expliquer à l'aide d'un autre indice morphologique : le *Landscape Shape Index (LSI)*. Cet indice est beaucoup utilisé en écologie du paysage car il permet de décrire la complexité d'un paysage en le comparant à la figure parfaite du cercle (Gyenizse et al., 2014). Cet indice est basé sur le rapport entre le périmètre (P) et l'aire (A)

de l'objet étudié (voir équation 4.1). Plus l'indice est élevé, plus le périmètre est long au regard de la surface, et donc plus le paysage est tortueux et complexe. Bien que nous nous soyons efforcés de concevoir des tissus théoriques qui épousent les caractéristiques des tissus réels, ces derniers sont plus complexes en témoigne la Figure 4.18. Les dimensions de boîte et de corrélation sont donc également sensibles à la complexité des tissus bâti.

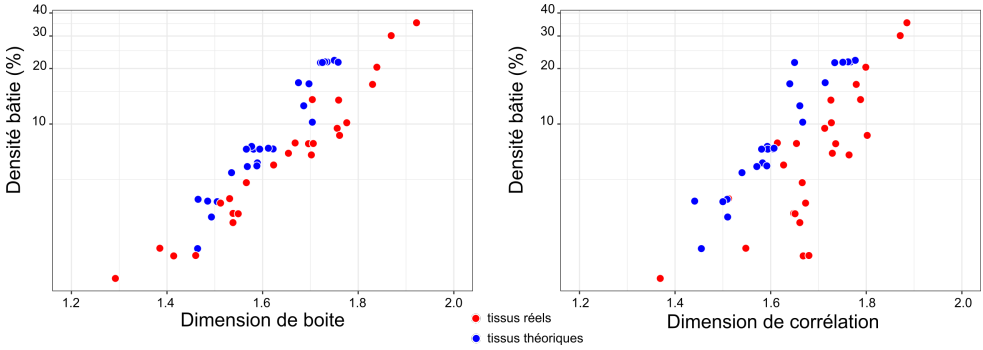


Figure 4.17. Relation entre le logarithme de la densité bâtie et les dimensions de boîte (gauche) et de corrélation (droite) des cinquante tissus bâtis

$$LSI = \frac{P}{(2 + (A \times \pi)^{1/2})} \tag{4.1}$$

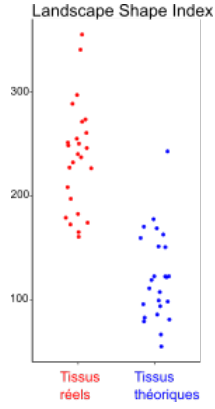


Figure 4.18. Landscape Shape Index des tissus réels et théoriques

Dans la section 4.2.3, nous avons vu que la délimitation des zones d'étude des tissus de notre échantillon est susceptible d'influencer les valeurs de dimensions fractales estimées. Afin de réduire cette possible influence, nous avons choisi de considérer des zones d'étude dont les étendues sont du même ordre de grandeur. Concernant les tissus bâtis théoriques, nous avons en outre contrôlé la position de chaque zone d'étude pour éviter que les structures spatiales fractales concernées ne soient tronquées. Ces précautions prises, la Figure 4.19 montre que, dans le cas des tissus bâtis réels, la différence de valeur entre D_B et D_C a tendance à être d'autant plus grande que les zones d'étude sont étendues.

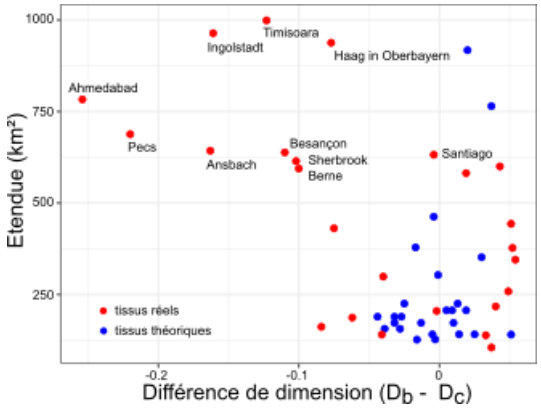
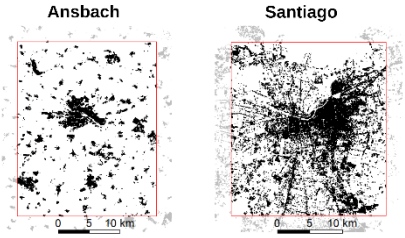


Figure 4.19. Relation entre la différence des dimensions de boîte et de corrélation et l'étendue des cinquante tissus bâtis

Prenons par exemple les tissus d'Ansbach et Santiago dont les étendues spatiales et les densités bâties sont assez comparables mais le LSI est différent, celui de Santiago, étant plus élevé (Fig 4.20). Leurs dimensions de corrélation sont assez proches bien que le tissu de Santiago, étant plus complexe (LSI plus grand), ait une dimension de corrélation légèrement plus élevée que celle d'Ansbach. En revanche, leurs dimensions de boîte sont bien différentes, la dimension de boîte d'Ansbach étant nettement inférieure à sa dimension de corrélation. Le tissu bâti de Santiago apparaît plus homogène tandis que celui d'Ansbach comporte un nombre important d'espaces non bâtis de tailles variées et dont la distribution spatiale est irrégulière. Ce pourrait donc être la présence d'espaces non bâtis de tailles et de distribution spatiale irrégulière qui provoque des différences entre la dimension de boîte et de corrélation. Une telle configuration apparaitrait d'autant plus souvent que les zones d'étude considérées sont grandes. La Figure 4.21, qui reprend les huit tissus bâtis réels dont la dimension de corrélation est supérieure à la dimension de boite

et dont l'étendue des zones d'étude est relativement large, tend à conforter visuellement cette hypothèse. Il ne nous est malheureusement pas possible de confirmer mathématiquement notre hypothèse. Il est en réalité très difficile de caractériser les espaces vides (lacunarités). Nous avons utilisé un grand nombre d'indices empruntés à l'écologie du paysage mais aucune corrélation significative n'est apparue.



	Ansbach	Santiago
Etendue spatiale (km²)	643	631
Densité bâtie (%)	2,12	5,9
Landscape Shape Index	161	271
D_B	1,385	1,548
D_C	1,623	1,627

Figure 4.20. Comparaison des tissus d'Ansbach et Santiago

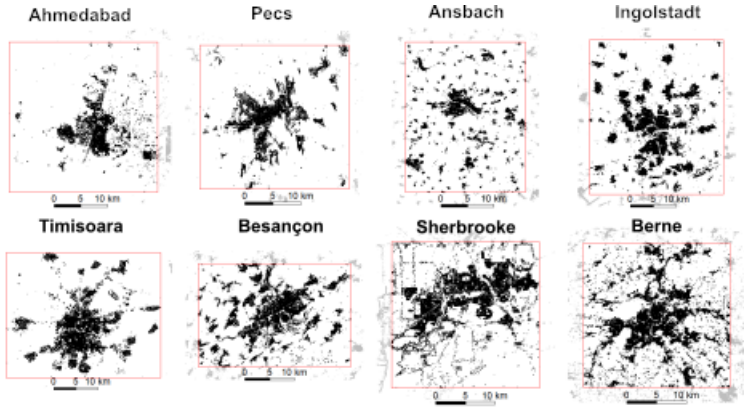


Figure 4.21. Les huit tissus bâtis pour lesquels la dimension de corrélation est supérieure à la dimension de boîte (différence supérieure ou égale à 0,1)

4.5 Conclusion

Les dimensions fractales calculées par les méthodes de boîte et de corrélation considérant l'emprise au sol des bâtiments sont fréquemment utilisées pour la caractérisation de tissus bâtis. La diversité des méthodes et de leurs paramètres de calcul rendent l'interprétation de leurs résultats et les comparaisons entre études hasardeuses. Dans ce chapitre, nous avons établi un protocole de mesure qui vise à minimiser les biais dans le calcul de dimension de boite et de corrélation. Sur base de se protocole, nous avons ensuite chercher à identifier ce que mesuraient chacune des deux dimensions (D_B et D_C). Nous avons pour cela utilisé vingt-cinq tissus bâtis théoriques et autant des tissus bâtis réels.

L'estimation des dimensions fractales de boîte et de corrélation est dépendante d'une série de paramètres. Les analyses effectuées dans la deuxième section de ce chapitre montrent que pour des tissus bâtis (à la fois théoriquement construits et réels) ayant une étendue spatiale du même ordre de grandeur, il est possible de fixer bon nombre de ces paramètres afin de permettre les comparaisons entre tissus bâtis, considérant l'une ou l'autre dimension, sans biaiser les estimations des dimensions fractales. Il est notamment intéressant d'effectuer les analyses sur base de fichiers raster à résolution fine (par exemple 4 m.). Une taille minimale de fenêtre de comptage de 30 m. permet l'obtention de résultats robustes. Partant de cette taille minimale, à condition que tous les tissus aient une étendue géographique du même ordre de grandeur, le nombre et la taille des fenêtres de comptage intermédiaires peuvent être fixés en adoptant un coefficient multiplicateur identique pour tous les tissus et appliqué selon une séquence géométrique jusqu'à une même taille de fenêtre maximale. Enfin, un ajustement logarithmique de la courbe de comptage empirique à une courbe théorique doit être effectué afin d'obtenir des résultats fiables et proches de la dimension de similarité.

Les dimensions de boite et de corrélation sont toutes deux le reflet de la densité bâtie et d'une organisation spatiale plus au moins complexe des bâtiments. Cependant, selon les cas, ces deux dimensions peuvent prendre ou non des valeurs différentes pour un même objet. Elles ne sont donc pas sensibles aux mêmes types d'irrégularités et ne sont, par conséquent, pas directement comparables. Nous émettons comme hypothèse que la présence d'espaces non bâtis de tailles et de distribution spatiale irrégulière provoque des différences entre la dimension de boîte et de corrélation

5

Delineation of cities based on scaling properties of urban patterns: a comparison of three methods

In this chapter, we compare three methods for the morphological delineation of cities based on the scaling properties of urban patterns. The main question is: do the results depend on the extent of the study area or on other biases such as the MAUP? The methods are critically compared and illustrated by applications to theoretical urban patterns and to the urban agglomeration of Brussels. We use different data and consider two study areas (the whole of Belgium and the former province of Brabant). The results show that it is impossible to draw an unambiguous morphological boundary for an urban agglomeration.

5.1 Introduction

As we saw in Section 1.2, there are four categories of methods for delineating urban morphological spaces. In this chapter, we focus on three methods from the fourth category (analysis of statistical distribution across scales). These methods have the advantage of being applied to disaggregated data and do not require a priori thresholds. It is well-known that case studies in geography are highly dependent on the size, shape and position of the spatial units used for the analysis (Openshaw, 1983; Briant et al., 2010), as well as the spatial extent of the study area (Goodchild, 2016). Therefore, we seek to determine the impact of these effects on the delineation of morphological agglomerations for three different methods: *Natural Cities*, *MorphoLim* and *Hierarchical Percolation*. Data used are of two types: the footprint of buildings (2D) or their centroid (1D), and the road intersections (1D). Such data are easy to collect, freely available in many countries and often easy to handle.

The first method derives *Natural Cities* by clustering points (street nodes or building centroids). It combines several methodological processes proposed in a series of papers headed by Prof. Bin Jiang (Jiang and Jia, 2011; Jiang, 2013, 2015, 2019; Jiang and Miao, 2015). Two thresholds are successively identified: first, the mean distance between the points of interest, which enables us to create clusters of points; second, the mean area of the clusters of points calculated via the recursive *head/tail breaks* method applied to the rank-size distribution of the clusters. The *Natural Cities* identified on this basis are the larger clusters (area larger than the nested means). The other two methods under consideration are the *Hierarchical Percolation* (Arcaute et al., 2016; Masucci et al., 2015) and fractal-based *MorphoLim* methods (Tannier et al., 2011; Tannier and Thomas, 2013). Both involve transforming the data (buildings mapped in 2D or road intersections) using a step-by-step dilation. The resulting dilation curves represent the size of the dilation buffer on the x-axis and the change in the number of clusters (in the case of *MorphoLim*) or the number of road intersections in the biggest cluster (in the case of *Hierarchical Percolation*) on the y-axis. A crucial threshold is then identified on the dilation curve. This threshold corresponds to a break point in the fractal organisation of the built pattern (*MorphoLim*) or to the condensation threshold of a logistic function (*Hierarchical Percolation*).

This chapter is structured as follows: Section 5.2 critically analyse each method with example of theoretical urban patterns, Section 5.3 is the application of the three methods to delineate the morphological agglomeration of Brussels (Belgium), and Section 5.4 concludes and discusses the impossibility of defining a single unambiguous morphological limit of a

city. The theoretical and empirical urban patterns analysed in Section 5.2 and Section 5.3 are freely available in the form of shapefiles (see section entitled 'Data and codes availability statement' at the end of this chapter).

5.2 Description and comparison of the three methods with theoretical urban patterns

5.2.1 *Natural cities*

The data used to identify *Natural Cities* are street nodes including intersections and end points (Jiang and Jia, 2011) (Fig.5.1a), although other types of points could be used, e.g. social media users' locations (Jiang and Miao, 2015). The methodological process adopted to delineate *Natural Cities* begins with the creation of a *Triangular Irregular Network* (TIN) as proposed by (Jiang and Miao, 2015) (Fig.5.1b). On this basis, we plot the rank-size distribution of the edges of the TIN according to their length and we verify that this distribution is heavily right-skewed, with a minority of large values in the head and a majority of small values in the tail. If the condition is confirmed, we apply the head-tail division rule set out by Jiang and Jia (2011): *"given a variable X, if its values x follow a heavy tailed distribution, then the mean (m) of the values divides all the values into two parts: a high percentage in the tail, and a low percentage in the head"* (Fig.5.1c). Thus, edges shorter than the mean length are selected (Fig.5.1d). We then transform the sets of contiguous edges into polygons, i.e. morphological clusters. We plot the rank-size distribution of the clusters according to their area. On the condition that this distribution is heavily right-skewed, we apply the *head/tail breaks* method proposed by Jiang (2013, 2015, 2019). The values of the distribution are separated into two groups: those above the mean and those below the mean; the partitioning for the values above the mean then continues iteratively until the head part of the distribution is no longer heavy-tailed, i.e. more than 40% of the clusters have an area higher than the mean (Jiang, 2019) . By the end of this process, the largest clusters across scales (i.e. the *Natural Cities*) are identified.

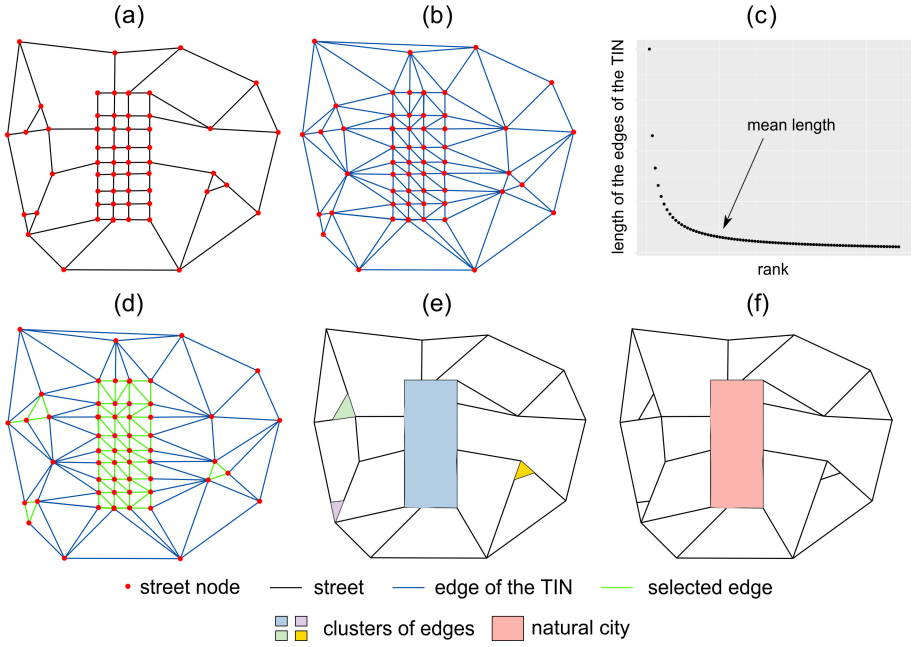


Figure 5.1. Methodological process adopted to delineate *Natural Cities*: (a) identification of the street nodes; (b) creation of a *Triangular Irregular Network* (TIN); (c) rank-size distribution of the edges of the TIN according to their length; (d) selection of the edges shorter than the mean; (e) transformation of the selected edges into polygons; (f) selection of the largest clusters (i.e. the *Natural Cities*).

In order to select the shortest edges of the TIN, we apply the non-recursive head-tail division rule and not the recursive *head/tail breaks* method because the rank-size distribution of the edges of the TIN is short-tailed and not long-tailed for the two types of data used and the two study areas considered in Section 5.3. This distinction between short-tailed and long-tailed distributions is pointed out in Jiang (2019): *"if the notion of far more smalls than larges recurs on at least twice, the data are long-tailed (...); otherwise, they are short-tailed."* In other words, if a distribution is no longer heavy-tailed after the first *head/tail break*, this means that it is short-tailed. Thus, we use the mean to select the shortest edges of the TIN. For right-skewed data distributions, the median is substantially smaller than the mean. Incidentally, using the median instead of the mean would lead to the creation of smaller *Natural Cities*. In Fig.5.2, it can be seen that the delineation of the *Natural Cities* is more relevant with the median than with the mean. Nevertheless, this observation should be confirmed in the case of real-world urban patterns (see Section 5.3.5).

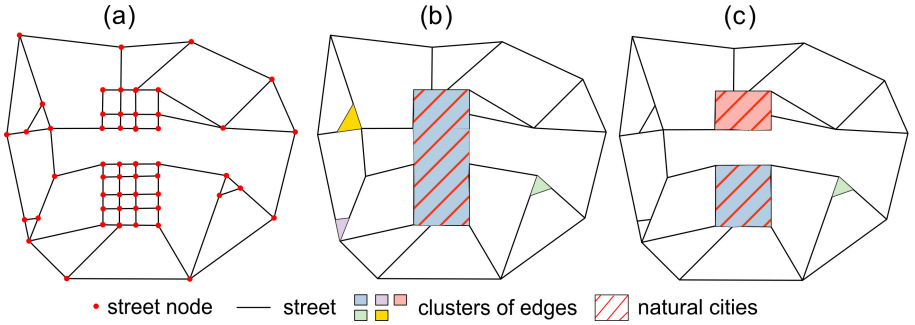


Figure 5.2. A road network (a) and two variants of the *Natural Cities* method: delineation based on the mean length of the edges of the TIN (b); delineation based on the median length of the edges of the TIN (c).

The delineation of *Natural Cities* also crucially depends on the geographical extent of the study area. Expanding the study area may lead to the introduction of supplementary "rural" street nodes being more distant from each other than "urban" street nodes: this mechanically increases the mean length of the edges of the TIN (Fig.5.3b). Conversely, the introduction of new "urban" street nodes shortens the mean length of the edges of the TIN (Fig.5.3c).

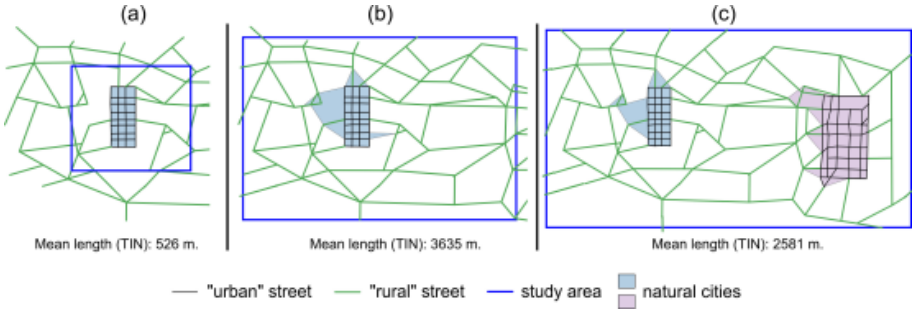


Figure 5.3. Effect of the geographical extent of the study area on the delineation of *Natural Cities*.

Finally, concerning the selection of the largest clusters across scales (i.e. the *Natural Cities*), it is worth noticing that Wang et al. (2015) and Liu et al. (2019) propose to use the maximal entropy instead of the *head/tail breaks* to identify the optimal size of clusters. This alternative is not considered in this thesis but could be worth testing further.

5.2.2 *MorphoLim*

Data used to delineate morphological agglomerations with *MorphoLim* (https://sourcesup.renater.fr/www/morpholim/) are the building footprints (2D). Considering that buildings are spatially distributed according to a fractal pattern, *MorphoLim* seeks to identify clusters of buildings whose spatial distribution logic differs the most from that of the surrounding buildings. For this purpose, a step-by-step dilation is applied to each building on the map.

To understand this principle, let us take the example of a highly schematic building fabric of 144 buildings. All buildings are identical in shape, area and orientation, and all buildings have their nearest neighbour at a distance d equal to their side length (Fig.5.4a). The first coverage step is to cover each building with a square of length $l_1 = 2 * d$. The neighbouring buildings then touch each other without overlapping. These buildings merge into aggregates squares with a side length of $L_1 = 2 * l_1$ (Fig.5.4b). At this stage, the number of aggregates (L_1) is 36 or $\frac{1}{4}$ of the initial number of buildings.

In the second covering step (Fig.5.4c), the L_1 aggregates are covered by squares with length $l_2 = 2 * l_1 + 2 * d$. In the same way as in step 1, as the L_1 aggregates have their nearest neighbour at distance l_1, the nearest l_2 squares touch each other and merge to give new aggregates L_2 with size $L_2 = 2 * l_2$. At this stage, the number of aggregates is equal to 9, i.e. $\frac{1}{4}$ of the number of aggregates in the previous step (L_1).

The same principle is applied to the third step of coverage. The aggregates L_2 are covered by squares of size $l_3 = 2 * l_2 + 2 * d$. The closest cover squares (distance equal to l_2) merge. However, it can be seen in Fig.5.4d that not all l_3 clusters are located at a distance l_2 from their nearest neighbour. Therefore, the number of L_3 aggregates (7) is not equal to $\frac{1}{4}$ of the number of L_2 aggregates (9). The scale's relation observed in step 1 and 2 of coverage stops in step 3. Step 3 marks a break in the self-similarity of the built fabric. It is at this step that the morphological limit of the aggregates is drawn. The most important aggregate can be assimilated to a city while the other aggregates can be assimilated, according to their importance, to boroughs or villages. Note that if the built fabric in Fig.5.4 was a perfectly self-similar fractal figure, no change would have been detected in the coverage process and the ratio of the number of clusters in one step to the number of clusters in the previous step would have been the same in all steps.

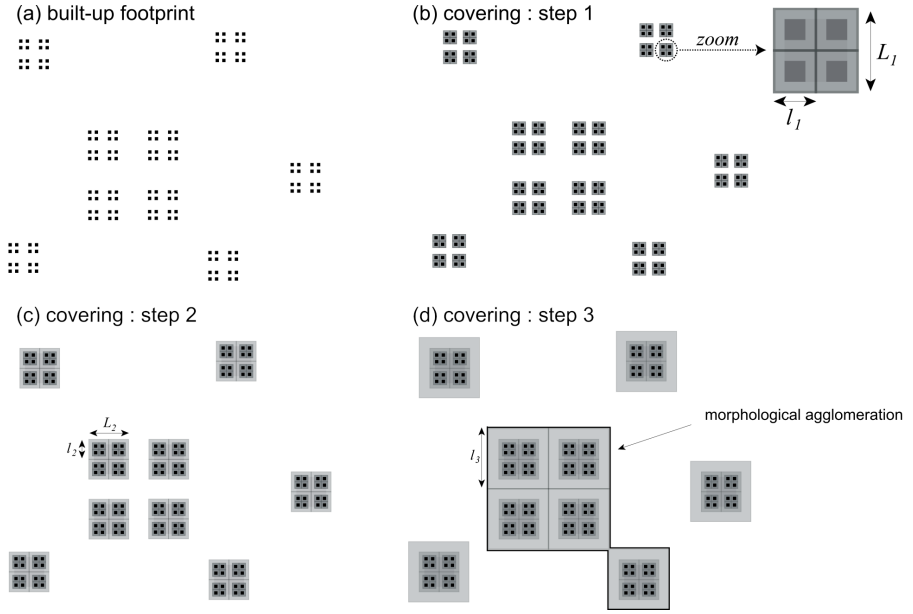


Figure 5.4. Covering of a theoretical urban pattern. Source:Tannier (2017)

For real urban built-up fabrics, it is clearly impossible to use such a covering process since the spatial distribution of buildings is not strictly self-similar. It is possible, however, to use the dilation procedure proposed by Minkowski (1903) and developed by Bouligand (1929). The buildings (or the pixels in the case of non-vector images) are surrounded by a buffer zone that becomes larger and larger (dilation) in an iterative process. At each stage of the dilation, buildings (or aggregates of buildings) merge if they touch each other and form a new (larger) aggregate. The number of built clusters is counted after each dilation step and the results are presented on a log–log plot, where the $x-axis$ represents the width of the dilation buffer and the $y-axis$ the corresponding number of built clusters (Fig.5.5a). If a pattern is fractal, then the number of clusters is related to their size by a power-law function corresponding to a linear relationship on a log–log plot and the slope of the curve is the fractal dimension (Tannier et al., 2011). If the pattern is not purely fractal or even multifractal, the dilation curve is not linear and the point at which the dilation curve deviates most from a straight line is a crucial distance threshold. On the curvature function of the dilation curve, which measures how far the curve deviates from a straight line at each point (Lowe, 1989), it is possible to identify the point characterised by the maximum curvature (Fig.5.5b). The maximum curvature of the dilation curve reveals a major spatial discontinuity across scales. The corresponding distance threshold separates two morphological

spatial subsets that are distinct in fractal terms: below that threshold, built elements are organised according to the same spatial logic and belong to the same morphological subset. Mapping the urban boundaries then consists in applying to the building map a buffer with a diameter equal to the distance threshold. On this map, the largest built cluster(s) is (are) identified by visual analysis of the rank-size distribution of all built clusters. The largest built cluster(s) is (are) considered forming the urban morphological agglomeration (Tannier and Thomas, 2013).

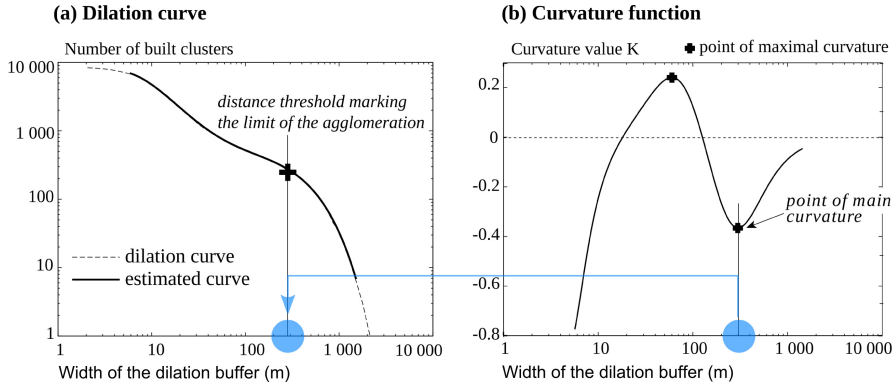


Figure 5.5. *MorphoLim*: identification of the dilation threshold marking the limit of the morphological agglomeration (blue points).

Yet the separation of two morphological spatial subsets that are distinct in fractal terms, indicated by the maximum curvature of the dilation curve, does not always correspond to the separation between the city and its periphery, as Fig.5.6 shows. Figures 5.6a and 5.6b represent two theoretical built patterns with a densely built-up urban core surrounded by a less dense suburban area that is itself surrounded by an even less dense rural area. The two figures are really similar except that the urban core in Fig.5.6a is denser than that in Fig.5.6b. As a consequence, the maximum curvature in Fig.5.6a indicates the boundary of the urban core, while in Fig.5.6b it indicates the boundary of the suburban area. In Fig.5.6a, the built shape of the suburban area is actually more similar to the built shape of the rural area than to the shape of the urban core. In other words, the strongest morphological difference concerns the urban core with respect to its suburban and rural periphery. Thus, in this case, the delineation of the city excludes the suburban area. In contrast, in Fig.5.6b, the built shape of the suburban area is closer to that of the urban core. Consequently, the delineation of the city includes the suburban area.

The delineation of a morphological agglomeration using *MorphoLim* can

depend on the geographical extent of the study area. Fig.5.7 illustrates this phenomenon. If we look at Fig.5.7a and Fig.5.7b, the delineation of the morphological agglomeration does not change with the expansion of the study area: the morphological agglomeration comprises the dense urban core and the suburban area surrounding it. This is because the expansion of the study area in Fig.5.7b with respect to Fig.5.7a consists in the addition of supplementary buildings in the rural area. Since the spatial distribution of those supplementary buildings conforms with the spatial distribution of buildings located in the rural area of Fig.5.7a, the morphological difference between the morphological agglomeration and its "rural" periphery is the same in both cases and their delineation is identical. In contrast, the expanded study area in Fig.5.7c, compared with Fig.5.7b, contains a second dense urban core. Thus, the strongest morphological difference is now between, on the one hand, the dense urban cores and, on the other hand, their suburban and rural peripheries. As a result, the suburban area, which surrounds the urban core located on the left of Fig.5.7c, is no longer included in the morphological agglomeration.

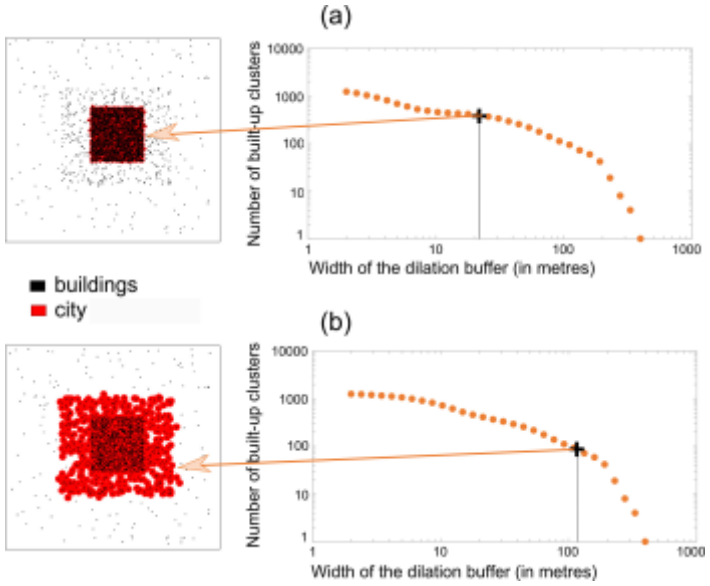


Figure 5.6. Analysis of two theoretical built patterns using *MorphoLim*. The urban core of pattern (a) is denser than the core of pattern (b).

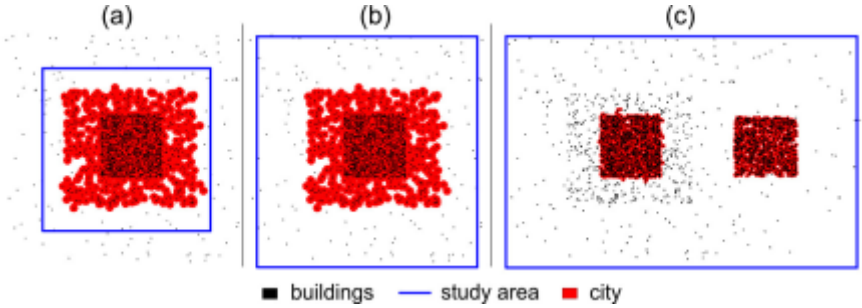


Figure 5.7. Effect of the geographical extent of the study area on the identification of city boundaries using *MorphoLim*.

5.2.3 *Hierarchical Percolation*

Hierarchical Percolation applies to street intersections (Arcaute et al., 2016; Masucci et al., 2015) (Fig.5.8). A street network is defined as a planar graph where the street intersections are the vertices and the street segments are the edges. Thus, intersection points are connected only if there is a street connecting them. Two street intersections belong to the same morphological cluster if the distance between them is below a given distance threshold. Progressively increasing the distance threshold via a dilation process enlarges the size of the clusters, until eventually a giant component appears, spanning the entire street network. The size of the largest cluster $N_{max}(\tau)$, measured in terms of the number of street intersections, increases with respect to the dilation distance (τ) according to a logistic growth function, where C is the carrying capacity, r the growth rate and τ_0 the inflection point (equation 5.1). The R code written to calculate the carrying capacity is provided (see section entitled "Data and codes availability statement").

$$N_{max}(\tau) = \frac{C}{1 + e^{-r(\tau - \tau_0)}} \tag{5.1}$$

The city condensation threshold is defined as the threshold where the maximum cluster size $N_{max}(\tau)$ intersects the carrying capacity of the logistic function. The largest cluster identified at this condensation threshold forms the city.

The geographical extent of the study area is critical when delineating cities by *Hierarchical Percolation*, as the expansion of the study area may involve the introduction of supplementary cities. In the example shown in Fig.5.9, we see that a first condensation threshold is reached at a $500m$ dilation distance. A second condensation threshold then appears at about $1000m$. Hence, as mentioned by Masucci et al. (2015), identification of the largest clusters (i.e.

the cities) within a pattern containing several cities cannot be completely automated and requires visual inspection.

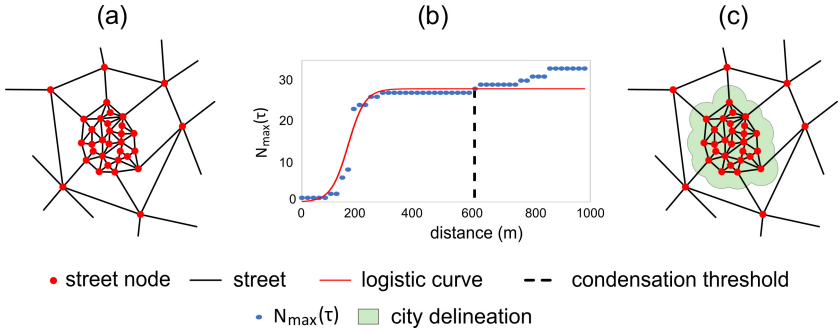


Figure 5.8. Hierarchical Percolation method applied to a theoretical street network. (a) identification of street intersections; (b) fitted logistic function ($x - axis$: size of the dilation buffer; $y - axis$: number of street intersections in the biggest cluster); (c) city delineation.

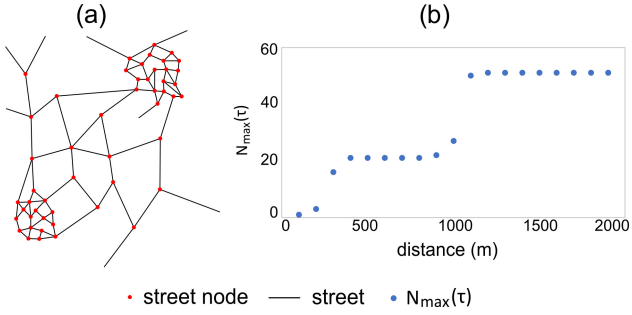


Figure 5.9. A polycentric street pattern (two centres) (a), and a double condensation threshold in the logistic growth function of the size of the largest cluster (b).

5.2.4 *Comparing the methods*

All three methods involve the use of fully disaggregated data and follow a bottom-up approach. Thus, they avoid biases associated with the MAUP. Nevertheless, their results are influenced by the existence of natural barriers, especially wide rivers or forests, or major transport routes. In the case of *Natural Cities*, the distances between street nodes separated by any such barrier can be above the mean (Fig.5.10a). In the case of *MorphoLim*, a barrier can separate a morphological agglomeration into two clusters. In Fig.5.10b, the two clusters have almost the same size. Yet if one of these clusters were smaller, it would not appear at the top of the rank-size

distribution of built clusters and so would not be identified as a part of the morphological agglomeration. In the case of *Hierarchical Percolation*, the barriers are reflected as jumps in the logistic function (Fig.5.10c). To overcome this limitation, Masucci et al. (2015) have artificially added street intersections in the middle of the bridges on the River Thames in London (UK). Interestingly, city delineation obtained using the three methods can take into account ribbons of buildings or street nodes linking two cities (Fig.5.11).

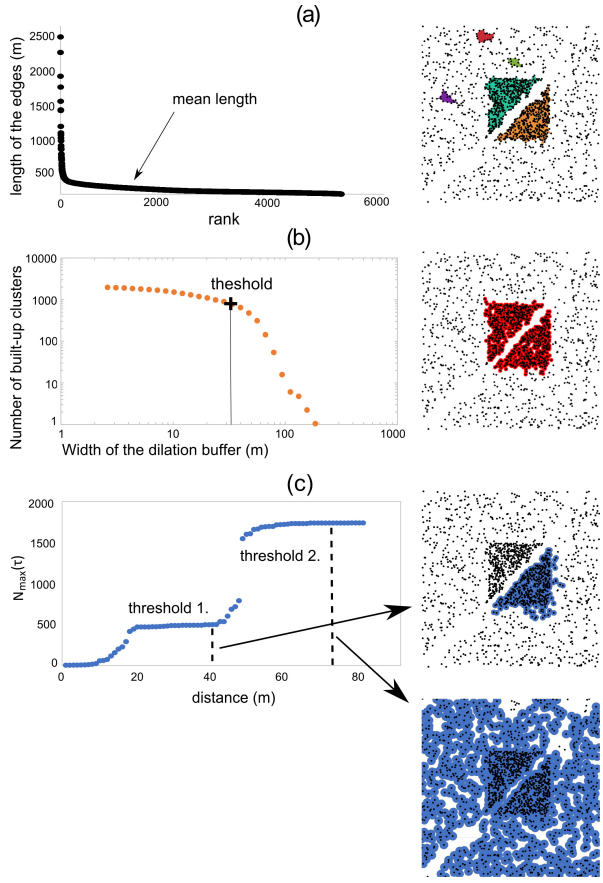


Figure 5.10. Morphological agglomeration within a theoretical urban pattern crossed by a large non built area: (a) *Natural Cities*; (b) *MorphoLim*; (c) *Hierarchical Percolation*.

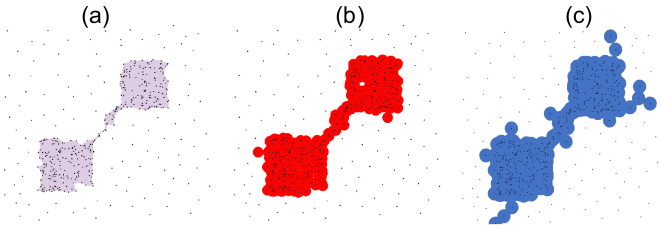


Figure 5.11. Morphological agglomeration of a theoretical urban pattern where a built ribbon links two urban entities by means of (a) *Natural Cities*, (b) *MorphoLim*, and (c) *Hierarchical Percolation*

All three methods are also sensitive to the geographical extent of the study area. Yet the underlying mechanism is different. For *MorphoLim* and *Hierarchical Percolation*, the extent of the study area does not change the dilation threshold or the percolation threshold unless a new built structure appears with the extension of the study area. In contrast, city delineations obtained using the *Natural Cities* method varies with respect to the extension of the study area even if no new built structure is included.

In the case of *Natural Cities*, distances separating street nodes are measured on a *Triangular Irregular Network* (TIN). In the case of *Hierarchical Percolation* and *MorphoLim*, distances separating buildings are Euclidean distances. City delineation with *Natural Cities* results from direct analysis of the distribution of distances that separate street nodes. With *MorphoLim* and *Hierarchical Percolation*, the distribution of distances is not directly analysed: city delineation appears with a jump in the size of the largest cluster of street intersections in the course of a dilation process (*Hierarchical Percolation*) and a maximum deviation from scale invariance of the number of built clusters in the course of dilation (*MorphoLim*). Although the criteria (measurement and analysis of distances) are different between the three methods, they all look for heterogeneity, i.e. spatial differentiation. Consequently, if buildings or street intersections are uniformly distributed in space, i.e. if distances separating buildings do not vary, no morphological agglomeration can be identified, whatever the method chosen. Besides, with *Hierarchical Percolation* and *MorphoLim*, no morphological agglomeration can be detected if local variations in distances separating the elements (buildings or street intersections) are always the same across scales.

Finally, each method involves some arbitrariness. As for *Natural Cities*, the *head/tail breaks* are defined here as more than 40% of the clusters with an area greater than the mean. Note that another percentage (20% or 30%) could also be applied. As for *Hierarchical Percolation*, the arbitrariness refers to the

visual inspection of the curve representing the growth of the size of the largest cluster in the course of dilation to identify the condensation thresholds. As for *MorphoLim*, the arbitrariness refers to the choice of the appropriate polynomial to fit the empirical dilation curve before calculating the curvature function.

5.3 Delineation of the morphological agglomeration of Brussels

The three methods under consideration have been applied to delineate the morphological agglomeration of the Brussels metropolitan area. While Brussels is undoubtedly a city, its spatial delineation is not clear. From an administrative and political standpoint, Brussels is referred to as Brussels-Capital Region (Fig.5.12a). Yet, the city extends over a much larger area across the borders of BCR into the Flemish and Walloon regions. The boundaries of this large metropolitan area are not officially or scientifically fixed: indeed, there is no governance at the scale of the metropolitan area and scientists do not concur about its limits (see, for example Dujardin et al., 2007; Boussauw et al., 2012; Thomas et al., 2012; Van Hecke et al., 2009). Hence, the delineation of the morphological agglomeration of Brussels raises not only interesting political issues but also the question of statistical homogeneity throughout the three administrative regions concerned.

5.3.1 Data and study area

The *Hierarchical Percolation* and *MorphoLim* methods necessitate adopting a large study area comprising at least one urban agglomeration and its hinterland (i.e. suburban or rural areas under the influence of the urban core). The *Natural Cities* method works better with big data, hence a large study area, as the statistical distribution of the data is more likely to be heavy-tailed (Jiang, 2015). Additionally, we have shown that all three methods are sensitive to the geographical extent of the study area. For those reasons, we delineate the morphological agglomeration of Brussels within two different study areas: (1) the former province of Brabant (Adam et al., 2017) and (2) the entire country (whole of Belgium). The former province of Brabant (3357 km^2) was created in 1815 and is today divided into three entities: BCR (1.2 million inhabitants), the province of Flemish Brabant in the north (1.1 million inhabitants) and the province of Walloon Brabant in the south (0.4 million inhabitants). It includes about ten regional cities (Fig.5.12b). The entire country of Belgium (30689 km^2) has 11.46 million inhabitants (2019).

Data about buildings come from the Land Registry Administration of Belgium (©2009 Administration Générale de la Documentation Patrimoniale). In this registry, all buildings with a footprint of more than $4m^2$ are mapped in 2D (polygons), regardless of their function (e.g. residential, commercial or industrial). For our study, to avoid noise due to small buildings such as garages or garden sheds, all entities under 12 m^2, which represent less than 5% of all buildings in the whole of Belgium, were removed from the database. Road network data come from the *Geofabrik platform* of *OpenStreetMap* [1]. OSM data have been chosen because of their large availability across the world. All cycle paths and car parks were removed from the database and all streets with the same name were merged to avoid their over-representation. Within cities, it can happen that street segments are duplicated two or three times in OSM data. Yet, a visual inspection of the data has shown that this is rarely the case in Belgium. Street segments have been transformed into street nodes according to the tutorial given by Ren (2018). All nodes of degree 2 were also removed.

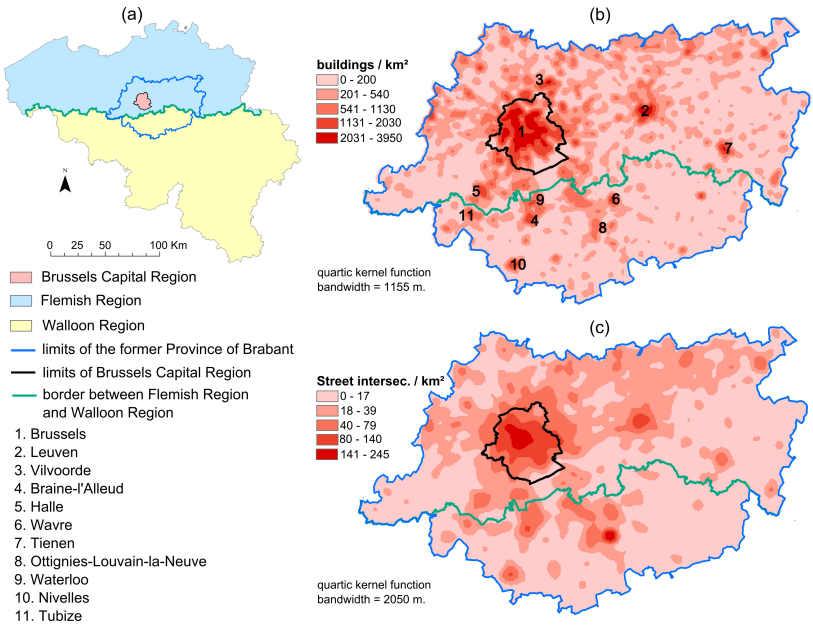


Figure 5.12. Location of the former province of Brabant around the Brussels-Capital Region in Belgium (a). Smoothed densities of building centroids (b) and road intersections (c) in the province of Brabant. Note: the bandwidths correspond to the optimal values proposed by default in the ArcGis software application.

[1] http://download.geofabrik.deaccessed08/21/2018

5.3.2 Natural Cities method applied to Brussels

The *Natural Cities* method is successively applied to street nodes, which include street junctions and end points (a), and to building centroids (b). The basic condition for applying the *Natural Cities* method is that the rank-size distribution of the edges of the TIN according to their length is heavily right skewed. This condition is fulfilled both for the province of Brabant and for Belgium.

a) Deriving *Natural Cities* from street nodes

When considering both the province of Brabant and the whole of Belgium, a single Natural City forms the morphological agglomeration of Brussels (Fig.5.13). This *Natural City* is centred on Brussels-Capital Region. Its size is close to that of BCR but its shape differs. The *Natural City* is slightly larger for the whole of Belgium than for the province of Brabant ($148km^2$ versus $111km^2$; see table 5.1).

b) Deriving *Natural Cities* from building centroids

When considering the province of Brabant, four small *Natural Cities* make up the morphological agglomeration of Brussels. Three are located within Brussels-Capital Region and the fourth covers the regional cities of Braine-l'Alleud and Waterloo (see Fig.5.12b and Fig.5.14a). When considering the whole of Belgium, only one *Natural City* located within Brussels-Capital Region makes up the morphological agglomeration of Brussels (Fig.5.14b).

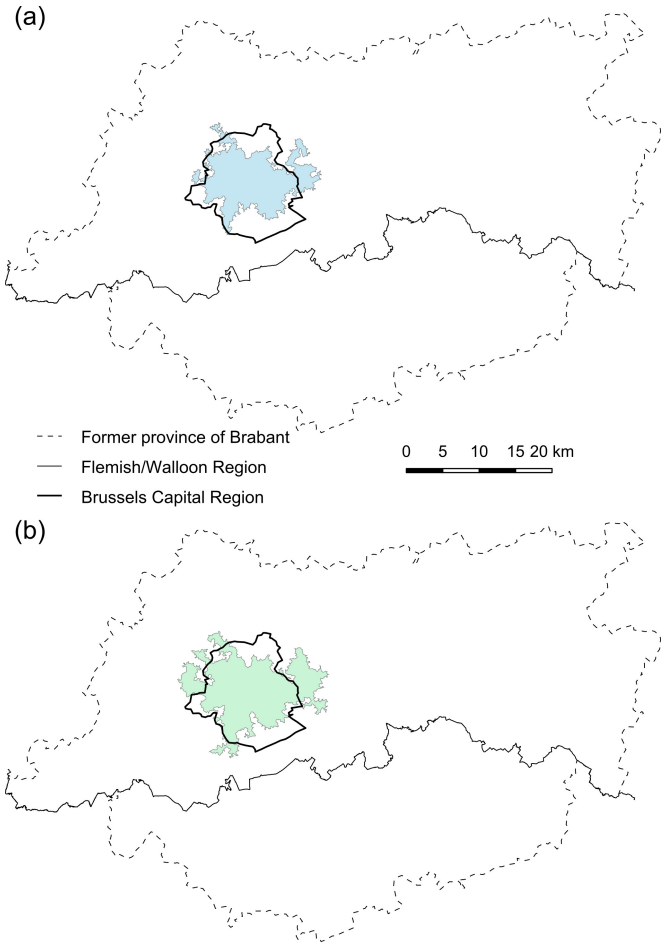


Figure 5.13. Morphological agglomeration of Brussels obtained using *Natural Cities* computed on street nodes (a) for the province of Brabant and (b) for the whole of Belgium.

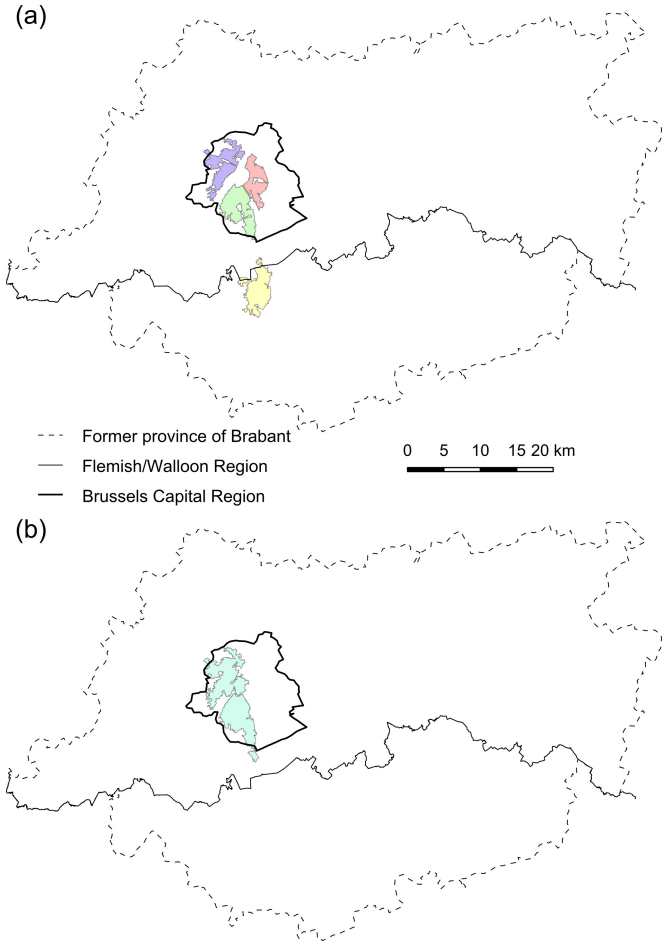


Figure 5.14. Morphological agglomeration of Brussels obtained using *Natural Cities* on the basis of building centroids: (a) for the province of Brabant and (b) for the whole of Belgium.

5.3.3 *MorphoLim* method applied to Brussels

The analysis was first performed on the province of Brabant. Fig.5.15a shows that the dilation curve is quite close to a straight line. The maximum curvature value is low (−0.087). The spatial distribution of buildings in the cluster at the top of the rank-size distribution (Fig.5.15b) therefore differs little from the spatial distribution of buildings in the rest of the province of Brabant. This largest cluster covers a major part of the study area (Fig.5.16a) and incorporates all regional cities in Fig.5.12b apart from Tienen and Nivelles. Indeed, Nivelles is quite confined within itself and Tienen is connected by a few built ribbons to some nearby villages but not to the other regional cities.

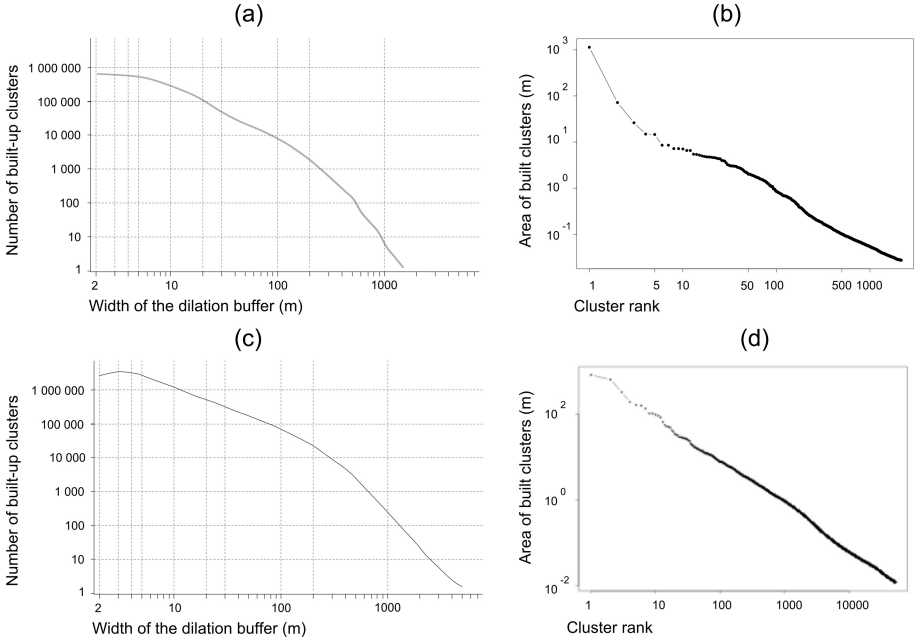


Figure 5.15. Dilation curve and rank-size distribution of built clusters obtained using *MorphoLim* for the province of Brabant (a & b) and for the whole of Belgium (c & d).

The results are markedly different when the study area is the whole of Belgium: the dilation curve exhibits a stronger main curvature (equal to −0.11) (Fig.5.15c); the cluster at the top of the rank-size distribution is not so large (Fig.5.15d); the morphological agglomeration spreads towards the south (Walloon Brabant) and towards the north-west (Flemish Brabant) beyond the border of the province where it stops at the edge of the city of Ghent (Fig.5.16b).

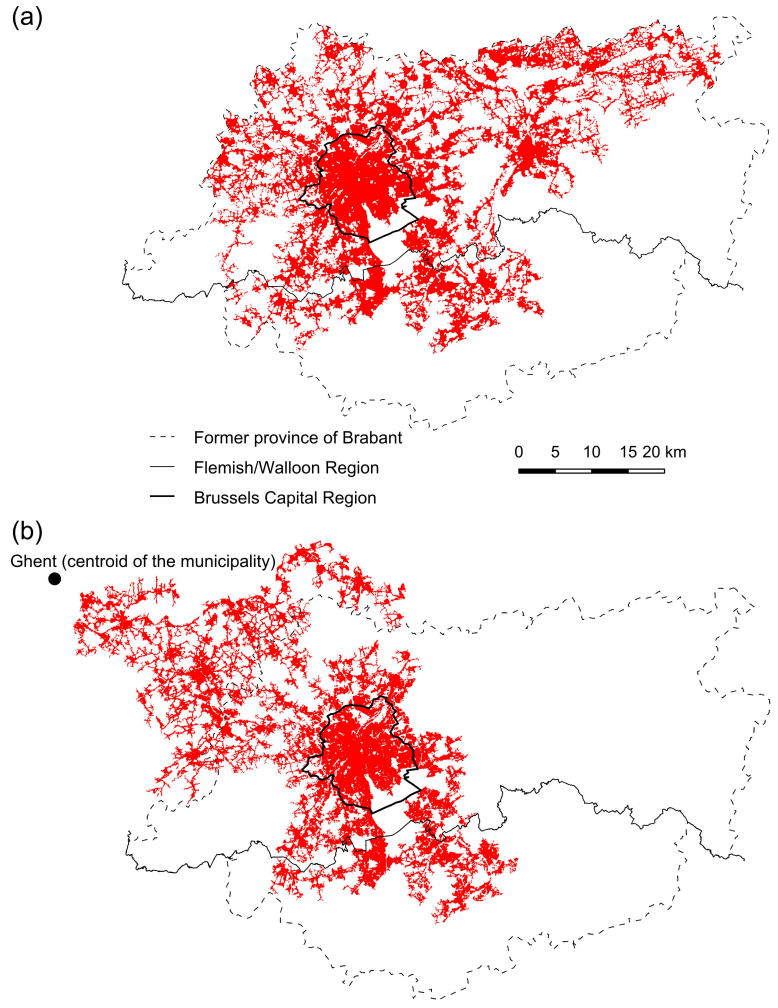


Figure 5.16. Morphological agglomerations of Brussels obtained using *MorphoLim*; (a) computed for the province of Brabant; (b) computed for the whole of Belgium.

5.3.4 *Hierarchical Percolation* method applied to Brussels

Hierarchical Percolation applies to road intersections (without the street end points unlike the *Natural Cities* method). Adaptation of the method enables us to also apply it to building centroids (Behnisch et al., 2019).

a) *Hierarchical Percolation* applied to road intersections

In the case of the province of Brabant, the first condensation threshold appears at $140m$. The morphological agglomeration of Brussels extends to the close periphery of Brussels-Capital Region and a little to the south to Braine-l'Alleud (Fig.5.17a). With the second threshold ($170m$), the agglomeration extends beyond the administrative limits of BCR in all directions and encompasses Braine-l'Alleud in the south. The results obtained for Belgium are very similar (Fig.5.17b): a first percolation threshold appears at 140 m and the morphological agglomeration of Brussels extends to the close periphery of BCR and to Braine-l'Alleud.

Turning now to the largest condensation thresholds, the morphological agglomeration of Brussels is no longer identifiable. In the case of the province of Brabant (threshold: $220m$), the resulting morphological agglomeration covers almost the entire study area. In the case of Belgium (threshold: $440m$), the resulting morphological agglomeration covers almost the entire country.

b) *Hierarchical Percolation* applied to building centroids

Similarly to the above-mentioned results, several percolation thresholds emerge at different dilation steps. For the province of Brabant, three thresholds are identified (Fig.5.18a). If we consider the first one ($30m$), the agglomeration of Brussels is almost entirely limited to BCR and only extends beyond it to the south where it encompasses Braine-l'Alleud. Considering the second threshold ($90m$), the agglomeration of Brussels is much larger and covers almost all of the western half of the study area. With the third threshold ($100m$), the two largest cities in the study area (Brussels and Leuven) merge and the morphological agglomeration covers all regional cities apart from Tienen and Nivelles.

With the extension of the study area to the whole of Belgium, the first percolation threshold appears at $60m$ (Fig.5.18b). At this stage, Brussels covers the western half of the study area and encompasses all regional cities apart from Leuven, Tienen and Nivelles. With the second threshold ($70m$), the agglomeration of Brussels encompasses the city of Ghent but still not Leuven.

Finally, with the next threshold ($160m$), all of Flanders is attached to the agglomeration of Brussels because of the existence of built ribbons that link the Flemish cities.

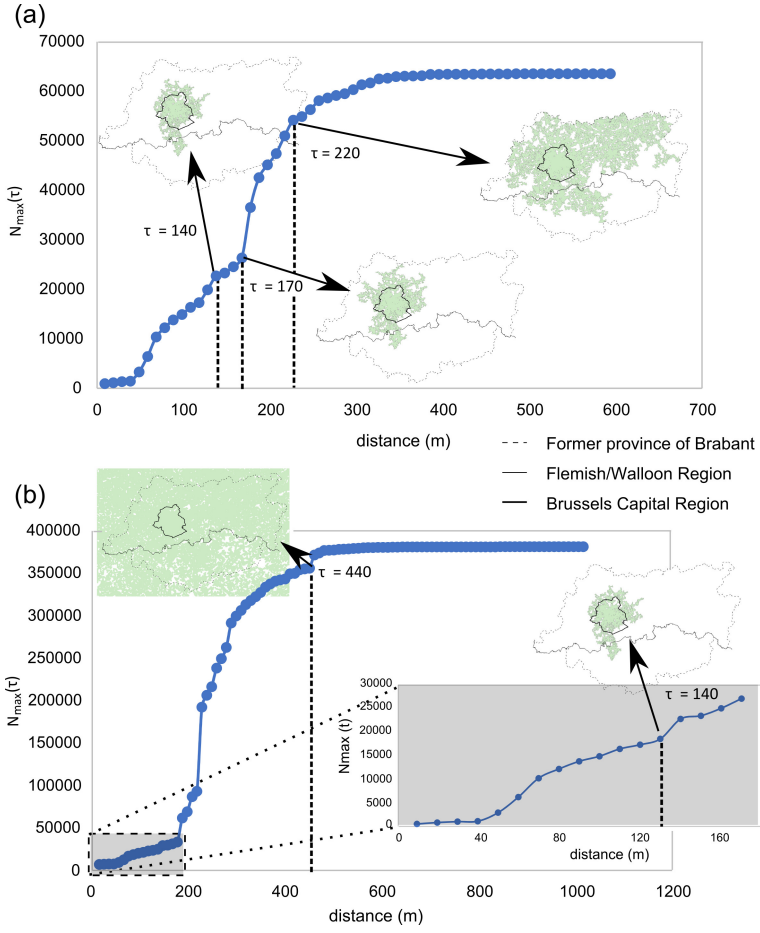


Figure 5.17. Logistic functions and morphological agglomerations of Brussels obtained using *Hierarchical Percolation* on the basis of road intersections; (a) for the province of Brabant; (b) for the whole of Belgium. $x - axis$: width of the dilation buffer; $y - axis$: number of road intersections in the largest cluster; τ: condensation threshold.

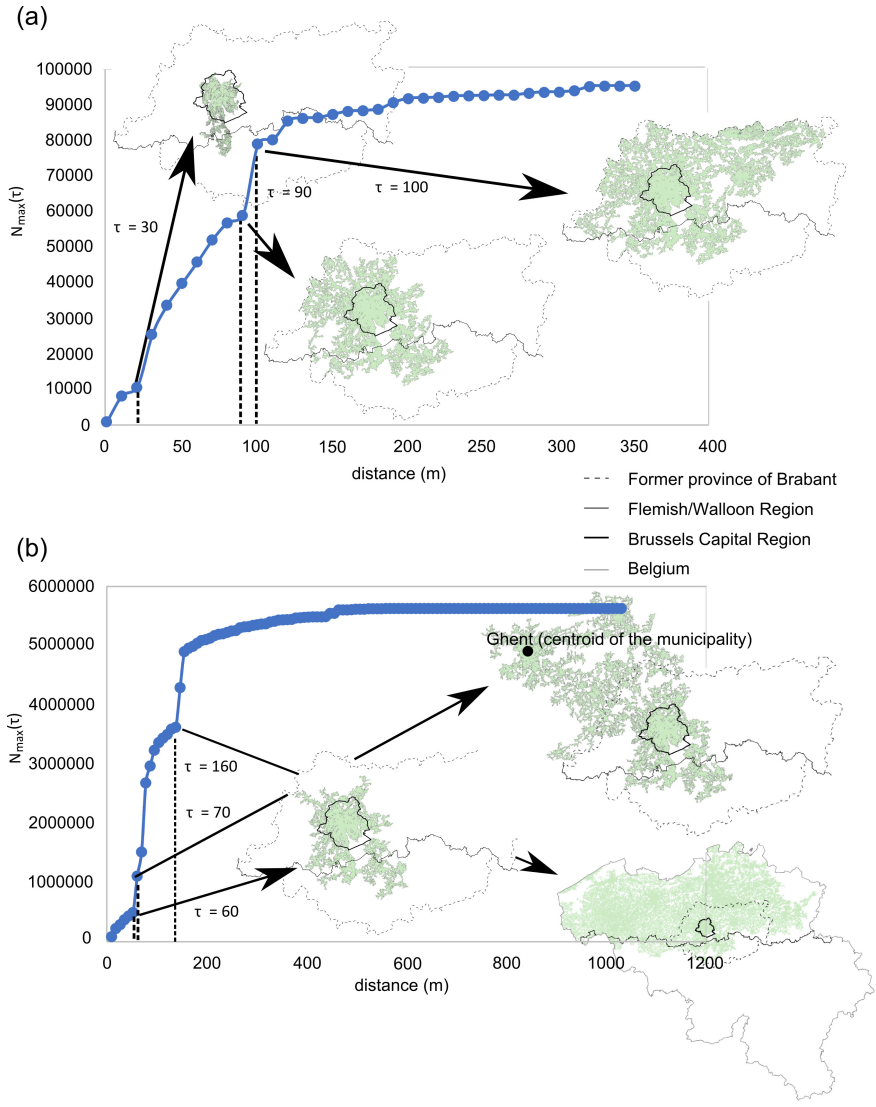


Figure 5.18. Logistic functions and morphological agglomerations of Brussels obtained using *Hierarchical Percolation* on the basis of building centroids; (a) for the province of Brabant; (b) for Belgium. $x - axis$: width of the dilation buffer; $y - axis$: number of building centroids in the largest cluster; τ: condensation threshold.

Method	Data	Study area	Threshold (m)	Area of Brussels' agglo. (km^2)	Perimeter of Brussels agglo. (km)
Natural Cities	Road inters.	Pr. of Br.	212(*)	110	169
		Belgium	250(*)	148	204
	Building centroids	Pr. of Br.	48(*)	67	329
		Belgium	54(*)	46	187
MorphoLim	Buildings footprints	Pr. of Br.	181(+)	653	5299
		Belgium	121(+)	502	5186
Hierarchical Percolation	Road inters.	Pr. of Br.	140(.)	271	1038
			170(.)	395	1262
			220(.)	1395	4165
		Belgium	140(.)	269	1030
			440(.)	21580	20876
	Building centroids	Pr. of Br.	30(.)	128	1261
			90(.)	710	3053
			100(.)	1176	5093
		Belgium	60(.)	395	2073
			70(.)	1158	6818
			160(.)	8779	25463

Table 5.1. Summary of results obtained using the three delineation methods. (*) Mean length of the TIN; (+) distance at which the dilation curve deviates most from a straight line; (.) percolation threshold

5.3.5 Discussion

By using three methods for delineating the morphological agglomeration of Brussels, we obtain a series of morphological limits that differ markedly. With *Hierarchical Percolation*, *MorphoLim* and *Natural Cities* applied to street nodes, the morphological agglomeration is made up of a single morphological cluster. In contrast, the application of the *Natural Cities* method to building centroids results in the identification of several small *Natural Cities*. In general, all *Natural Cities* are quite small. It would therefore be irrelevant to use the median instead of the mean, as considered in Section 5.2.1

With *MorphoLim*, the morphological agglomeration of Brussels is smaller if the study area is Belgium (Fig.5.16b) than if it is the province of Brabant (Fig.5.16a); see also Tab.5.1. This is because a crucial change in distances separating buildings occurs at a lower distance threshold when considering all of Belgium (we will see in the following paragraph that limiting the study area to Brabant for the use of *MorphoLim* is not really relevant). The opposite is found for the *Hierarchical Percolation* method applied to building centroids when the first percolation threshold is considered (see Fig.5.18a and Fig.5.18b, and Tab.5.1). Yet in the case of street intersections, the percolation thresholds and the area of the morphological agglomerations are the same whichever study area is under consideration (province of Brabant or whole of Belgium) when the first percolation threshold is considered (see Fig.5.17a and Fig.5.17b, and Tab.5.1). For *Natural Cities*, the effect of the expansion of the study area also depends on the data used. With the street intersections, we observe an increase in the mean length of the edges of the TIN and a decrease in the minimal area of the clusters. With the building centroids, we again observe an increase in the mean length of the edges of the TIN, but this time an increase in the minimal area of the clusters (see Fig.5.13 and Fig.5.14, and Tab.5.1).

Among all delineations, some are undoubtedly irrelevant if the objective is to define a large agglomeration for Brussels. First, *Natural Cities* derived from building centroids are much too small (Fig.5.14). They delineate large neighbourhoods within BCR as well as the morphological agglomeration of two regional cities (Braine-l'Alleud and Waterloo), but not one inclusive morphological agglomeration as expected (see for example Dujardin et al., 2007; Vandermotten, 2016). Second, the boundaries of the morphological agglomeration are sometimes artificially truncated by the border of the province of Brabant when this study area is chosen. This is the case with the delineation obtained using *MorphoLim* (Fig.5.16a) as well as the application of the largest percolation thresholds of *Hierarchical Percolation* (Fig.5.17a - threshold $220m$, Fig.5.18a - threshold $100m$). In those cases, the province of Brabant is too small with respect to the chosen method or parameters and a larger study area (e.g. the whole of Belgium) would be required. Third, when the study area is the entire country, the largest percolation thresholds do not enable us to delineate the morphological agglomeration of Brussels because the morphological cluster being identified covers either all of Flanders (for building centroids) or almost all of Belgium (for street intersections).

Finally, nine delineations of the morphological agglomeration of Brussels appear to be geographically relevant: the *Natural Cities* identified on the basis of street nodes for the two study areas (province of Brabant and

Belgium), the delineation obtained using *MorphoLim* applied to the whole of Belgium, the delineations obtained using the *Hierarchical Percolation* method applied to road intersections (province of Brabant - thresholds 140m and 170m, and Belgium - threshold 140m) and to the centroids of buildings (province of Brabant - threshold 30m, and Belgium - thresholds 60m and 70m). In six cases out of nine, the limit of the morphological agglomeration of Brussels remains in the close periphery of BCR (*Natural Cities*, Fig.5.13a and Fig.5.13b), or spreads a little into the Walloon region to Braine-l'Alleud (*Hierarchical Percolation*, Fig.5.17a - threshold 140m and 170m, Fig.5.17b - threshold 140m, Fig.5.18a - threshold 30m). In another case, the morphological agglomeration of Brussels is larger and extends clearly over the borders of BCR into the Flemish and Walloon regions (Fig.5.18b - threshold 60m). It encompasses seven regional centres: Vilvoorde, Halle, Tubize, Waterloo, Braine-l'Alleud, Wavre and Ottignies-Louvain-la-Neuve. In the last two cases, the agglomeration of Brussels is even much larger. It spills over the border of the province of Brabant and reaches the edge of the city of Ghent in the north-west (MorphoLim, Fig.5.16b) or even encompasses the city of Ghent (*Hierarchical Percolation*, Fig.5.18b - threshold 70m).

Finally, two opposing views of the settlement pattern of Belgium are supported by the results obtained. On the one hand, the limit of the morphological agglomeration of Brussels remains in the close periphery of BCR or spreads a little into the Walloon region to Braine-l'Alleud (six relevant delineations of Brussels out of nine). This confirms the idea that Belgian cities are still separate entities (Van Criekingen et al., 2007). In contrast, in three delineations out of nine, the morphological agglomeration of Brussels extends clearly over the borders of BCR into the Flemish and Walloon regions, or spills over the border of the province of Brabant and reaches the city of Ghent in the north-west. Moreover, the morphological agglomeration even covers either all of Flanders or almost all of Belgium, with the largest percolation thresholds. This supports the idea that Belgium is one large conurbation (Vandermotten et al., 2008) in opposition to Van Criekingen et al. (2007). In actual fact, Belgian cities are separate entities but they have strong morphological ties that take the form of built ribbons, especially in Flanders.

Each method provides a specific delineation of the morphological space that could be associated with particular functional uses. *Natural Cities* could be used when the users look for the location of morphological hot spots since this method highlights places that have a high concentration of road intersections. The method can then be used for example by a transport manager (e.g. bus or taxi) who wants to rethink the structure of his network

and stations. The use of *MorphoLim* would be more advisable when an urban planner wants to delimit the space into two distinct parts (since *MorphoLim* draws a border between two different organisations of built spaces). Let's imagine that the Belgian State allocates subsidies for the development of urban parks. A delimitation with *MorphoLim* would make it possible to identify urban areas (eligible for the subsidy) as opposed to rural areas (not eligible). *Hierarchical Percolation* has applications quite similar to those of *MorphoLim*. However, the hierarchical approach allows a less binary and more evolving vision of the space. Let us assume the creation of a traffic tax for cars circulating in Brussels. This tax could be higher when closer to the CBD. The *Hierarchical Percolation* method can then be used to draw the various boundaries that are synonymous with a shift to a higher tax scale.

5.4 Conclusion

Conventional methods that integrate a fixed threshold set ex ante have the advantage of being easy to apply and their results are easy to handle. Yet such methods overlook the specific features of the patterns under consideration that appear across scales and that are revealed by the three methods compared in this chapter. Each of these methods has a specific way of detecting the urban boundary— the nested mean value in a *head/tail breaks* distribution for *Natural Cities*, a maximum deviation from a hyperbolic distribution for *MorphoLim* and a condensation threshold in a logistic growth function for *Hierarchical Percolation*. Cities with the same built density or the same density of street nodes may display different dilation thresholds (*MorphoLim*) or percolation thresholds (*Hierarchical Percolation*), or even mean edge lengths (*Natural Cities*). All three methods involve the use of fully disaggregated data and follow a bottom-up approach. Consequently, they avoid biases of the MAUP. Nevertheless, they are all sensitive to the geographical extent of the study area.

The three methods were applied to two study areas (the province of Brabant and the whole of Belgium) with two types of data (buildings and street intersections). The resulting delineations of the morphological agglomeration of Brussels differ markedly. Accordingly, it is impossible to draw a single unambiguous morphological boundary for an urban agglomeration in general, and for Brussels in particular. This is consistent with the observations by Chen (2016): *"Urban form has no characteristic scale, and thus an urban boundary cannot be identified exactly"* and by Jiang (2018): *"(...) things are not measurable, or measurement depends on the measuring scales because of the fractal nature of geographic phenomena (e.g.*

Goodchild and Mark, 1987; Batty and Longley, 1994; Chen, 2011)". Indeed, if a city is a boundless fractal, none of the three methods can enable its boundary to be identified. As suggested by Chen (2016), *"(...) geographical phenomena fall into two groups: one is with characteristic scale (scaleful group), and the other, without characteristic scale (scale-free group). The former can be described with characteristic length such as average value, standard deviation, and eigenvalue, while the latter should be described with scaling exponent such as fractal dimension."* Nevertheless, the three methods make it possible to delineate a morphological agglomeration of Brussels because local deviations from scale invariance actually exist (Chen, 2016; Sémécurbe et al., 2019).

As stated by Jiang (2018), geography has three fundamental issues regarding geographical space: how it looks, how it works, and what it ought to be. In the present chapter, we compare three methods with a view to seeing how they delineate the built-up realities on the basis of the example of the morphological agglomeration of Brussels. However, we are not comparing these methods with a view to finding out how a morphological agglomeration should be, nor are we trying to determine whether or not results comply with norms or ideas of goodness or optimality, as done for instance in Chen (2016) and Jiang (2019). With the example of Belgium, we here compare methodologically and empirically three methods for delineating morphological agglomerations; Belgium is sufficiently diverse in terms of urban textures for trusting the present results. Comparing with case studies in other countries/continents opens a new avenue for further research in human and urban geography.

Street nodes and building centroids do not convey the same information about the spatial organization of built patterns. This is very clear in the case of the *Natural Cities* method, as the *Natural Cities* delineated on street nodes are much larger than the *Natural Cities* delineated on building centroids. Distances that separate street intersections are actually above the mean in between individual cities, whereas distances that separate building centroids are above the mean everywhere except in a few very dense intra-urban built clusters. Moreover, the nature of the data used exerts a joint effect with the size of the study area on city delineations. This joint effect varies with the method under consideration. With this in mind, the crucial interest of the *Natural Cities* method applied to street nodes is that it enables clear identification of the morphological limits of individual cities, in this case Brussels but it could be other cities as well. Considering the Belgian case, changing the study area has an impact on the resulting delineations. Indeed, expanding the study area expands Brussels' *Natural City* just a little bit. In

contrast, the *MorphoLim* and *Hierarchical Percolation* methods applied to the building data for the whole of Belgium enable the identification of an extensive morphological agglomeration of Brussels that includes many regional cities, among which Brussels, linked by built ribbons. However, the application of those two methods to building data considering only the province of Brabant is not relevant because the boundaries of the morphological agglomeration are truncated by the border of the study area. Finally, the *Hierarchical Percolation* method applied to street intersections is interesting, as it enables the identification of three percolation thresholds, and thus three nested morphological agglomerations of growing size that show how the peripheries of Brussels are more or less closely linked to the core agglomeration from a morphological point of view. Conversely, this method does not provide a single clear limit of the morphological agglomeration of Brussels.

To conclude, geographers and planners should be more than ever conscious of the importance of their choices about the data used, the extend of the study area and the delineation method. This chapter has indeed shown that the delineation of an urban space depends on these three aspects. They all have to be carefully selected with respect to both the objectives of each delineation and the specific features of the area under study. This is therefore also true for delineation methods that are based on statistical analysis of a pattern across scales.

In the following chapter, we focus in particular on *MorphoLim*. Knowing how this method works, how to characterise and understand its results will be our main objective.

Data and codes availability statement: data and codes that support the findings of this chapter are available on figshare.com: https://doi.org/10.6084/m9.figshare.11405598

6

Characterisation of urban agglomerations delineated by *MorphoLim*

In this chapter, we seek to understand the characteristics of agglomeration delimited by *MorphoLim*. To do so, we study the delineation of 82 French urban agglomerations. We use variables describing (1) the built-up areas of the agglomerations, (2) the size of the agglomerations and (3) the shape of their boundaries. We find that some variables co-vary but that, overall, the three types of variables each carry part of the information that enables us to describe the delimited morphological agglomerations. This chapter continues the work initiated by Tannier and Thomas (2013).

6.1 Introduction

Le chapitre 5 nous a montré que différentes méthodes de délimitation des espaces urbains, même lorsqu'elles sont basées sur l'analyse de distributions statistiques, mènent à des résultats différents. Il est alors important de mettre en perspective les résultats d'une méthode à la lumière de la méthode elle-même. Nous voulons donc dans ce chapitre mesurer les caractéristiques des espaces délimités par *MorphoLim*. Pour mémoire, *MorphoLim* délimite morphologiquement les villes en effectuant une dilatation pas à pas du tissu bâti et en repérant la distance de dilatation à laquelle une rupture majeure apparaît en ce qui concerne la répartition spatiale des bâtiments (grand changement des distances entre les bâtiments). La délimitation effectuée correspond à l'identification de la limite entre deux sous-ensembles morphologiques distincts d'un point de vue fractal, dont l'un est la ville. Nous devrions donc mesurer des disparités importantes dans l'organisation spatiale des bâtiments de part et d'autre des frontières tracées par *MorphoLim*.

Cette disparité dans l'organisation spatiale des bâtiments découle donc directement de la méthode mais n'est, à priori, pas la seule caractéristique des espaces délimités par *MorphoLim*. Comme nous l'avons vu dans l'introduction (section 1.2), la forme des frontières d'une ville peut être le reflet des forces internes à celle-ci (force économiques, sociales, politiques, etc.). En effet, ces forces interagissent ensemble dans un espace fini et donc impact leurs bordures (Medda et al., 1998; Batty and Longley, 1994). En plus de mesurer les caractéristiques des espaces délimité par *MorphoLim*, nous allons donc également mesurer, dans ce chapitre, les caractéristiques morphologiques des frontières. Est-ce que caractériser une frontière, revient à caractériser l'espace à l'intérieur de cette frontière ?

Afin d'analyser les caractéristiques morphologiques des agglomérations délimitées par *MorphoLim*, nous allons mesurer différents indices ayant trait à la répartition spatiale des bâtiments dans les sous-ensembles morphologiques, d'une part, et à la forme de la limite tracée entre ces sous-ensembles (i. e. l'enveloppe morphologique des villes), d'autre part. Nous allons appliquer la méthode *MorphoLim* à un large échantillon d'aires urbaines françaises afin de mettre en évidence des éventuelles covariations entres les différents indices calculés. Nous tenterons ensuite d'établir une typologie morphologique des villes étudiées. Tannier and Thomas (2013) ont déjà effectué ces analyses pour six villes théoriques et dix-huit villes belges. Elles montrent notamment que la relation entre la forme du périmètre d'une agglomération et la morphologie bâtie au sein de cette agglomération ne peut pas facilement et simplement être caractérisée. En effet, pour les dix-huit villes belges, aucune corrélation

n'apparaît par exemple entre la densité bâtie des villes et la complexité de leurs frontières mesurées à l'aide d'une dimension fractale de boîte. Tannier and Thomas (2013) observent cependant, parmi d'autres éléments sur lesquels nous reviendrons plus en détail dans la suite du chapitre, que plus la surface d'une agglomération est grande, plus son périmètre présente de nombreuses indentations. Par ailleurs, Sémécurbe et al. (2019) ont montré, pour 20 villes françaises, que le seuil de distance identifié au moyen de *MorphoLim* et caractérisant chaque ville n'est pas corrélé statistiquement à la surface de chaque aire urbaine. A partir d'un échantillon plus grand de 82 villes françaises, nous chercherons à confirmer ou infirmer ces différents résultats.

Dans ce chapitre, nous commençons par présenter les 82 aires urbaines retenues pour l'analyse (Section 6.2). Nous présentons à la section 6.3 la méthode que nous avons suivie pour obtenir les différents résultats (agglomérations) que nous présentons ensuite en à la section 6.4. La section 6.5 conclu.

6.2 Les 82 aires urbains étudiées

Nous allons délimiter des agglomérations morphologiques au départ d'aires urbaines françaises. Selon la définition de l'INSEE (*Institut National de la Statistique et des Études Économiques, France*), une aire urbaine est *« un ensemble de communes, d'un seul tenant et sans enclave, constitué par un pôle urbain de plus de 10 000 emplois, et par des communes rurales ou unités dont au moins 40% de la population résidente ayant un emploi travaille dans le pôle ou dans des communes attirées par celui-ci »* (INSEE, 2016). La France comptait, en 2020, 161 aires urbaines en métropole et en territoire d'outre-mer. Nous avons choisi d'étudier uniquement les aires urbaines de plus 100 000 habitants situées en France métropolitaine. Nous supprimons également de notre échantillon les aires urbaines qui s'étendent vers d'autres pays, par exemple celle de Lille qui s'étend en Belgique ou celle de Strasbourg qui s'étend en Allemagne. Nous conservons finalement un échantillon de 82 aires urbaines (Fig 6.1). Les données, c'est-à-dire tous les bâtiments dont l'emprise au sol est supérieure à 1 m^2, proviennent de la BD TOPO éditée en mars 2021 par l'*Institut National de l'Information Géographique et Forestière, France* (`geoservices.ign.fr`).

L'aire urbaine de Paris est de loin la plus grande en termes de superficie (17 278 km^2) et de nombre de bâtiments (3 654 901). Les aires urbaines de Lyon, Bordeaux et Toulouse ont une superficie comprise entre 5 500 km^2 et 6 000 km^2 pour un nombre de bâtiments compris entre 890 000 et 985 000. Mis

à part ces quatre aires urbaines particulièrement grandes, les autres ont des superficies comprises entre 157 km^2 et 3 755 km^2 et un nombre de bâtiments compris entre 30 976 et 838 781. Il y a donc une variabilité notable de superficie et de nombre d'entités bâties entre les différentes zones d'étude.

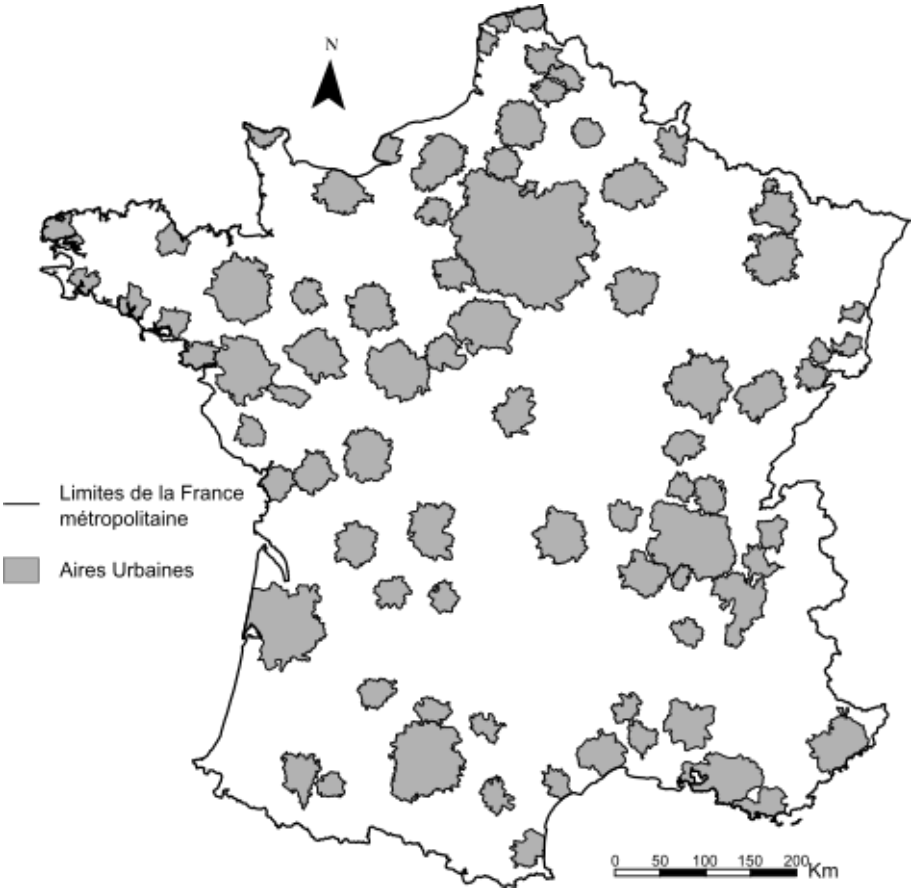


Figure 6.1. Les 82 aires urbaines de l'échantillon

6.3 Méthodologie

6.3.1 Délimitation des agglomérations morphologiques

Nous utilisons le programme *MorphoLim* conformément aux recommandations de Tannier et al. (2011) et Tannier and Thomas (2013). Une dilatation des bâtiments de chaque aire urbaine est effectuée suivant la même progression géométrique (coefficient de 1,2). Les courbes empiriques des nombres d'agrégats à chaque étape de dilatation sont ajustées à l'aide de fonctions polynomiales. Le degré du polynôme est propre à chaque aire urbaine et est déterminé en minimisant visuellement, à l'aide d'un graphique, un critère d'information bayésienne (BIC). Le degré du polynôme doit être suffisamment grand pour ajuster au mieux la courbe de dilatation mais ne pas être trop grand afin de limiter le biais de sur-ajustement. Le seuil de distance de dilatation de chaque tissu est obtenu grâce à la maximisation d'une fonction de courbure (point de courbure principale en valeur absolu) appliquée à la courbe de dilatation estimée. Si le point de courbure principale indique un seuil à une distance inférieure à 30 m. ou supérieur à 1000 m., le deuxième maximum (en valeur absolue) de la fonction de courbure est sélectionné. Un seuil de distance inférieur à 30 m. décrirait en effet une rupture dans les distances entre bâtiments due à la largeur des routes tandis qu'un seuil supérieur à 1000 m. est au-delà des ordres de grandeurs des distances inter-bâtiments intra-urbaines (Tannier, 2017).

Une fois le seuil de distance déterminé, *MorphoLim* produit deux shapefiles. Le premier est nommé « buffer » et correspond au tissu bâti dilaté selon une zone tampon d'épaisseur égale au seuil de distance divisé par deux. Le deuxième shapefile est nommé « enveloppe » et correspond au shapefile « buffer » érodé en appliquant une zone tampon d'épaisseur égale au seuil de distance divisé par deux. Les surfaces des agrégats du shapefile « buffer » sont calculées afin de produire les distributions rang-taille des agrégats de chaque aire urbaine. Une analyse visuelle de ces distributions permet de sélectionner le ou les agrégats principaux qui forment l'agglomération morphologique. Ensuite, pour l'analyse des caractéristiques de l'agglomération morphologique, on considère les agrégats de bâtiments érodés du shapefile « enveloppe » qui sont contenus dans les limites de l'agglomération morphologique identifiées à partir du shapefile « buffer ».

6.3.2 Caractéristiques des tissus bâtis des 82 villes étudiées

Étant donné que la limite morphologique des villes identifiée avec *MorphoLim* correspond à une rupture vis-à-vis d'une distribution spatiale fractale des bâtiments, il serait pertinent de mesurer la dimension fractale des tissus bâtis de part et d'autre de cette limite. Toutefois, s'il est possible de calculer la dimension fractale du tissu bâti de l'agglomération morphologique, il n'est pas possible de mesurer la dimension fractale du bâti situé hors de l'agglomération morphologique car le vide laissé par cette dernière au milieu du tissu étudié fausserait les mesures. La différence de fractalité du tissu bâti de part et d'autre de la limite ne peut donc être évaluée qu'en comparant la dimension fractale de l'ensemble de l'aire urbaine à celle de son agglomération morphologique.

Comme Tannier and Thomas (2013), nous avons choisi de calculer la dimension fractale de boîte (box-counting dimension). Dans nos analyses, nous adoptons les mêmes paramètres de mesure pour l'estimation de la dimension de boîte des aires urbaines ($D_s(AU)$) et des agglomérations morphologiques ($D_s(AM)$) afin de permettre leur comparaison (voir Chapitre 4). Les tissus bâtis sont rastérisés à une résolution de 4 m. Les fenêtres de comptage s'étendent de 30 m. à 2 600 m. selon une progression géométrique d'un coefficient de 1.5. L'ajustement de la courbe empirique s'opère via une transformation logarithmique. Ces paramètres sont pour certains légèrement différents de ceux utilisés par Tannier and Thomas (2013). Nous ne comparerons donc pas les valeurs des dimensions obtenues avec celles des villes belges mais nous observerons si la relation entre les dimensions des aires urbaines et des agglomérations morphologiques sont identiques en Belgique et en France.

En complément de $D_s(AU)$ et $D_s(AM)$, des indices morphologiques plus classiques sont calculés : la densité bâtie de l'agglomération morphologique (Dens), la surface de l'agglomération (S) et la part des bâtiments de l'aire urbaine qui font partie de l'agglomération morphologique (%B).

6.3.3 Caractéristiques morphologiques du périmètre des 82 villes étudiées

Afin de caractériser la forme des frontières des agglomération (limites morphologiques), nous utilisons trois indices : la dimension fractale de boîte ($D_b(AM)$), la dendricité (δ) et la compacité (CO).

Les limites morphologiques (périmètres) des villes identifiées au moyen de *MorphoLim* se présentent sous la forme d'un ensemble de lignes fermées. Plus un périmètre est composé de nombreuses lignes fermées de longueurs variées et présentant de nombreuses indentations, plus sa dimension fractale est élevée. La dimension fractale de boîte de ces périmètres est estimée à partir des mêmes valeurs de paramètres que les dimensions de boîte des tissus bâtis. Une seule différence concerne la taille minimale des fenêtres de comptage. Selon les recommandations de Tannier and Thomas (2013), pour une ville donnée, la plus petite taille de fenêtre de comptage doit être supérieure au seuil de distance de dilatation appliqué pour délimiter la ville. En deçà de cette distance, l'enveloppe de la ville ne présente pas de caractéristique fractale. Le seuil de distance le plus grand de l'échantillon des 82 aires urbaines est celui de Marseille-Aix-en-Provence avec une valeur de 457 mètres. On fixe dès lors la taille minimale des fenêtres de comptage à 450 m. pour toutes les analyses fractales des périmètres des villes de l'échantillon.

L'indice de dendricité (δ) exprime le rapport entre la fractalité du tissu bâti de l'agglomération morphologique et la fractalité de son périmètre. Selon l'équation 6.1, un indice δ proche de zéro indique que la dimension fractale du tissu bâti est bien supérieure à celle du périmètre. Inversement, un indice δ supérieur à 1 indique que la dimension fractale du tissu bâti est inférieure à celle du périmètre : le tissu est relativement uniforme tandis que l'enveloppe morphologique est découpée en plusieurs lignes fermées et présente de nombreuses indentations de tailles différentes. Si δ est proche de 1 cela signifie que les deux dimensions sont identiques. Sur la figure 6.2, les dimensions de boite sont très proches pour les tissus bâtis mais pas pour les limites morphologiques. La $D_b(AM)$ de la limite morphologique (b) est plus élevée que celle de la limite morphologique (a). Il en découle que la dendricité de (b) est plus élevée.

$$\delta = 2 - \frac{D_s(AM)}{D_b(AM)} \tag{6.1}$$

L'indice de compacité (CO) exprime le rapport entre l'aire d'une agglomération morphologique et la longueur de son périmètre (équation 6.2). CO tend vers 1 lorsque le périmètre est lisse, comme dans le cas de figures

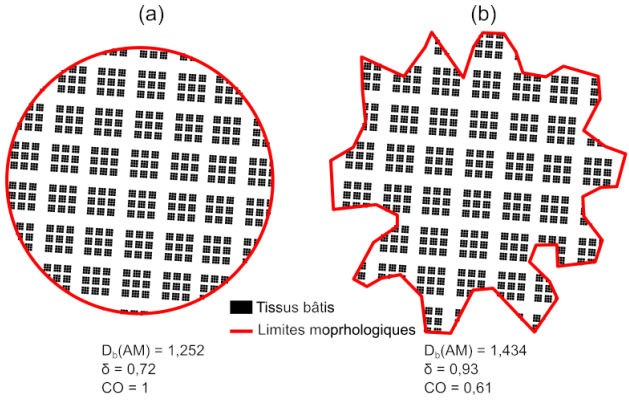


Figure 6.2. Dimension fractale de boite, indice de dendricité et indice de compacité pour deux agglomérations morphologiques théoriques

géométriques simples telles que le cercle ou le carré. CO tend vers 0 lorsque le périmètre est tortueux (nombreuses lignes fermées et indentations de tailles variées). En effet, dans ce cas, le périmètre est très long relativement à la surface de l'agglomération.

$$CO = \frac{2(\pi * A)^{1/2}}{P} \tag{6.2}$$

avec A l'aire de l'agglomération morphologique et P la longueur de son périmètre. La rapport entre l'aire et la longueure du périmètre de l'agglomération (a) (fig. 6.2) est plus élevé que ce même rapport pour l'agglomération (b). L'indice de compacité est donc plus élevé pour (a).

Des corrélations entre les différents indices morphologiques sont calculées avec le coefficient de corrélation de Pearson. Comme cet indice ne mesure que des corrélations linéaires entre deux variables, nous nous assurons de l'absence d'autres relations non linéaires via l'observation visuelle des représentations graphiques des relations entre les variables (non présentées dans ce chapitre).

6.4 Résultats

6.4.1 Délimitation des agglomérations morphologiques

Il a été possible de délimiter l'agglomération morphologique à l'intérieur de chacune des 82 aires urbaines. Les valeurs de courbure principale varient de -0,07 (Toulon) à -0,89 (La Rochelle) (voir la Fig 6.3 pour un échantillon de 6 courbes de dilatation présentant des points de courbure principale bien différents). Les valeurs de courbure principale les plus proches de 0 témoignent d'une structure très peu hiérarchique du tissu bâti dans laquelle les formes bâties de l'agglomération morphologique se distinguent assez peu de celles de sa périphérie. Inversement, une valeur de courbure principale éloignée de 0 montre que le tissu bâti de l'agglomération morphologique se distingue nettement de celui de sa périphérie d'un point de vue fractal.

Dans la grande majorité des cas, la courbe de dilatation empirique a été correctement estimée avec des polynômes de degré 4 à 6. Pour dix aires urbaines, il a toutefois été nécessaire d'augmenter le degré du polynôme à 7, 8, voire 9 pour Annecy et Rennes. Sur l'échantillon de 82 aires urbaines, le degré du polynôme n'est pas corrélé aux valeurs de courbure principale non plus qu'au seuil de distance de dilatation.

Le large spectre de seuils de distance de dilatation, de 68 m. à 457 m., montre la diversité morphologique des aires urbaines françaises sélectionnées. Les seuils de distance ne sont pas corrélés à la superficie des aires urbaines. Ces résultats sont conformes aux résultats obtenus par Sémécurbe et al. (2019) pour vingt aires urbaines de France métropolitaine. Nous avons déjà indiqué que des méthodes utilisant un seuil de distance fixé a priori pour délimiter des ensembles morphologiques ne permettent pas toujours de tenir compte des spécificités morphologiques des différents espaces urbains. Nous ajoutons ici qu'un seuil de 100 à 200 mètres, généralement adopté dans les études (Zahn, 1971; Yu et al., 2014; Caruso et al., 2017) pourrait n'être pertinent que pour 48% des aires urbaines de notre échantillon. En effet, seules quarante aires urbaines sur les 82 ont un seuil de distance compris entre 100 et 200 m.

Pour soixante-neuf aires urbaines, un agrégat morphologique apparaît clairement plus grand que les autres. Les agglomérations morphologiques de ces aires urbaines sont donc monocentriques. L'agglomération morphologique de dix aires urbaines est constituée de deux agrégats. Dans ces cas, l'agglomération est monocentrique si les deux agrégats sont contigus (Bordeaux, Perpignan, Nantes, Agen, Blois, Angers et Toulouse). En revanche, si les deux agrégats sont distants l'un de l'autre, l'agglomération est

dite polycentrique (Béthune, Quimper, Vienne). L'agglomération morphologique de Saint-Etienne est constituée de trois agrégats spatialement distincts. L'agglomération de Tours est composée de quatre agrégats contigus et l'agglomération d'Avignon en compte huit non contigus. La Figure 6.4 montre des exemples de distributions rang-taille correspondant à ces différents cas de figure.

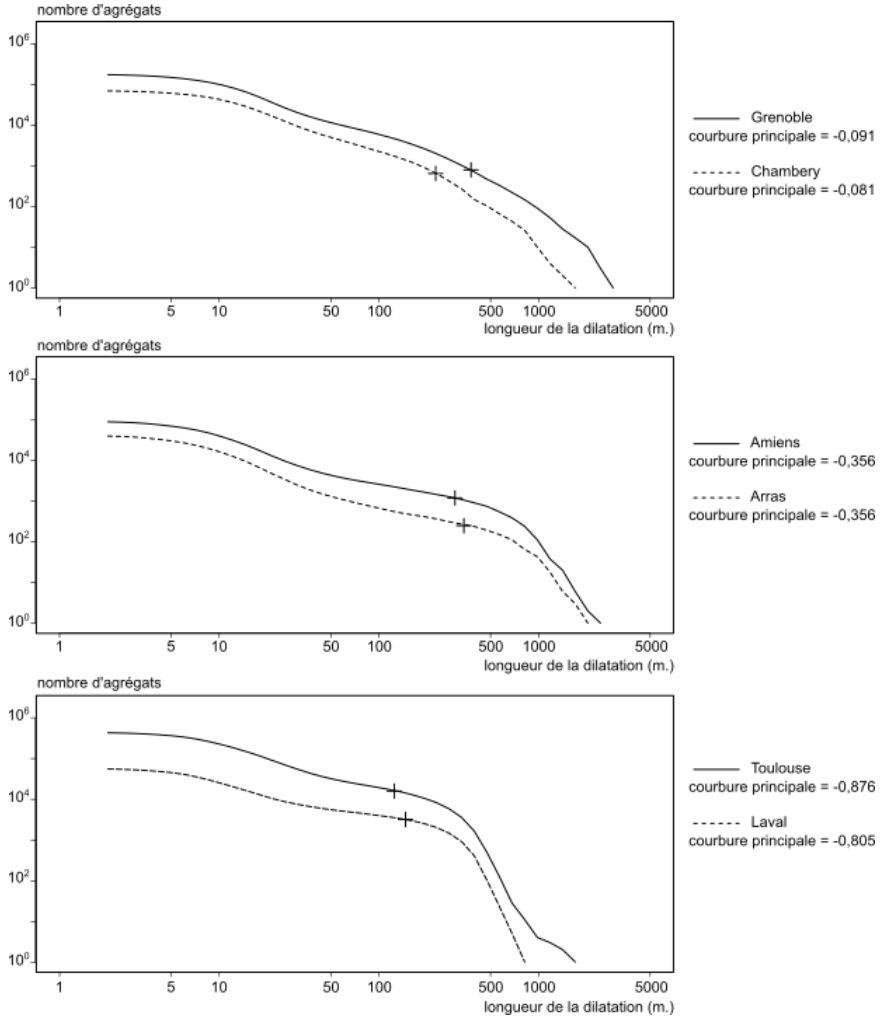


Figure 6.3. Courbes de dilatation et points de courbure principale (+) de six aires urbaines françaises

Les agglomérations morphologiques ont des superficies variant de 1 654 km^2 pour la plus grande (Paris) à 12 km^2 pour la plus petite (Lorient).

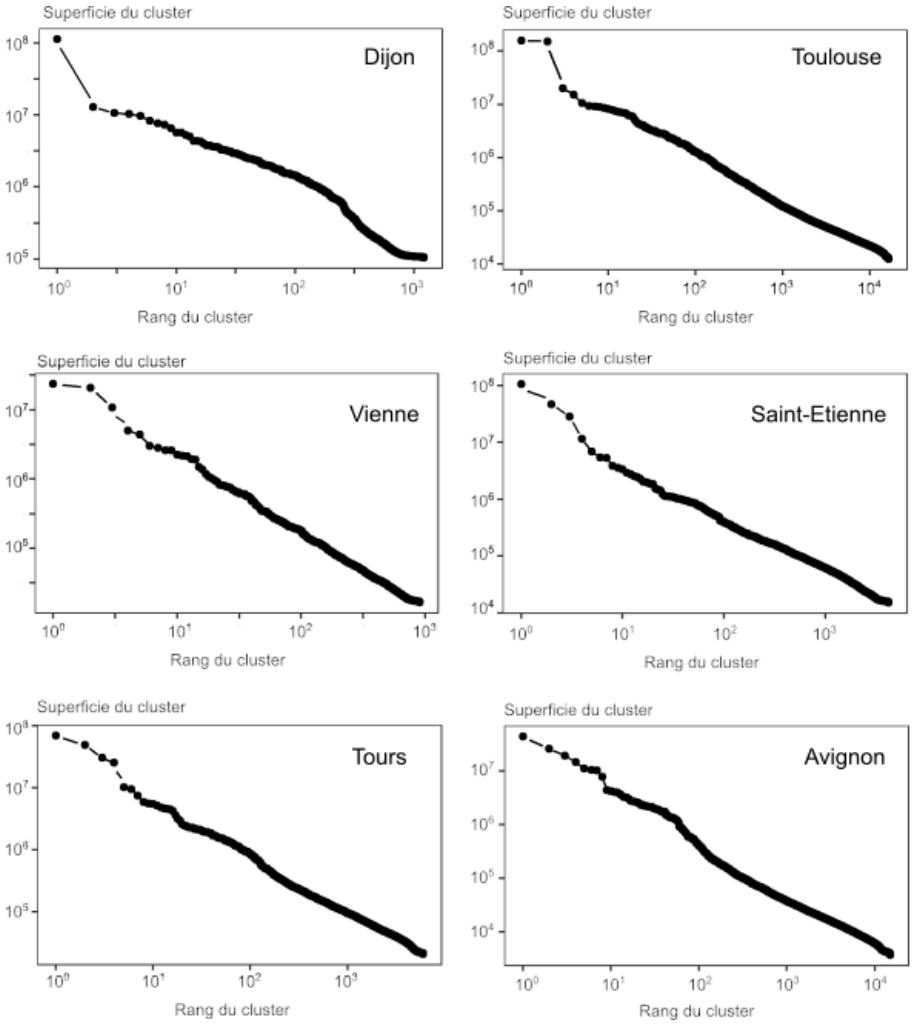


Figure 6.4. Distribution rang-taille des agrégats bâtis de six aires urbaines morphologiquement bien différentes

6.4.2 Morphologie des agglomérations et de leurs périphéries

Les dimensions fractales de boîte du tissu bâti des aires urbaines ($D_s(AU)$) et des agglomérations morphologiques ($D_s(AM)$) ont été calculées considérant douze tailles de fenêtres de comptage s'étalant sur trois ordres de grandeurs (de 30 m. à 2600 m.). Ceci répond aux recommandations établies pour l'estimation de dimensions fractales (voir Chapitre 4). L'estimation des dimensions fractales est globalement assez robuste au regard des p-valeurs et des R^2. Notons tout de même que, pour les tissus bâtis de certaines aires urbaines, le R^2 de l'estimation est légèrement en-dessous du seuil de 0,99. Il s'agit des aires urbaines de Toulouse, Rennes, Angers, Le Mans, Limoges, Laval, Agen, Roanne, Brive-la-Gaillarde, Montauban, Périgueux et Albi. Les valeurs de $D_s(AU)$ sont comprises entre 1,232 et 1,617. Les valeurs de $D_s(AM)$ sont, elles, comprises entre 1,524 et 1,784.

Comme attendu (Thomas et al., 2007; Tannier and Thomas, 2013; Sémécurbe, 2020), la densité bâtie des agglomérations morphologiques est proportionnelle à leur dimension fractale de boîte ($D_s(AM)$) (indice de corrélation de Pearson = 0,63, Tab. 6.1). La densité bâtie est également inversement proportionnelle (corrélation = -0,61) à la part de bâtiments de l'aire urbaine localisés dans l'agglomération morphologique (%B). Cependant, contrairement à ce qu'observent Tannier and Thomas (2013) pour la Belgique, $D_s(AM)$ et %B ont également une relation inversement proportionnelle (corrélation = -0,44). Ainsi, quand une agglomération morphologique est relativement petite par rapport à son aire urbaine, son tissu bâti a tendance à être plus dense et plutôt uniforme. A l'inverse, si une agglomération morphologique rassemble une grande partie des bâtiments de son aire urbaine, la distribution spatiale de ses bâtiments sera plus hétérogène et moins dense. Notons que c'est la différence de taille relative d'une agglomération morphologique par rapport son aire urbaine qui est ici commentée. En effet, comme le montrent Tannier and Thomas (2013) la superficie d'une agglomération n'est pas corrélée à sa dimension fractale. Il en est de même pour la superficie totale de l'aire urbaine.

	D_s(AU)	D_s(AM)	Dens	%B	S	L	D_b(AM)	δ
D_s(AM)	n.s							
Dens	n.s	0,63						
%B	0,50	-0,44	-0,61					
S	0,28	n.s	-0,30	0,44				
L	0,35	n.s	-0,25	0,42	0,89			
D_b(AM)	0,31	0,36	0,27	n.s	n.s	0,25		
δ	0,24	-0,28	n.s	n.s	n.s	0,26	0,79	
CO	-0,46	n.s	n.s	-0,27	-0,34	-0,53	-0,66	-0,69

Table 6.1. Coefficients de corrélation de Pearson entre les différents indices morphologiques caractérisant les 82 aires urbaines étudiées. ns : non significatif à un seuil de significativité égal à 0,05 ; S = superficie des AM ; Dens= densité bâtie des AM ; %B = part des bâtiments de l'AU qui sont dans l'AM ; L = longueur du périmètre.

Tannier and Thomas (2013) observent encore pour dix-huit agglomérations belges que la différence de dimension fractale entre l'agglomération morphologique et l'aire urbaine n'est pas particulièrement liée à la valeur de l'indice de courbure principale ou à %B. Nos résultats montrent, inversement, que ces mesures sont statistiquement liées entre elles dans le cas des 82 aires urbaines françaises. Tout d'abord, l'indice de courbure principale est corrélé à la différence de dimension fractale avec un coefficient de corrélation de Pearson égal à 0,64. Donc, en général, pour les aires urbaines françaises, plus la différence de dimensions est marquée entre une agglomération morphologique et son aire urbaine, plus l'indice de courbure principale est, en valeur absolue, élevé. L'indice de courbure principale mesure en effet la rupture la plus importante dans la logique d'emboîtement d'échelles fractal d'un tissu bâti. Si la rupture est très marquée (indice de courbure élevé), la différence de dimensions de part et d'autre du point de rupture est donc également importante. Notons que $D_s(AU)$ ne mesure pas la dimension fractale du tissu bâti hors agglomération morphologique mais celle de l'ensemble de l'aire urbaine. La différence entre les dimensions fractales des tissus bâtis de l'agglomération morphologique et hors agglomération pourrait donc être encore plus marquée que la différence avec la dimension fractale de l'aire urbaine dans son ensemble. Par ailleurs, au sein d'une même aire urbaine, les logiques de répartitions spatiales des bâtiments sont parfois très différentes entre l'agglomération morphologique et hors agglomération, ce qui peut expliquer pourquoi les estimation des dimensions $D_s(AU)$ sont globalement moins robustes (R^2 moins élevés) que celles de $D_s(AM)$. De plus, comme la part de bâtiments de l'aire urbaine localisés dans l'agglomération est inversement proportionnelle à la différence

de dimension fractale (-0,70), la différence entre $D_s(AM)$ et $D_s(AU)$ est potentiellement d'autant plus forte qu'une agglomération occupe une part restreinte de la superficie de l'aire urbaine.

Nous avons établi une première typologie des 82 aires urbaines étudiées basée uniquement sur les caractéristiques morphologiques des tissus bâtis. Cette typologie résulte d'une classification (dissimilarité euclidienne et agrégation de Ward) à partir des dimensions fractales de boîte des aires urbaines et des agglomérations morphologiques. Six groupes ont pu être identifiés (Tab. 6.2, Fig 6.5 et Fig. 6.6a). Les groupes 1, 3 et 6 rassemblent des agglomérations morphologiques dont les dimensions de boîte sont relativement identiques. A l'inverse, les dimensions de leurs aires urbaines sont bien différentes (Fig. 6.5). Les tissus bâtis appartenant au groupe 1 ont des $D_s(AU)$ légèrement plus faibles que leurs $D_s(AM)$. La différence entre $D_s(AM)$ et $D_s(AU)$ s'accentue dans le groupe 3 avec une diminution des dimensions des AU. Le groupe 6 est celui pour lequel la différence est la plus marquée entre $D_s(AM)$ et $D_s(AU)$, avec les $D_s(AU)$ les plus faibles. Globalement, les tissus bâtis des groupes 1, 3 et 6 sont plutôt concentrés dans des agrégats distants les uns des autres, cette tendance s'accentuant du groupe 1 au groupe 6 (groupe 1 – $D_s(AU)$ autour de 1,51 ; groupe 3 – $D_s(AU)$ autour de 1,41 ; groupe 6 – $D_s(AU)$ autour de 1,26). Les tissus bâtis des agglomérations morphologiques sont, eux, davantage connexes tout en présentant des lacunes internes de différentes tailles ($D_s(AM)$ autour de 1,64). Les groupes 2 et 5 ont des $D_s(AU)$ très proches les unes des autres mais des $D_s(AM)$ différentes. La différence entre $D_s(AU)$ et $D_s(AM)$ est moins marquée pour les tissus du groupe 2 que pour ceux du groupe 5. Dans l'aire urbaine, les tissus bâtis des groupes 2 et 5 sont assez proches de celles des groupes 1, 3 et 6 ($D_s(AU)$ autour de 1,41). Il en est de même pour les agglomérations morphologiques du groupe 2 ($D_s(AM)$ autour de 1,58) tandis que celles du groupe 5 remplissent davantage l'espace et leurs bâtiments sont répartis spatialement de manière plus uniforme ($D_s(AM)$ autour de 1,75). Enfin, dans les aires urbaines du groupe 4, les bâtiments ont tendance à être concentrés dans des agrégats distants les uns des autres ($D_s(AU)$ autour de 1,33) tandis que les tissus bâtis dans les agglomérations morphologiques sont davantage connexes tout en présentant des lacunes internes ($D_s(AM)$ autour de 1,65). La Fig. 6.7 présente un exemple de tissu caractéristique de chaque groupe.

Groupes	Aires urbaines	D_s(AM) – D_s(AU) (intervalle)	Description
1	Creil, Marseille-Aix-en-Provence, Béthune, Thionville, Douai - Lens, Le Havre, Saint-Nazaire, Avignon, Saint-Brieuc, Toulon, Mulhouse, Nice, Lyon	0,045 – 0,22	Faible différence entre Ds(AU) et Ds(AM). Ds(AU) élevées par rapport à celles de l'ensemble des villes.
2	Grenoble, Belfort, Calais, Bordeaux, Vienne, Montpellier, Ales, Chambéry, Quimper, Chalon-sur-Saône, Macon	0,123 – 0,218	Faible différence entre Ds(AU) et Ds(AM). Ds(AU) et Ds(AM) moyennes par rapport à celles de l'ensemble des villes.
3	Montauban, Saint-Etienne, Tarbes, Rouen, Boulogne-sur-Mer, Valence, Annecy, Caen, Cherbourg-en-Cotentin, Colmar, Brive-la-Gaillarde, Pau, Albi, Perpignan, Vannes, Nîmes	0,225 – 0,316	Différence moyenne entre Ds(AU) et Ds(AM). Ds(AM) plutôt élevées et Ds(AU) moyennes par rapport à celles de l'ensemble des villes.
4	Metz, Montbéliard, Périgueux, Tours, Roanne, La Roche-sur-Yon, Arras, Agen, Besançon, Poitiers, Niort, Angoulême, Bourg-en-Bresse, Evreux, La Rochelle, Béziers, Limoges, Cholet, Nancy, Laval, Orléans	0,255 – 0,391	Forte différence entre Ds(AU) et Ds(AM). Ds(AU) faibles par rapport à celles de l'ensemble des villes.
5	Lorient, Nantes, Paris, Brest, Clermont-Ferrand, Dunkerque, Le Mans, Toulouse, Angers, Rennes	0,281 – 0,390	Forte différence entre Ds(AU) et Ds(AM). Ds(AM) très élevées par rapport à celles de l'ensemble des villes.
6	Charleville-Mézières, Saint-Quentin, Troyes, Blois, Beauvais, Bourges, Carcassonne, Amiens, Chartres, Dijon, Reims	0,292 – 0,467	Forte différence entre Ds(AU) et Ds(AM). Ds(AU) très faibles par rapport à celles de l'ensemble des villes.

Table 6.2. Les aires urbaines classées selon leur dimension fractale de boîte $D_s(AU)$ et celle de leurs agglomérations morphologiques $D_s(MA)$

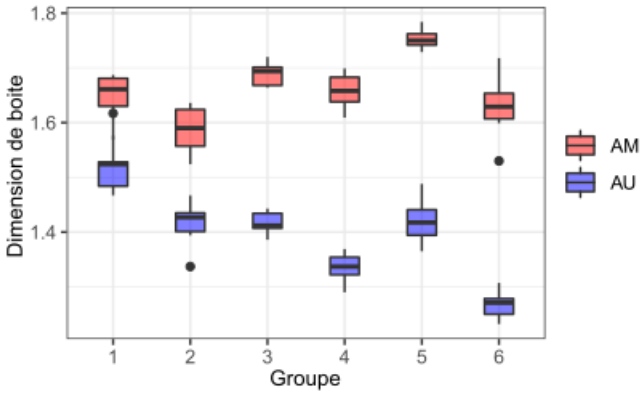


Figure 6.5. Dimensions fractales de boîte des aires urbaines (AU) et des agglomérations morphologiques (AM) pour chaque groupe issu de la classification effectuée sur ces dimensions

6.4.3 Caractéristiques morphologiques des périmètres des villes

Les valeurs de l'indice de dendricité sont globalement proches de 1, comprises dans un intervalle allant de 0,77 à 1,07, suggérant ainsi que les périmètres des agglomérations ont une fractalité comparable à celle des tissus bâtis. Cela est confirmé par l'indice de compacité qui est compris entre 0,022 et 0,32 avec une valeur moyenne de 0,1 pour l'ensemble des agglomérations morphologiques. Les dimensions de boîte calculées sur les périmètres sont aussi bien supérieures à 1, dans un intervalle compris entre 1,371 pour le périmètre le moins complexe (Troyes) et 1,709 pour le périmètre le plus complexe (Limoges). Contrairement à une partie des résultats obtenus par Tannier and Thomas (2013) pour dix-huit villes belges, les trois indices sont corrélés entre eux (Tab. 6.1). En règle générale, plus un périmètre est complexe ($D_b(AM)$ élevée), plus l'agglomération morphologique est dendritique (δ élevée) et moins elle est compacte (CO faible).

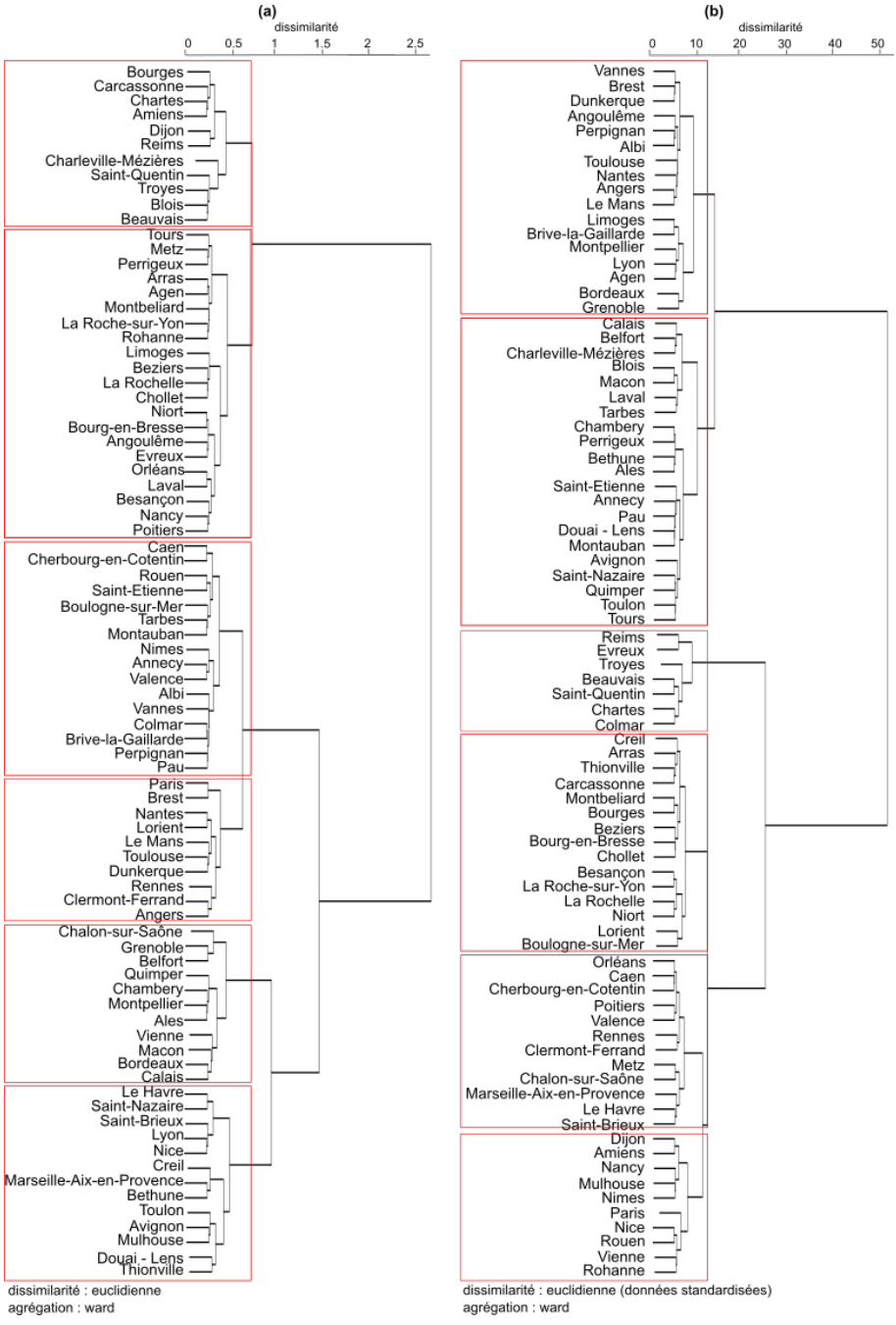


Figure 6.6. Dendrogramme des classifications de 82 aires urbaines françaises au regard des caractéristiques de leur tissu bâti (a) et de leur périmètre (b)

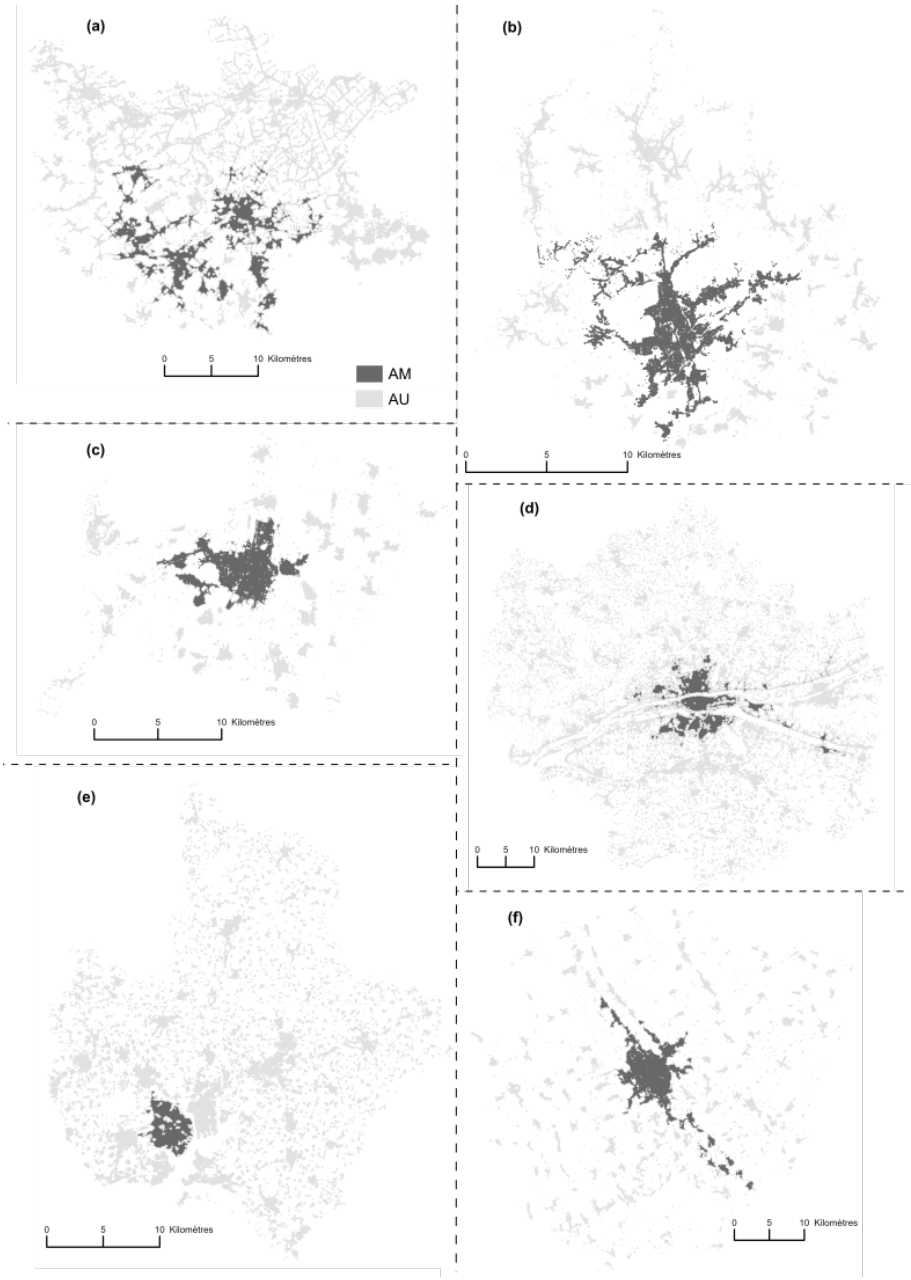


Figure 6.7. Six tissus bâtis représentatifs de chaque groupe de la classification selon $D_s(AM)$ et $D_s(AU)$: (a) Béthune – groupe 1 ; (b) Belfort – groupe 2 ; (c) Colmar – groupe 3 ; (d) Tours – groupe 4 ; (e) Lorient – groupe 5 ; (f) Troyes – groupe 6

Il est important de garder en tête que le périmètre d'une agglomération morphologique n'est en général pas composé d'une ligne unique mais bien d'un ensemble de lignes fermées plus ou moins nombreuses et indentées, dont le nombre et la longueur renseignent sur la fragmentation et la complexité de ce périmètre. Le périmètre des 82 agglomérations morphologiques de notre échantillon comporte un nombre de lignes fermées (fragments) qui varie entre 10 et 2364 (valeur médiane = 152). La relation statistique entre le nombre de fragments et les valeurs des trois indicateurs morphologiques considérés est assez nette, comme le montre la Figure 6.8.

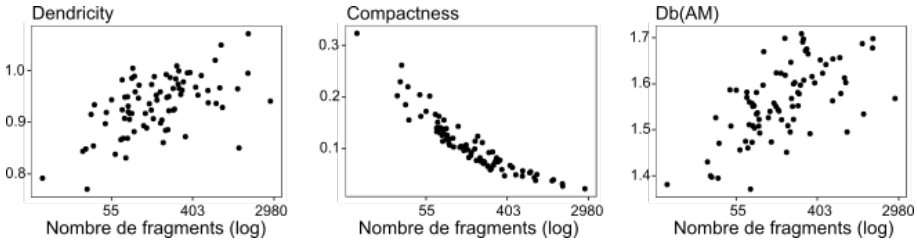


Figure 6.8. Relations entre le nombre de lignes fermées (fragments) composant le périmètre de chaque agglomération morphologique et les différents indices la caractérisant.

Contrairement aux résultats obtenus par Tannier and Thomas (2013), la dimension fractale du périmètre ($D_b(AM)$) n'est pas corrélée à la surface des agglomérations. En ce sens, nos résultats sont conformes à l'hypothèse émise par Chen (2011) pour qui une augmentation de la superficie d'une agglomération n'implique pas une augmentation de la complexité de son périmètre. Tannier and Thomas (2013) observent par ailleurs une corrélation négative entre la dendricité et la densité bâtie des agglomérations belges. Un tel résultat ne se retrouve pas dans le cas des 82 agglomérations françaises étudiées.

Une classification des aires urbaines sur la base des indices standardisés de dendricité, de compacité et de dimension fractale de leur périmètre permet d'identifier six groupes (Fig. 6.6b). Du groupe 1 au groupe 6, les valeurs de l'indice de dendricité δ et de la dimension fractale des périmètres $D_b(MA)$ croissent. Ces deux valeurs sont fortement corrélées (Tab. 6.1), c'est-à-dire que plus un périmètre est fragmenté et présente de nombreuses indentations (δ élevée), plus il est caractérisé par une valeur élevée de dimension fractale. A l'inverse, les valeurs de l'indice de compacité décroissent du groupe 1 au groupe 6. Cet indice est en effet inversement proportionnel aux deux autres (Tab. 6.1) : plus un périmètre présente de nombreuses indentations et fragments d'ampleur variée, moins la valeur de l'indice de compacité est

élevée. Une exception est toutefois à noter pour le groupe 4 pour lequel l'indice de compacité apparaît anormalement élevé au regard des valeurs de dendricité et $D_b(MA)$. En effet, bien que les périmètres des agglomérations morphologiques du groupe 4 aient des valeurs de $D_b(MA)$ et δ plus ou moins égales à celles du groupe 3, leur indice de compacité est bien supérieur. Cette différence s'explique par la moindre fragmentation des périmètres des villes du groupe 4 comparées à celles du groupe 3. Les enveloppes du groupe 4 sont donc plus compactes que celles du groupe 3. La Figure 6.9 présente un exemple de périmètre caractéristique de chaque groupe de villes.

La typologie obtenue sur la base des trois indices caractérisant les périmètres (Tab. 6.3) est bien différente de celle obtenue sur la base des indices caractérisant les tissus bâtis. Cela montre que les caractéristiques morphologiques d'un tissu bâti ne sont pas identiques à celles de son périmètre, ce qui a déjà été remarqué par Tannier and Thomas (2013). On observe en effet une très faible corrélation entre les dimensions $D_s(AM)$ et $D_b(AM)$ ou entre la densité bâtie d'une agglomération et la $D_b(AM)$ de son périmètre. Comme Tannier and Thomas (2013), nous n'observons pas de lien statistique entre $D_s(AM)$ et CO (compacité), ni entre δ (dendricité) et S (superficie des agglomérations) (Tab. 6.1).

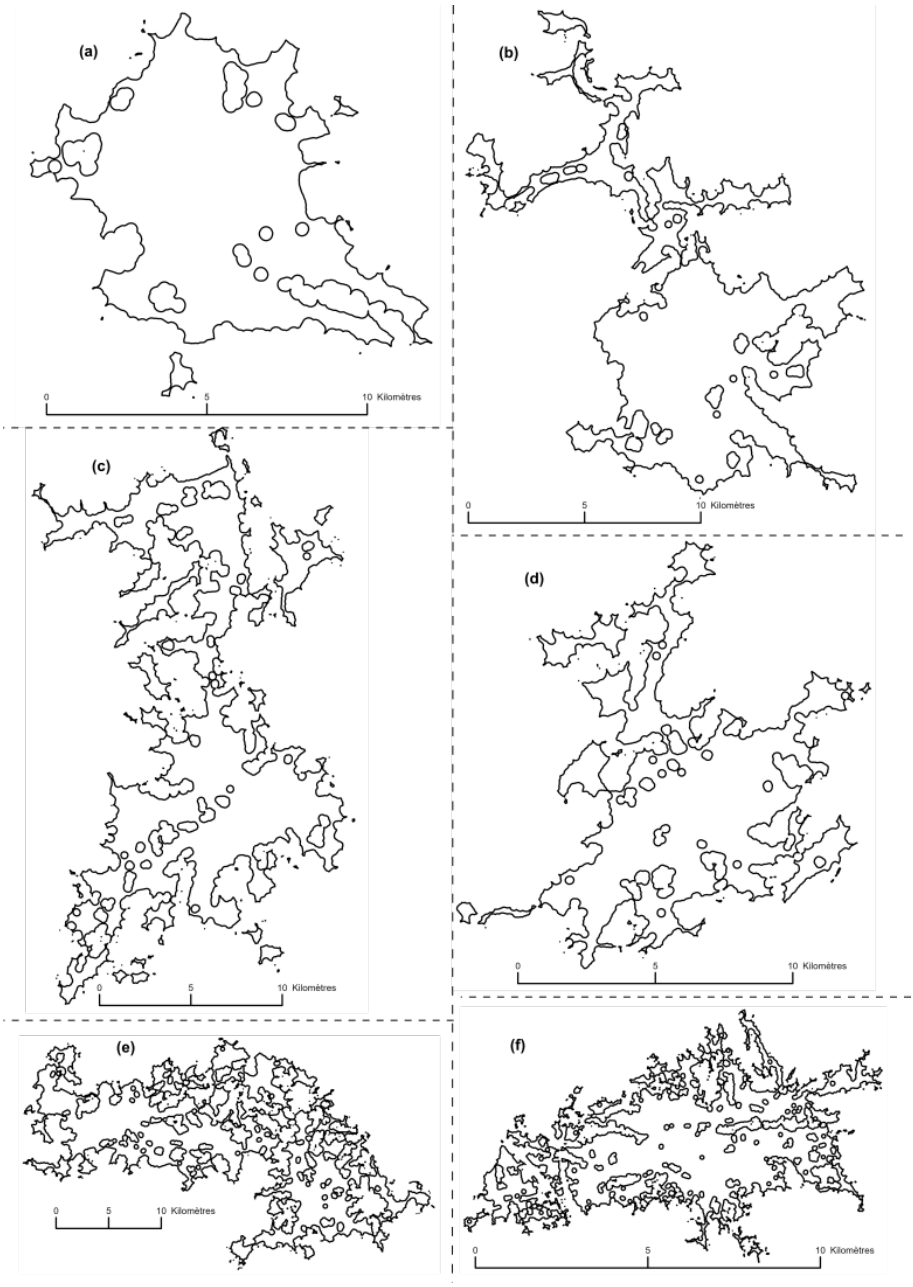


Figure 6.9. Six périmètres d'agglomérations morphologiques représentatifs de chaque groupe de la classification selon δ, CO et $D_b(MA)$: (a) Reims – groupe 1 ; (b) Nancy – groupe 2 ; (c) Metz – groupe 3 ; (d) Besançon – groupe 4 ; (e) Douai-Lens – groupe 5 ; (f) Brive-la-Gaillarde – groupe 6

Groupes	Aires urbaines	Indices (intervalle)
1	Reims, Troyes, Chartres, Colmar, Beauvais, Saint-Quentin, Evreux	δ : 0,77 - 0,85 CO : 0,14 - 0,32 D_b(AM) : 1,371- 1,456
2	Paris, Nice, Rouen, Nancy, Dijon, Amiens, Mulhouse, Nîmes, Vienne, Roanne	δ : 0,85 - 0,90 CO : 0,03 - 0,17 D_b(AM) : 1,451 - 1,534
3	Marseille-Aix-en-Provence, Rennes, Clermont-Ferrand, Orléans, Caen, Metz, Le Havre, Poitiers, Valence, Saint-Brieuc, Chalon-sur-Saône, Cherbourg-en-Cotentin	δ : 0,89 - 0,96 CO : 0,05 - 0,11 D_b(AM) : 1,493 - 1,619
4	Besançon, Lorient, La Rochelle, Montbéliard, Béziers, Niort, Bourges, Boulogne-sur-Mer, Thionville, Arras, Creil, Bourg-en-Bresse, La Roche-sur-Yon, Cholet, Carcassonne	δ : 0,89 - 0,95 CO : 0,11 - 0,20 D_b(AM) : 1,471 - 1,587
5	Toulon, Douai-Lens, Avignon, Saint-Etienne, Tours, Béthune, Pau, Annecy, Chambéry, Saint-Nazaire, Calais, Blois, Quimper, Laval, Tarbes, Belfort, Ales, Charleville-Mézières, Montauban, Perigueux, Macon	δ : 0,94 - 1,00 CO : 0,02 - 0,13 D_b(AM) : 1,523 - 1,624
6	Lyon, Toulouse, Bordeaux, Nantes, Grenoble, Montpellier, Angers, Le Mans, Brest, Perpignan, Limoges, Dunkerque, Angoulème, Vannes, Agen, Brive-la-Gaillarde, Albi	δ : 0,93 - 1,07 CO : 0,03 - 0,10 D_b(AM) : 1,603 - 1,709

Table 6.3. Les aires urbaines classées selon la dendricité, la compacité et la dimension fractale standardisées des périmètres de leur agglomération morphologique.

6.4.4 Typologie à partir des indices morphologiques caractérisant les tissu bâti et leur périmètre

Une analyse en composantes principales (ACP) suivie d'une classification ont été effectuées sur un ensemble de variables caractérisant à la fois les tissus bâtis et les périmètres des villes. Les variables (préalablement standardisées) suivantes ont été utilisées : dimension fractale de boîte du tissu bâti ($D_s(AM)$), superficie (S), densité bâtie de l'agglomération morphologique ($Dens$), longueur du périmètre (L), dimension fractale de boîte du périmètre ($D_b(AM)$), pourcentage des bâtiments de l'aire urbaine présents dans l'agglomération morphologique ($\%B$), dendricité (δ), compacité (CO).

Trois facteurs principaux de l'ACP permettent d'expliquer 86% de la variance des données (Tab. 6.4). Le facteur 1 est fortement corrélé aux variables « Superficie » et « Longueur du périmètre ». Ce premier facteur décrit donc la taille des villes (i. e. des agglomérations morphologiques). Le facteur 2 est corrélé aux variables $Dens$, $D_s(AM)$ et $\%B$. Ce facteur 2 synthétise donc l'information ayant trait aux caractéristiques des tissus bâtis. Enfin, le facteur 3 synthétise les variables relatives aux caractéristiques des périmètres puisqu'il est corrélé aux variables $D_b(AM)$, δ et CO. La séparation des variables décrivant les caractéristiques bâties de celles décrivant les périmètres dans deux facteurs différents démontre encore une fois leur différence et donc leur complémentarité pour décrire la forme des villes.

Une classification (dissimilarité euclidienne, agrégation de Ward, nombre de classes défini par la minimisation d'un indice d'entropie et la maximisation des indices de Dunn et de Calinski et Harabasz) opérée sur les notes factorielles de l'analyse en composantes principales aboutit à une typologie des 82 aires urbaines en six groupes (Fig. 6.10, Tab. 6.5 et Fig. 6.12). Chaque groupe est caractérisé par la taille des agglomérations (facteur 1) qu'il rassemble ainsi que par les caractéristiques des tissus bâtis (facteur 2) et les caractéristiques des périmètres (facteur 3) de ces agglomérations (Fig. 6.11). Le groupe 1 est constitué uniquement de Paris qui est de loin l'agglomération la plus grande. Le groupe 2 rassemble les agglomérations de taille moyenne ayant un tissu bâti ainsi qu'un périmètre bâti assez complexe. Le groupe 3 est constitué des agglomérations de taille moyenne dont le tissu bâti apparaît peu complexe contrairement à leur périmètre. Les agglomérations des groupes 4, 5 et 6 sont de taille identique. Elles présentent des tissus bâtis de fractalité décroissante de 4 à 6. Les agglomérations du groupe 5 se différencient de celles des groupes 4 et 6 de par la moindre complexité de leur périmètre.

Variables de départ	**Facteur 1** Taille de l'agglomération morphologique	**Facteur 2** Caractéristiques morphologiques des tissus bâtis	**Facteur 3** Caractéristiques morphologiques des périmètres
S	**0,97**	-	-
L	**0,91**	-0,11	0,27
Dens	-0,22	**0,86**	-
D_s(MA)	0,17	**0,91**	-
%B	0,45	- **0,69**	-
D_b(MA)	-	0,34	**0,90**
δ	-	-0,24	**0,95**
CO	-0,38	-	- **0,81**
Part de la variance expliquée (total = 86%)	28%	28%	30%

Table 6.4. Saturations de l'analyse en composantes principales (corrélation entre les facteurs principaux et les variables d'entrée)

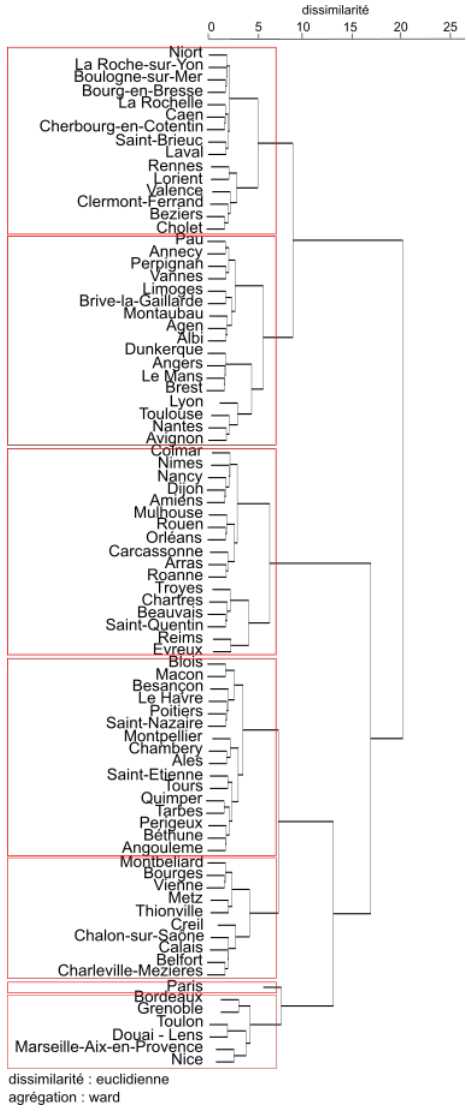


Figure 6.10. Dendrogramme de la classification de 82 aires urbaines françaises au regard des notes factorielles d'une analyse en composantes principales effectuée sur un ensemble de variables caractérisant à la fois les tissus bâtis et les périmètres des villes.

Groupes	Agglomérations Morphologiques	Caractéristiques
1	Paris	La plus grand AM, tissu bâti uniforme, périmètre assez peu complexe
2	Lyon, Toulouse, Nantes, Avignon, Angers, Le Mans, Brest, Perpignan, Limoges, Dunkerque, Pau, Annecy, Agen, Brive-la-Gaillarde, Montauban, Albi	AM de taille moyennement grande, tissu bâti uniforme, périmètre complexe
3	Marseille - Aix-en-Provence, Bordeaux, Nice, Grenoble, Toulon, Douai – Lens	AM de taille moyennement grande, tissu bâti présentant des concentrations locales de bâtiments à différentes échelles, périmètre complexe
4	Rennes, Clermont-Ferrand, Caen, Lorient, La Rochelle, Valence, Saint-Brieuc, Béziers, Niort, Boulogne-sur-Mer, Cherbourg-en-Cotentin, Laval, Bourg-en-Bresse, La Roche-sur-Yon, Cholet	AM de taille moyenne, tissu bâti plutôt uniforme, périmètre relativement peu complexe
5	Rouen, Nancy, Orléans, Dijon, Reims, Amiens, Mulhouse, Nîmes, Troyes, Chartres, Arras, Colmar, Beauvais, Saint-Quentin, Evreux, Roanne, Carcassonne	AM de taille moyenne, tissu bâti moyennement uniforme, périmètre très peu complexe
6	Montpellier, Saint-Etienne, Tours, Metz, Béthune, Le Havre, Poitiers, Besançon, Chambéry, Saint-Nazaire, Angoulême, Montbéliard, Bourges, Thionville, Chalon-sur-Saône, Calais, Blois, Quimper, Creil, Tarbes, Belfort, Vienne, Alès, Charleville-Mézières, Périgueux, Mâcon	AM de taille moyenne, tissu bâti présentant des concentrations locales de bâtiments à différentes échelles, périmètre relativement peu complexe

Table 6.5. Aires urbaines françaises classées selon la dendricité, la compacité et la dimension fractale standardisées des périmètres de leur agglomération morphologique.

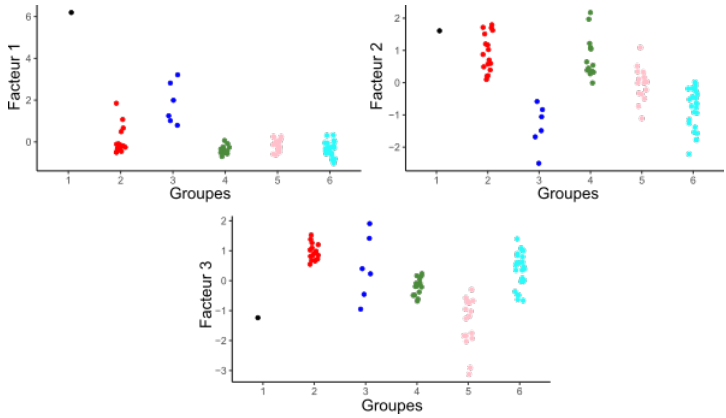


Figure 6.11. Notes factorielles des agglomérations morphologiques en fonction des groupes de la classification (troisième typologie)

6.5 Conclusion

Au moyen de la méthode *MorphoLim*, nous avons pu identifier la limite morphologique de 82 villes françaises monocentriques ou polycentriques. Nous avons vu que plus la différence d'organisation spatiale des bâtiments était marquée entre une agglomération morphologique (i. e. une ville) et son aire urbaine, plus *MorphoLim* avait de chances de mesurer un indice de courbure principale élevé en valeur absolue, ce qui marque une rupture nette dans la logique fractale d'emboîtement d'échelles de la distribution spatiale des bâtiments. Entre notre étude et celle de Tannier and Thomas (2013), nous avons relevé des points communs mais aussi des différences concernant la covariation entre les variables morphologiques caractérisant les villes. Ces différences concernent surtout les indices incluant une mesure de dimension fractale. Une des raisons qui peut expliquer les différences de covariation des indices entre les deux études réside dans les différentes manières qu'elles ont eu de mesurer les dimensions fractales. En effet, dans le présent chapitre, nous avons strictement adopté les mêmes paramètres d'estimation des dimensions fractales pour toutes les zones d'étude (cf. Chapitre 4). Dans les analyses de Tannier and Thomas (2013), en revanche, le nombre de fenêtres de comptage variait légèrement d'une ville à l'autre. Quoi qu'il en soit, une source majeure de différence entre nos résultats et ceux de Tannier and Thomas (2013) est sûrement que les aires urbaines étudiées, à savoir les aires urbaines belges, d'une part, et les aires urbaines françaises, d'autre part, présentent des caractéristiques morphologiques bien différentes. Tannier et al. (2011) délimitaient déjà des villes belges et françaises à l'aide de *MorphoLim*

et notaient que les villes belges avaient généralement un indice de courbure principale nettement inférieur à celui des villes françaises. Une délimitation moins marquée des villes belges par rapport aux villes françaises, en raison d'un étalement urbain important (Caruso, 2001), pourrait être à l'origine des différences de résultats entre la présente étude et celle de Tannier and Thomas (2013). Notons encore que les analyses de 2013 ont été effectuées sur un échantillon de dix-huit villes tandis que notre étude a concerné un échantillon de 82 villes. L'échantillon plus important amène sans doute à des résultats statistiquement plus robustes.

Comme énoncé dans le chapitre précédent (Chapitre 5), il est fondamental d'être conscient de l'importance des choix méthodologiques (données, méthodes et étendues des zones d'étude) présidant à la délimitation morphologique des villes. Chaque aspect doit être prudemment considéré afin de proposer une délimitation de la ville qui réponde aux questions et aux objectifs visés. Dans le présent chapitre, nous nous sommes particulièrement intéressé aux résultats proposés par *MorphoLim*. Nous avons notamment montré que *MorphoLim* trace une frontière entre deux espaces dont la logique d'emboîtement d'échelle est la plus différente. Lorsqu'il y a une différence claire entre un espace urbain et son espace rural, l'indice de courbure de *MorphoLim* est élevé. A l'inverse, lorsque l'urbanisation est plus diffuse, l'indice de courbure est faible (proche de 0). Nous avons également vu dans ce chapitre qu'une agglomération urbaine peut être décrite à l'aide de trois types de variables. Des variables décrivant sa taille (superficie et longueur du périmètre), son tissu bâti (densité bâtie, $D_s(AM)$, %B) et la forme de son périmètre ($D_b(AM)$, δ, CO). Ces trois catégories de variables apportent chacune des informations complémentaires et permettent ensemble de mieux décrire chaque agglomération que séparément.

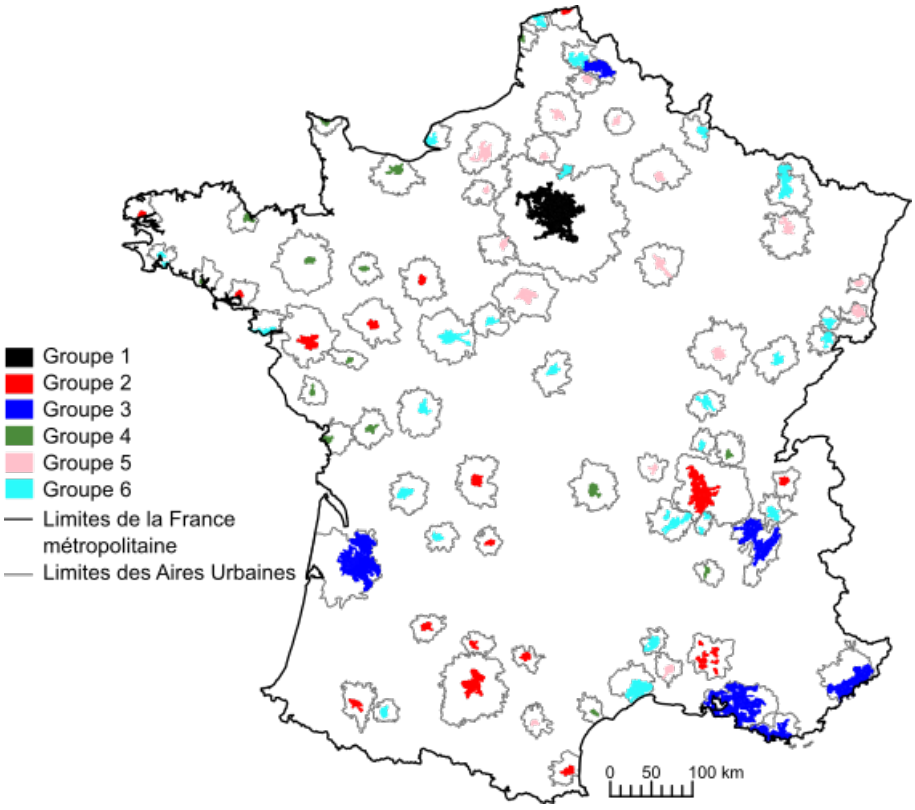


Figure 6.12. Typologie de 82 aires urbaines françaises selon leur forme. (Analyse en composantes principales suivie d'un classification ascendante hiérarchique sur 9 variables caractérisant les tissus bâtis et les périmètres des villes).

7

Delineation of a « diffuse city » at ground level (2D) and according to the number of floors (3D)

This chapter uses *MorphoLim* to delineate the agglomeration of Valencia, an urban area on the Spanish Mediterranean coast, which has been affected by an intense process of urban growth. The third dimension is incorporated into the analysis, applying the methodology in a differentiated way to buildings according to the number of floors. The use of the third dimension provides a new perspective in the morphological analysis of urban spaces, helps to differentiate the compact urban space from that affected by processes of dispersed urbanisation and improves the identification and characterisation of morphologies.

7.1 Introduction

Chapter 5 and Chapter 6 showed that *MorphoLim* can be used to delineate a more or less concentrated/diffuse urban agglomeration within an urban area. But can this method be used to delimit an agglomeration when the urbanisation is very diffuse as is the case for the urban area of Valencia in Spain? Can the use of the height of buildings (and not just the footprint as in the previous chapters) help? These are the questions raised by this chapter.

In Spain, as in many other countries, there have been several territorial processes in recent years that have led to a change in the city model. Improvements in communication and transportation systems, new residential preferences, and, in some cases, private interests not controlled by public policies, have led to the extension of suburbanisation processes. These processes giving rise to what has been called the "*dispersed city*" or "*diffuse city*" (Nel.Io, 1996; Indovina, 1998; Indovina and Nel.Io, 2012). The new city model however generates serious problems in the field of mobility, environmental pollution and social cohesion, as well as in the governance of urban spaces. The new "*dispersed city*" exceeds the usual political-administrative boundaries, requiring new planning frameworks that transcend the municipal sphere and allow intervention in this largely disjointed, fragmented and contradictory space (Romero, 2009). It is then necessary to delimit the territorial area affected by the new urban process.

However, one of the first repercussions of the new urban processes is precisely the growing difficulty in delimiting metropolitan spaces. Consequently, the delineation of cities and, in particular, metropolitan agglomerations, has become an object of reflection in Spain both in academic and at different political-administrative level (Cheshire and Gornostaeva, 2002; Parr, 2007; Redfearn, 2007; Bode, 2008; Cörvers et al., 2009). As we have seen in Section 1.2, two approaches dominate for delimiting urban spaces: the functional approach (see e.g. for Spain Marmolejo Duarte et al., 2010; Roca Cladera et al., 2012; Feria et al., 2015; Feria Toribio and Martínez Bernabeu, 2016; Feria Toribio et al., 2018) and the morphological approach (see e.g. for spain Martinez de Lejarza and Martinez de Lejarza, 2002; Marmolejo et al., 2013; Arribas-Bel et al., 2019). We will see in Section 7.2 some examples of delineations (and the limits they encounter) of Spanish urban area. As we already seen, both approaches are fraught with bias. The lack of data, which when it exists does not cover the whole territory (as in the case of mobility surveys) or is not at the appropriate territorial scale, is the major bias of the functional approach. The use of a fixed threshold is one of the major bias the morphological approach.

The aim of this chapter is to morphologically delineate a Spanish agglomeration affected by diffuse urbanisation, without using an a priori fixed threshold. To this end, we use *MorphoLim* to delineate the morphological agglomeration of Valencia. Valencia is an urban agglomeration located on the Spanish Mediterranean coast that has been affected, since the last quarter of the 20^{th} century, by an intense urban growth characterized by the predominance of the dispersed city. A first objective of this chapter is to delimit the main morphological agglomerations of Valencia. Does *MorphoLim* provide a new and relevant delineation for a "diffuse city"?

In addition to the application in a geographical environment characterized by the intensity of its diffuse urbanization process, this chapter presents another contribution: the incorporation of the third dimension in the morphological delineation of the urban space. In the previous chapter, we focused on the study of urban space with a particular attention to the description and analysis of built-up pattern and the shape of the cities border. As in many studies using fractal dimension, the basic information was vector mapping representing the footprint of buildings in two dimensions (2D). However, urban morphology does not only refer to the shape and distance of the floor plan of the built space, but is also established in a three-dimensional space. Indeed, the height of the buildings is a fundamental morphological variable in the identification and differentiation of the urban space with respect to rural space, and also intervenes in its internal differentiation (CBD with respect to residential neighbourhoods, popular neighbourhoods with respect to middle and upper class urbanizations, etc.). Moreover, this aspect is linked to essential urban characteristics such as land value or urban social structure. Consequently, in this chapter we apply *MorphoLim* to different buildings spatial distribution by differentiating buildings according to their heights (number of floors). We incorporate somehow the third dimension (3D) in the delineation of urban space.

This chapter is structured as follows: Section 7.2 discusses the delineation issues for the metropolitan area of Valencia and the proposed alternatives found in the literature. Section 7.3 presents the materials and method used. Section 7.4 shows the results. Section 7.5 concludes.

7.2 Delineation issues and proposed alternatives

The Spanish Mediterranean Region is characterized since the beginning of the 20^{th} century by uncontrolled urban sprawl. The urban sprawl, driven by the housing-price bubble, leads to a territorial pattern characterized by the predominance of dispersed urbanization. It is the case of the Valencia Metropolitan Area, where, according to data from *Corine Land Cover* (National Geographic Institute), between 2000 and 2006 almost 46 million m^2 of metropolitan land were transformed into urbanized land. It come particularly from a growth of diffuse urbanization in the second metropolitan crown along new high-capacity road axes and new rail transport lines (González et al., 2018; Burriel, 2009; Rullan, 2011). This process has continued until today, and, despite the impact of the economic crisis in 2007-2008, the growth rate of developable land continued to exceed that of urbanized land until 2013. This is mainly due to the conservation of the expectations of the municipalities, which hoped to recover the pre-crisis rate of construction (Farinós Dasí et al., 2019).

The metropolitan area of Valencia is characterized by the conurbation of municipalities derived from agricultural and industrial growth in an environment with great agricultural and landscape richness: the *Huerta de Valencia*. As a result of the conurbation, a strong urban continuity appears and it is not easily possible to identify elements of discontinuity or morphological rupture that allow the delineation of the urban agglomeration. Indeed, both at an academic and political level, multiple non-coincident delineations are available. Some of them carried out with unclear criteria.

The most common delineation is the one made by the *Ministry of Development of the Government of Spain* (de Fomento, 2018). In this delineation, municipalities are aggregated based on different demographic, economic/housing indicators and information on land use and transport networks. But the methodology do not clearly specify the criteria and thresholds used. In the case of the metropolitan area of Valencia, this delimitation coincides with the historical region and conforms to the 44 municipalities belonging to the administrative metropolitan area (*Consell Metropolitá de l'Horta*) in force between 1986 and 2000. However, the subsequent expansion of the urbanized space makes this delineation, which corresponds rather to the situation existing in the years 1970-1980, insufficient as a framework for the analysis of current territorial processes.

The *Territorial Strategy of the Valencian Community* (a framework document for land use planning in the region) as well as the *Territorial Plan for the Metropolitan Area of Valencia* (currently under preparation) distinguish between an "*Integrated Urban Area*" of 58 municipalities and a broader "*Functional Area*" composed of 90 municipalities. The former has been delimited on the basis of mainly morphological criteria (specifically the continuity of the residential built-up area based on the 2006 land use mapping) using an aggregation threshold of 1 km (Salom, 2011). For its part, the functional area is based on criteria of residence-work mobility and commuting for the provision of different services, although administrative and other criteria have also been introduced in this delineation (see Valenciana, 2011, p.45). The population considered ranges, depending on the delineation, between 1.5 and 1.8 million inhabitants.

However, these two delineations present different issues: in particular, the use of the municipal boundaries as basic spatial units and the arbitrariness of the thresholds used. Therefore, other methodologies based on the aggregation of road network nodes have been made in accordance with the work of Jiang and Jia (2011) and Borruso (2003). The results were quite satisfactory in the delineation and characterization of different Spanish metropolitan areas, including Valencia (Salom and Albertos, 2010, 2014). These proposals made it possible to analyse the phenomenon of the "diffuse city" in the Valencian Community. They also identifying and characterizing not only the most important urban areas but also the urbanized continuums of the coast and the dispersed areas of the metropolitan peripheries. Despite the diversity of territorial patterns and urban morphologies, *threshold values* of road network density were identified that allowed differentiating urban land, continuous or discontinuous, from other land uses for all the cities studied. Based on these *threshold values* of road network density, and through the use of Geographic Information Systems, two relevant spatial areas were delimited. The first corresponds to an urban-metropolitan area that mainly reflects the result of continuous urbanization processes better adjusted to the traditional concept of city. A second area is an urban region that encompasses areas of dispersed residential, industrial and commercial urbanization and conurbations derived from the coalescence of small and medium-sized nuclei close to the metropolitan area.

The intensification of urban sprawl processes since 2007 (year of the data used by Salom and Albertos (2010, 2014)), as well as the collapse of the housing-price bubble since 2008 make it difficult to apply these methodologies to more recent years. Thus, an update of this method with more precise data updated to 2014 (Salom and Zornoza, 2015) showed how the methodology

continued to be valid for identifying compact urbanization areas. However, it presented overestimated results in the case of dispersed urbanization areas due to the continuity of urbanization in the coastal zone, the processes of intense suburbanization in inland areas and the notable presence of "non-functional" road networks (i.e. networks corresponding to urban spaces under construction but not occupied due to the collapse of the housing-price bubble).

Given the limitations of the methodologies described above to accurately capture the new urban realities, we decide to use *MorphoLim*. There are two reasons for our choice:

- with *MorphoLim*, it is possible to identify the limits of a agglomeration from the structure of the built space itself without requiring the application of a predetermined distance threshold
- morphological agglomerations within the metropolitan framework located at the margin or distant from the main nucleus and whose dynamics are sometimes relatively independent of the municipal political-administrative limits can be delineated by *MorphoLim*.

7.3 Materials and Methods

7.3.1 Study area and data input

The study area (see Fig.7.1) is the one defined from functional criteria by Feria Toribio et al. (2018). The main criterion applied for its delineation is the residence-work mobility flows although corrected by some demographic criteria. It includes 78 municipalities and totals 1 885 533 inhabitants in 2019, occupying an area of 2212 km^2. In this study area, we can identify different spaces. The central municipality, Valencia (794 288 inhabitants in 2019) with some small municipalities totally urbanised. The first metropolitan crown, constituted, especially in the western and southern sectors, by municipalities with over 20 000 inhabitants, which have become residential neighbourhoods, industrial spaces and working class dormitory neighbourhoods closely linked to the central municipality. In this fist crown there are for example the municipalities of Torrent (82 208 inhabitants) and Paterna (70 195 inhabitants). The northwest axis, characterized by the presence of municipalities of smaller demographic size (L'Eliana, 18 235 inhabitants; Lliria, 23 253 inhabitants) but which have experienced an intense process of suburbanization linked to their role as a residential area for the middle and upper classes. And finally a peripheral space of a semi-rural character functionally linked to the metropolitan space, where some cities

stand out that maintain their own industrial dynamics, such as Sagunto (66 140 inhabitants), and, to a lesser extent, Chiva (15 123 inhabitants).

The database comes from the cadastral cartography available online in the *Sede Electrónica del Catastro* (https://www.sedecatastro.gob.es/). The data are the "Urban sub-parcels" that represent the built-up volumes within a plot in 2019. Pre-processing of the data has been required in order to apply the method. Four steps are needed:

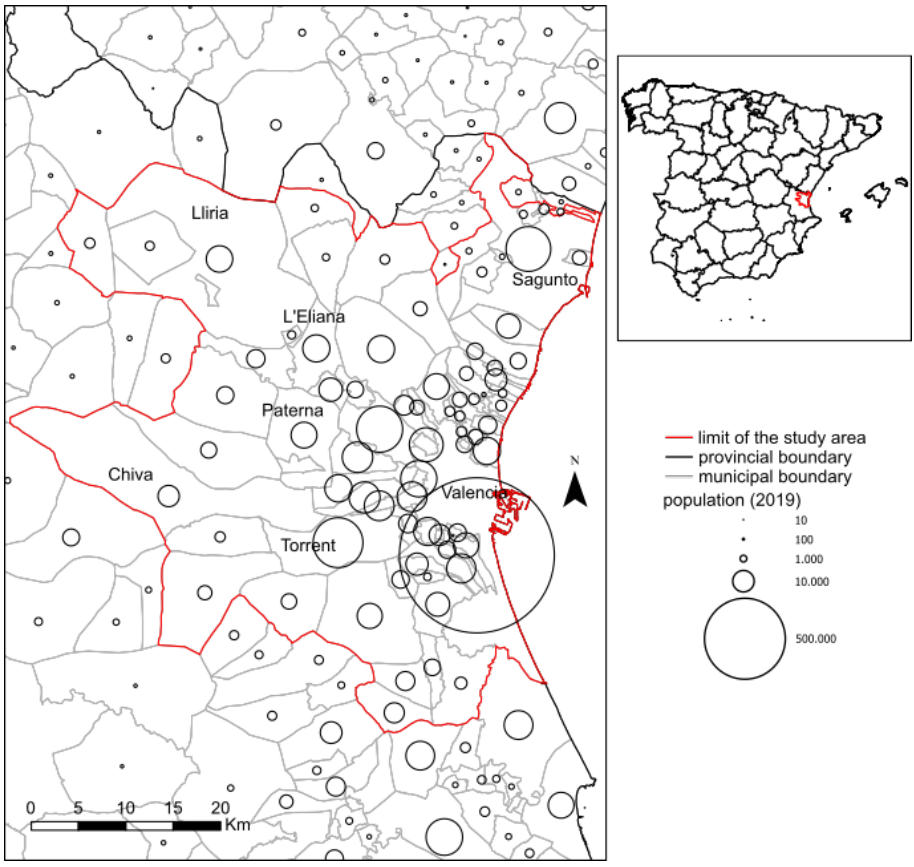


Figure 7.1. Delimitation of the Metropolitan Area of Valencia

1. Selection of the polygons corresponding to the buildings, eliminating unbuilt spaces such as plots and undeveloped land, as well as other urban uses not corresponding to buildings (gardens, roads, port, etc.).
2. Reclassification of each polygon to give it the maximum height (number

of floors)

3. Union of the built-up polygons belonging to the same plot. Each polygon (building) has been assigned the number corresponding to the maximum number of floors above ground level.

4. Union of all layers corresponding to the study area

The final database is composed of 360 008 polygons (buildings) with a height of 1 to 36 floors. The buildings cover an area of 525 km^2.

7.3.2 Methods

To delineate the agglomeration of Valencia, we use *MorphoLim* (Tannier et al., 2011; Tannier and Thomas, 2013). As a reminder (see Chapter 5), the methodology works in several steps:

1. application of Minkowski dilation over the buildings to draw a dilation curve ($x-axis$: dilation distance (in meters), $y-axis$: number of clusters)

2. detection of the distance threshold from the main change in behaviour on the dilation curve estimated with a polynomial regression

3. application of the distance threshold to define the urban border

4. selection of the agglomeration (the city) according to a rank/size distribution of the surface of each cluster

In this chapter, the method has been applied in two phases: first for the entire built urban space. Second, by considering exclusively, and successively, buildings with heights above a specific number of floors. We compute the analyses from 2 floors or more to 11 floors or more. For each of these groups of buildings, morphological agglomerations have been detected and analysed, which has made it possible to determine significant differences in their spatial patterns.

7.4 Results and discussion

7.4.1 Delineation at ground level (2D)

The maximum curvature value is -0.24 which is (in absolute value) smaller than the average values observed for French urban areas (-0.34 see Chapter 6). We see in the previous chapter that a low curvature value (in absolute value) is associated with a higher intensity of urban sprawl and lower contrasts between rural and urban space. Hence, if we compare the curvature value of Valencia with the mean curvature value of the French urban areas, we have that the Spanish city is indeed more diffuse. It should be noted, however, that Belgium shows a much more diffuse pattern (cf. Chapter 5).

The distance threshold detected is 72 meters (Fig.7.2a). This distance threshold is somewhat lower than the majority of Belgian and French cities which are around 100-200 m (Tannier and Thomas, 2013). The application of the distance threshold to urban areas results in the identification of 13 053 clusters of different sizes. Only 63 clusters are larger than 1 km^2 and only 9 are larger than 5 km^2. Each one of them responds to a discontinuity in the agglomeration. As can be seen in Fig.7.3, the resulting territorial pattern is highly fragmented, responding to the morphological heterogeneity of the metropolitan growth process. The rank-size distribution of the clusters shows no primacy (Fig.7.2b). It is therefore difficult to select a particular cluster to delimit the Valencia agglomeration at ground level.

It is possible to identify in the Fig.7.3 a series of clusters that act as small potential agglomerations. The city centre of Valencia does not appear to be a particularly larger cluster than the others. The metropolitan centre is in fact divided into three groups separated by the river axis and a main road. The northern part (Fig.7.4a) is separated from the southern part of Valencia by the river. The new course of the river, a deviation of the previous one towards the south built in 1969 to avoid flooding, constitutes the southern limit of this southern cluster of the centre of Valencia. The Eastern part of the city is separated from the rest by the river and a main road (V-21).

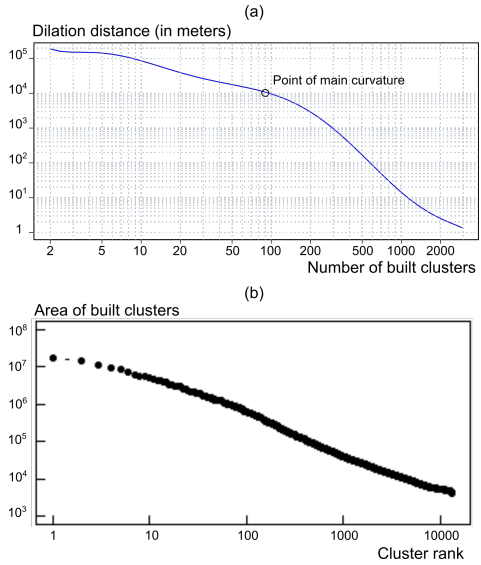


Figure 7.2. Dilation curve and rank-size distribution of the Valencia urban area at ground level

Besides the hydrological and road network, the impossibility to detect a particular cluster as the main cluster is also due to the very diffuse spatial distribution of buildings in the study area. Indeed, important morphological clusters can be detected which do not necessarily coincide with demographically important urban nuclei. These clusters are made up of the sum of a series of dispersed residential spaces belonging to one or more municipalities. In many cases, these clusters are even on the fringe of the main population nucleus of their respective municipalities. This is the case of the second-largest agglomeration in the region, located in the north-western axis of the metropolitan area (municipalities of La Eliana, Pobla de Vallbona, Ribarroja, San Antonio de Benageber and Bétera). In this cluster, the proportion of compact urbanisation is minimal and spread over several municipalities (Fig.7.4b). This pattern is repeated in many other cases. The situation of the cluster located in the municipality of Lliria, consisting of a group of urbanisations physically separated from the main urban centre of the municipality, is particularly significant (Fig.7.4c).

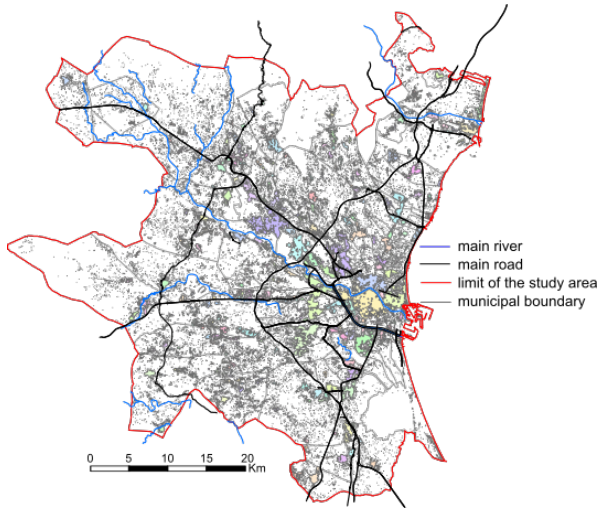


Figure 7.3. Morphological clusters of the urban metropolitan area of Valencia (ground level)



Figure 7.4. Example of main morphological clusters : (a) clusters of the centre of Valencia; (b) north-western axis cluster; (c) clusters of Lliria

7.4.2 Delineation according to the number of floors (3D): the city of the upper districts

The fragmentation of the space (due to the hydrological and road network) and the presence of spatially diffuse buildings in the Valencia metropolitan area make it difficult to delineate its morphological agglomeration. The height of the buildings (largely related to the cost of the land) is one of the morphological differential characteristics of urban spaces compared to rural spaces. It is also a characteristic that allows differentiating functional spaces within the city (business centre, commercial area, etc.). This is why we will apply *MorphoLim* to buildings in Valencia with at least a specific number of floors. We want to see if there are one or more agglomerations of neighbourhoods with multistorey buildings.

According to the land registry, the Metropolitan Area of Valencia counts 360 018 buildings ranging from 1 to 36 floors. There are far more buildings with few floors than buildings with several floors. Buildings with 1 and 2 floors account for 72.4% of the total (35.9% and 36.4%, respectively), while those with more than 5 floors account for only 5.5%. Those with more than 11 floors count for 0.32% and only three buildings exceed 26 floors (see Fig.7.5). For this reason, the morphological delimitation analysis was applied successively to groups of buildings with more than 1 to more than 11 floors, as the number of additional cases was minimal and the analysis did not provide any new information. Results related to the calculation of the maximum curvature of dilatation and the corresponding distance thresholds are shown in Tab.7.1.

As can be seen in Tab.7.1, the value of maximum curvature shows little variation and tends to remain around -0.2 and -0.3 except in the case of buildings with more than 11 floors. In this latter case, the buildings, due to their morphological characteristics, tend to be located in very specific urban spaces, linked to the urban CBD and the coastal area. The boundary between the agglomeration and its surroundings is then much more marked than in the rest of the built-up area. There is no trend in the main curvature, but there is a significant trend for the distance threshold to increase as building height increases. This trend could be related to the increase in the average distance between nearest neighbours as the number of floors increases (see tab 7.1). On average, the tallest buildings are therefore a little further apart than buildings with a few storeys. This increase in the distance between buildings could lead to the "phagocytization" of small urban centres on the outskirts of Valencia (see fig. 7.7(b) and fig. 7.7(b)) . This leads to an increase in the surface area of the morphological agglomeration. As the height of the buildings increases, the number of morphological clusters decreases. We

observe that the higher the buildings, the more they tend to be spatially clustered in a small number of places.

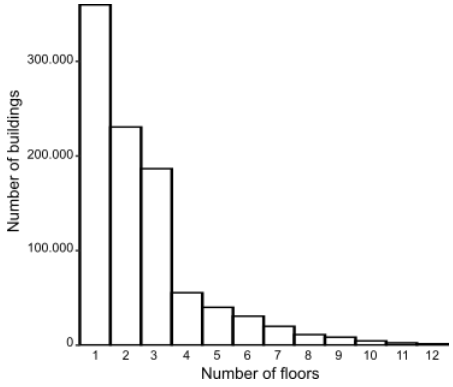


Figure 7.5. Number of buildings by number of floors in Valencia

The rank-size distributions of the different building heights show increasingly strong hierarchies in favour of the main cluster. From the analysis of the results obtained for a number of floors equal to 2 and more, we observe the presence of 7 larger clusters that mainly gather the different parts of the Valencia centre. From 4 floors onwards, it is possible to clearly identify a main morphological agglomeration which corresponds to the conurbation of the city of Valencia. The agglomeration delimited for buildings of 7 floors and more corresponds fairly well to the historical extent of *Horta*, which is the space that has traditionally been used to define the metropolitan area of Valencia, before considering the more recent processes of dispersed urbanisation. We therefore find a more historical delimitation of Valencia.

Therefore, the introduction of the third dimension introduces new relevant information for the delimitation of urban spaces. The analysis of the pattern of the "upper" city shows an identifiable and more marked boundary between the urban agglomeration and the periphery than the one resulting from the analysis of the total floor plan of the built space. We can advance as a hypothesis that the pattern on the surface is a combination of several vertical models, so that the consideration of the third dimension contributes to improve the analysis of the urban morphologies of spaces affected by intense processes of suburbanisation.

7. Delineation of a « diffuse city » at ground level (2D) and according to the number of floors (3D)

Number of floors	Main curvature	Distance threshold (in meters)	Average nearest neighbor (in meters)
=1	-0,247	72	17,4
>1	-0,196	117	16,6
>2	-0,175	264	18,3
>3	-0,336	662	20,6
>4	-0,258	1280	21,3
>5	-0,215	2372	22,8
>6	-0,194	1543	26,7
>7	-0,189	1529	31,8
>8	-0,171	6400	33,4
>9	-0,284	3462	37,4
>10	-0,331	3812	50,4
>11	-0,762	3490	61,5

Table 7.1. Main curvature and distance threshold obtained on buildings with different number of floors

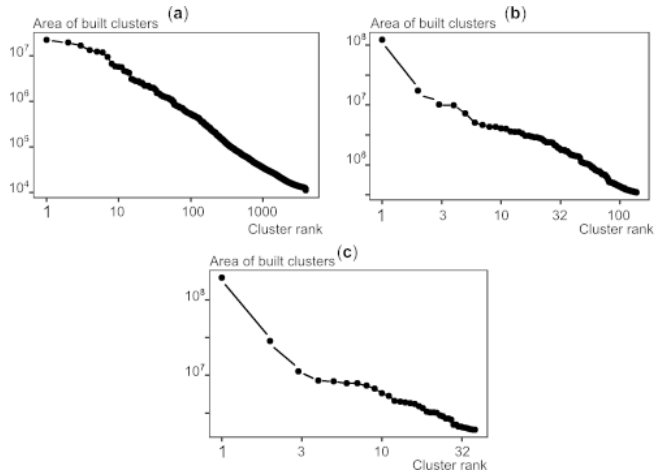


Figure 7.6. Rank-size distribution of morphological clusters of buildings of (a) 2 or more floors; (b) 4 or more floors; and (c) 7 or more floors



Figure 7.7. Morphological agglomeration for buildings of (a) 2 or more floors; (b) 4 or more floors; and (c) 7 or more floors

7.5 Conclusion

According to the results obtained, *MorphoLim* appears useful for analysing metropolitan spaces, such as the Valencian metropolitan area, affected by intense processes of dispersed suburbanisation and fragmentation.

As already developed in the two previous chapters, *MorphoLim* does not need a predetermined distance threshold but deduces it from the data itself. There is therefore no application of thresholds a priori. In the case of the urban area of Valencia, the usual distance threshold applied is 1000 metres as in the node density methodology (Salom and Zornoza, 2015) or in other delimitations based on the detection of conurbation processes (Integrated Urban Area of the Valencia Territorial Strategy). This threshold is much higher than the 71 m. which, according to the *MorphoLim*, is considered more appropriate for the study area. The application of the latter threshold tends to fragment the metropolitan area morphologically and makes it difficult to define a border for the whole area. This result is more in line with the morphological reality of the territory analysed.

The delineation of the metropolitan area of Valencia at a ground level shows a new vision of this space which highlights the importance of morphological clusters made up of dispersed and in some cases supra-municipal urban spaces mostly located on the fringes of the main compact nuclei. This is a particular challenge for territorial planning in the metropolitan area because of their environmental impact, the additional costs they generate in different areas of municipal management and because they are located outside the usual administrative framework of planning.

The introduction of the third dimension, based on the analysis of the height of buildings (number of floors), offers a new perspective in the morphological analysis of urban spaces. The analysis of the fractal model of the 'upper' city shows an identifiable and more marked boundary between the urban agglomeration and the periphery than when we analyse the area at ground level. The association of this building typology with, for example, the cost of land and some metropolitan functions (CBD, emblematic areas, tourist coastline) should make it possible to differentiate the compact urban space from that affected by urbanisation processes and improves the identification and characterisation of metropolitan morphologies.

8

A partition-free spatial clustering that preserves topology: application to built-up density

Urban density is central to urban research and planning and can be defined in numerous ways. Most measurements of urban density, however, are biased by arbitrarily chosen spatial units in the denominator and ignore the relative location of elementary urban objects within those units. We solve these two problems by proposing, in this chapter, a new graph-based density index, which we apply to the case of buildings in Belgium. The density of buildings is defined independently of the definition of any basic spatial unit (BSU) and preserves the built-up topology, i.e. the relative position of buildings.

8.1 Introduction

Density is the ratio of a mass - typically a number of individuals, jobs, or buildings - to a given two-dimensional reference, i.e. an area. In urban and planning practices, different density metrics are used (Churchman, 1999; Longley and Mesev, 2002; Angel et al., 2021) to represent the functioning of a city and, more particularly, to hint at the sustainability or liveability of urban forms (Pauleit and Duhme, 2000; Boyko and Cooper, 2011; Ewing et al., 2018; Rinkinen et al., 2021; Martino et al., 2021). Density metrics, however, are subject to two major issues: they use a reference area which is not necessarily related to research goals, and they ignore the relative locations of spatial elements within that reference area, thus aggregate without making sure of the internal homogeneity of the area. These problems were already stressed within the axiomatic approach to geographical space of Beguin and Thisse (1979) who showed that the elementary area considered to measure a density cannot be separated from the metric of the relative location of places, namely the topology.

We draw from this inherent property of geographic space and from more recent work by Caruso et al. (2017) to propose a method that computes a topology-based density index. The idea is to compute the index without using a reference surface but a graph that connects spatial objects (points) via edges. By weighting the edges of the graph with the distance between points, we can obtain a graph that preserves the relative position of the points with respect to each other. One can then cut the graph in order to obtain groups of points whose relative distances are homogeneous, thus having a homogeneous base (topology) for computing an aggregate index such as density. We suggest a novel spatial descending hierarchical clustering (SDHC) method to cut the graph, where the most locally dissimilar edges are removed iteratively. Rather than an exogeneous cutoff, we use the Cooks statistics on Moran scatterplot's regressions, thus directly using local topological information at each stage.

We focus here on the case where points are buildings. Buildings, together with plots and streets, are the elementary constituents of an urban space (Moudon, 1997). We illustrate our method by proposing a buildings' density index, which is a key index in urban planning. Hence, we move the measurement of the buildings' density from an area-based problem to a graph-based problem. Our density metric preserves the relative location, i.e. the topology, of the buildings.

We can see this research as a contribution to a strand of the urban literature where density is complemented by morphological indices (e.g. Galster et al., 2001; Berghauser Pont and Haupt, 2005; Sémécurbe et al., 2019; Fleischmann, 2019, 2021). Contrary to existing work in this domain, we emancipate from the definition of a reference spatial unit. We can also see our work as extending another line of research where topological graphs are used to capture the local spatial organisation of buildings (e.g. Caruso et al., 2017; Wu et al., 2018, for recent examples). Compared to this second strand, we move a step forward by proposing to use Moran's scatterplots not only for describing but also for cutting the graph, and by adding a density measure.

In the next section, we position our work within broader urban morphology research and with particular regards to recent methods applied to graphs of buildings. We then present data inputs and the different steps of our methodology in Section 8.3, then the results of an application to all buildings located in Belgium (Section 8.4). We discuss our findings in Section 8.5 and conclude in Section 8.6.

8.2 Methodological advances

The density of buildings is a major indicator for urban planning, but can only properly be used when two conditions are met: (i) density is complemented with indicators describing the relative spatial organisation of buildings and (ii) measurement biases due to the use of basic spatial units (BSU) are overcome. Regarding the first of the two conditions above, progress has been made since Galster et al. (2001) especially towards deriving composite multi-factorial indicators. The use of multiple attributes can depict how urban densities are experienced by inhabitants and users (Teller, 2021). Caruso et al. (2017) reviewed some of these multi-indicator contributions, mostly directed to capturing sprawl. They usually consider urban land pixels, not buildings. Teller (e.g. 2021); Godoy-Shimizu et al. (e.g. 2021); Berghauser Pont and Haupt (e.g. 2005) are examples of multi-factorial methods applied to building' distribution in relation to density. However, these indicators are defined from discret zone (BSU) and therefore do not meet the second condition.

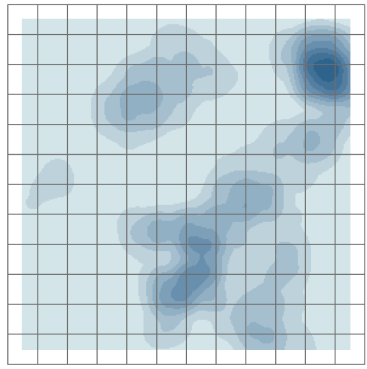
Regarding the second condition, i.e. avoiding the bias of a basic spatial unit (BSU), two strategies have been adopted so far: (i) spatial or statistical smoothing across the BSUs and (ii) clustering individual data with standard (k-means, etc.) or more complex classification methods (artificial neuronal networks methods, etc.). Bellefon et al. (2019) is a recent example of smoothing. The authors use a kernel function applied on a very fine grid resolution to avoid losing information about the internal spatial organisation

while aggregating. In essence however, kernels still depend on an *a priori* defined surface (bandwidth), even if it is often optimised for each case study (similarly to Geographically Weighted Regression approaches with optimum bandwidth (Brunsdon et al., 1996)). Although exceptions exist such as Araldi and Fusco (2019) (among others) in many cases, smoothing methods are applied uniformly across space and may not fit the local composition everywhere. Let us take Fig.8.1a as an example of a spatial pattern of buildings. If we compute density using a smoothing grid (kernel) (Fig.8.1b), we can see that the topology is lost and that the grid prevents from detecting groups of buildings with a similar topology. Referring to (ii), a recent example of density-based spatial clustering is the A-DBSCAN (*Approximate - Density-Based Spatial Clustering of Applications with Noise*) by Arribas-Bel et al. (2019), building on earlier work by Ester et al. (1996). Buildings are grouped according to a density criterion in order to draw city boundaries. Similar to other density-based methods, A-DBSCAN is not parameter free (minimal number of buildings and maximal distance between them in a cluster) and results can thus greatly vary from one user to another. Furthermore, these two criteria are applied uniformly over the study area, which therefore prevents the detection of locally specific patterns. If we now apply A-DBSCAN (Fig.8.1c) to our example, we see that buildings with different visual topologies are clustered into one large group (blue). Hence, with this method we can see that the topology is partly lost.

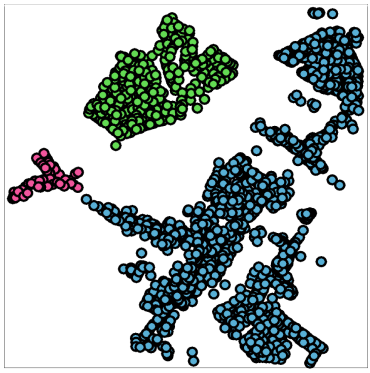
To avoid the vanishing of topologies due to spatial aggregation, Zhang and Kukadia (2005) suggest creating BSUs that make sense with regards to the initial spatial organisation of the disaggregated data. This goal is pursued by density-based or graph-based clustering methods (Wu et al., 2018; Deng et al., 2011). This is done by Fleischmann et al. (2020) or by Schirmer and Axhausen (2015) who perform clustering, local spatial statistics, and spatial smoothing within the topological constraints of building-based tessellation adjacencies or street network topology. In both publications, BSUs are adapted in size and shape to the urban intensity. We pursue the same objective in this paper by presenting a method that is able to provide BSUs that make spatial sense in terms of distribution of buildings (i.e. topology) using a graph-based clustering method.



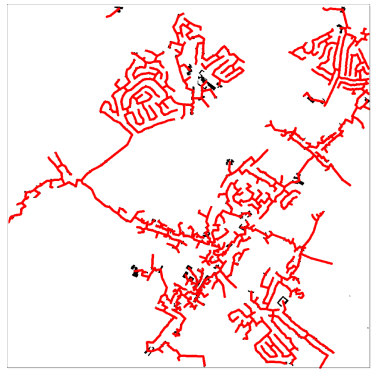
(a) Built-up footprints



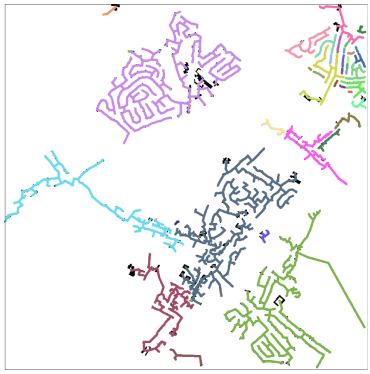
(b) Smoothing with a kernel function (bandwith = 250 m.) based on a grid



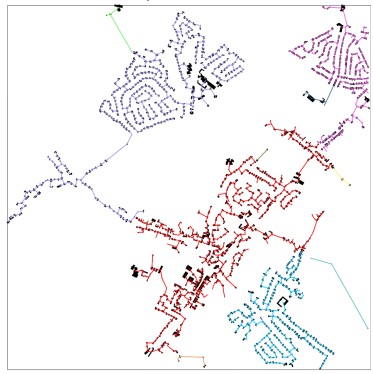
(c) Clustering : A-DBSCAN (distance = 200 m. ; min nodes = 30)



(d) Clustering : fixed distance threshold (distance $\leq$ 200 m.)



(e) Clustering : based on Gelstat theory ($p_1 = 2$, $p_2 = 2$, $p_3 = 3$)



(f) Clustering : SKATER (number of cluster = 10)

Figure 8.1. Different methods to calculate the density of buildings or to perform a spatial clustering of a built-up pattern

In graph-based methods, the nodes are typically the buildings (primal graph) and the edges the inter-building segments computed from their centroid or from their edge. The advantage is to conserve information about the absolute location of buildings as well as their relative location (Anders et al., 2001; Assunção et al., 2006; Caruso et al., 2017; Wu et al., 2018). Most graph-based methods start with a Minimum Spanning Tree (MST), which is easily partitioned into subgraphs, i.e. clusters. Each time one edge is removed, two distinctive subgraphs are created. If n edges are removed from the initial MST, $n + 1$ clusters of buildings appear. Different strategies are available to determine which edges should be removed to perform the spatial clustering. Caruso et al. (2017) remove all edges larger than an *a priori* threshold fixed at 200 metres. But this unique threshold cannot fit the local spatial organisation of buildings everywhere (Fig.8.1d). Zahn (1971); Yu et al. (2014), based on Gelstat theory, remove the edges that are the most different in the set of their contiguous edges according to three parameters determined *a priori* (p_1 the number of contiguous neighbours; p_2 a ratio between the length of an edge and the average length of its neighbours; p_3 a ratio based on the difference between the length of an edge and the average length of its neighbours to their standard deviation). Results will depend upon these thresholds. While Fig.8.1e shows interesting results overall, it can be seen that the pattern of buildings to the north-east of the area is very poorly captured by the method. Assunção et al. (2006) use a *Spatial 'K'luster Analysis by Tree Edge Removal* (SKATER) where they iteratively suppress the edge which, once suppressed, minimises the sum of the intra-cluster variances. In this case, the number of final clusters has to be fixed *a priori*.

Following the axiomatic of Beguin and Thisse (1979), we know that the denominator of our density metric must preserve the relative positions of buildings everywhere in space. We thus follow graph-based approaches where inter-building distances are used to build and then to prune the MST. We propose a new segmenting (clustering) algorithm that does not require an exogeneous threshold to be applied uniformly across the area, nor a number of clusters to be defined beforehand. Our strategy, inspired by the LISA approach of Caruso et al. (2017), is to compute a Moran scatterplot of inter-building distances for each graph (subgraph) and remove the main outlier for segmenting, rather than applying a distance threshold. As a result, the segmentation can be different across the area and catch distinct local topologies. Density is then computed, based on these topologically homogeneous clusters. The reader can already see that other morphometric indices than density could be calculated on these clusters.

8.3 Materials and Methods

8.3.1 Data input and study area

Our method is applied to all buildings located in Belgium. Data are provided by the Land Registry Administration of Belgium (© 2018 Administration Générale de la Documentation Patrimoniale). All buildings are used regardless of their function: each house (detached or semi-detached), office building, shop, garage, church or factory is kept in the database. In order to avoid the noise generated by very small buildings such as garden sheds, all built polygons smaller than $12m^2$ were removed from the database as was done by Montero et al. (2021). The database includes 5 726 804 buildings, which further leads us to chunk the data for computation (see Appendix C.1). Fig.8.2 shows the study area and a zoomed map of the buildings' footprint.

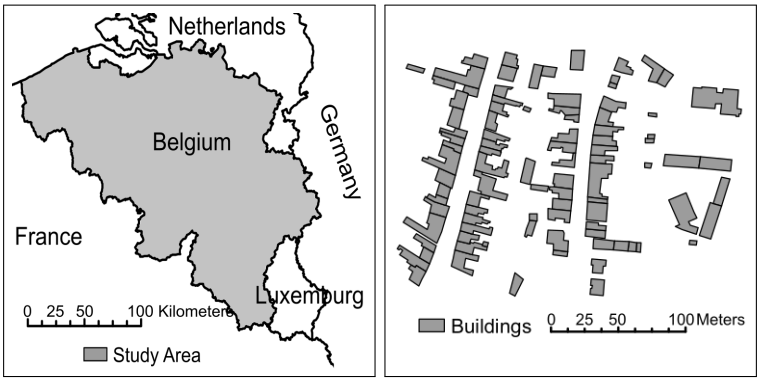


Figure 8.2. Study area (left) and an example of the footprint of buildings (right)

8.3.2 Methods

Our objective is to create a topology-based density index that preserves the local spatial organisation of the buildings. Hence, after pre-processing, our method comprises two main steps : a clustering method (step I) leading to topology consistent groups and a density computation (step II). The process and the outcome of each step are summarised in Fig.8.3 for a toy example (Fig.8.3a).

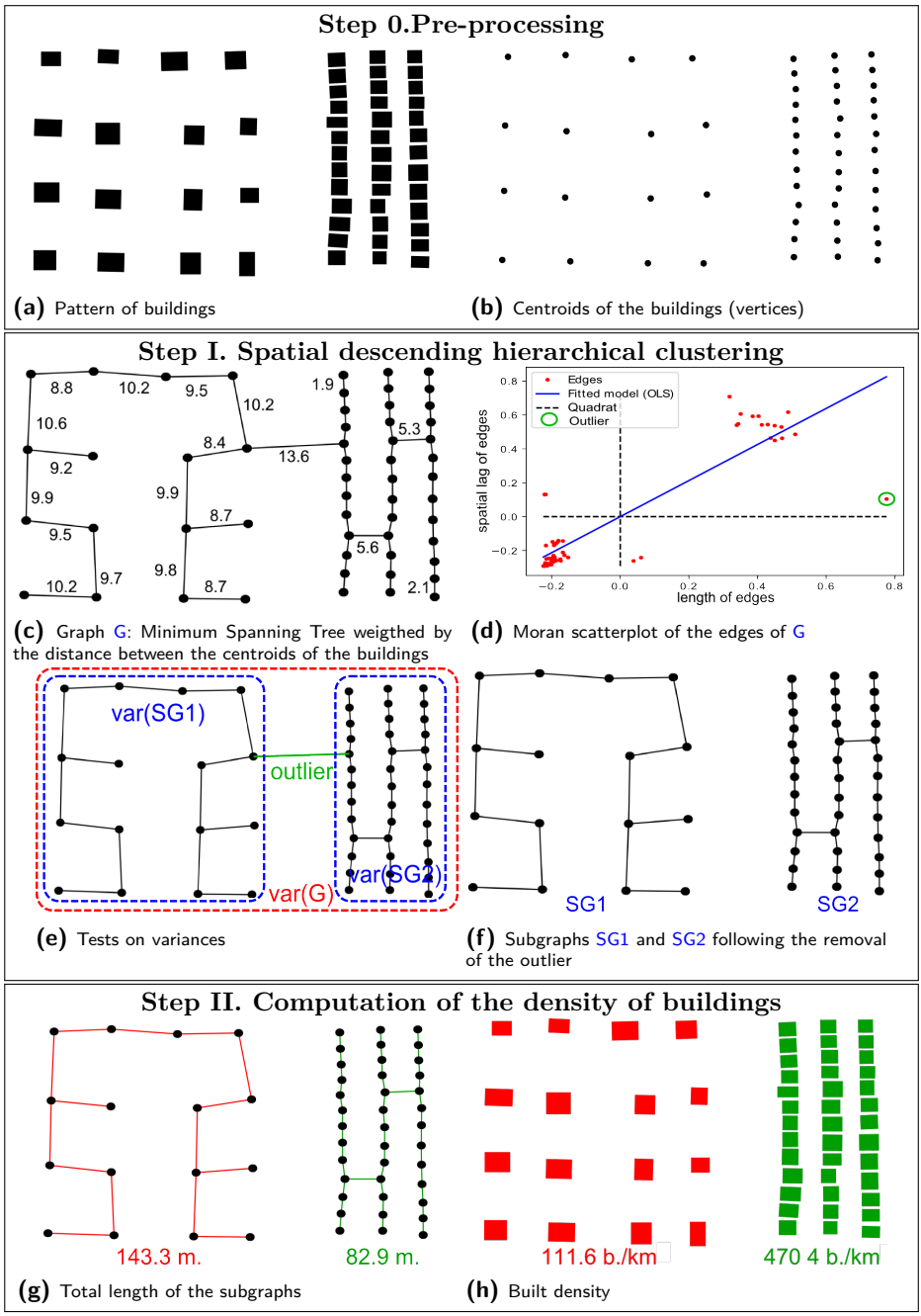


Figure 8.3. Application of the method on a simple spatial toy structure of buildings

Step 0 - Pre-processing.

The input data consists of polygons. Although the size and shape of the polygons can be heterogeneous across space and further impact the distribution of inter-building distances, we ignore those shapes by retrieving centroids (step 0, see Fig.8.3b). The distance between buildings is here the distance between the centroids of the buildings. We could have followed Yu et al. (2014), who measure the actual distance between buildings but this would not be appropriate in our case. Indeed, we want to measure the density of buildings per km of graph. By using distance between centroids, the user of the metric can then say "When I walk 1 km along the graph, I encounter x buildings". With the actual distance between buildings, this interpretation of the metric would no longer hold. Rather, the user should imagine teleporting him/her self from one end of the buildings to the other as he/she travels along the graph. Moreover, the real distance between two buildings may be zero (e.g. terraced house) which leads to an indeterminacy (denominator is null).

Step I - Spatial descending hierarchical clustering (SDHC)

This step starts with a minimum spanning tree (MST) where the nodes are the centroids of the buildings, and the edges are the inter-building segments. Euclidean distances are used as weights (Fig.8.3c) while computing the MST graph, which we denote by G. A descending hierarchical classification (SDHC) is then applied on G to iteratively define subgraphs SG by removing an edge out of each parent graph. The SDHC process is explained below and flow-charted on Fig.8.4. It is carried out on graphs of at least 30 vertices in order to avoid statistical problems due to small numbers.

A Moran scatterplot (Anselin, 1995) and a maximum Cook's distance are used to identify the edge that should be removed at each iteration of the SDHC. The Moran scatterplot (Fig.8.3d) shows how the edges are spatially associated locally. One point represents one edge of the MST. On the x-axis is the length of an edge, i.e. the distance between two connected buildings, and on the y-axis, its spatial lag, i.e. the weighted average distance of its contiguous edges. We voluntarily restrict the computation of the spatial lag of an edge to its contiguous neighbours of order 1 because we want to detect break between direct neighbours. A topological depth greater than 1 would lead to a smoothing of the spatial lag of each edge. Details about the spatial lag computation can be found in Appendix C.2.

A linear model (OLS) is then estimated on the scatterplot, the slope of which indicates the global spatial autocorrelation level (Moran's I, see Anselin (1995)). In most cases, we expect a positive slope, meaning that a long (short) separation between two buildings is found in neighbourhoods where distances between buildings are long (short) on average, i.e. in topologically homogeneous cases. The slope will be insignificant in cases where the relative distance between buildings and thus the topology are more heterogeneous. As pointed out by Anselin (1996), *"points in the scatterplot that are extreme with respect to the central tendency reflected by the regression slope may be outliers in the sense that they do not follow the same process of spatial dependence as the bulk of the other observations."*. We build on this property and use the maximum Cook's distance to identify the most extreme point (outlier) of each graph G, which is actually the edge to be removed to obtain two subgraphs ($SG1$ & $SG2$) (Fig.8.3f) featuring two topologically distinct clusters of buildings.

In order to determine whether the removal of the outlier leads to the creation of more homogeneous subgraphs, tests of variance (Brown and Forsythe, 1974) between the parent graph (G) and each of the child subgraphs ($SG1$ & $SG2$) are performed (Fig.8.3e). These tests measure whether the variance of the length of the edges in at least one of the two subgraphs is statistically different from that of the parent graph. If the null hypothesis (equality of variances) is rejected, the edge is removed, and the algorithm is re-run separately on each of the subgraphs. If the null hypothesis is not rejected, the edge is not removed, and the algorithm stops. Technically, in order to perform a relevant variance test, one needs first to make sure the parent graph is not very large (thus bearing a lot of heterogeneity) and that each child graph is at least made of more than one point. Hence the following three conditions are used to determine whether the observed outlier is removed or not (Fig.8.4):

- the total length of the MST is higher than 10 000 metres (see Appendix C.3 for details) (1)
- one of the subgraphs is made of a single vertice (2)
- the variance of the length of the edges in at least one of the two subgraphs is statistically different from that of the graph before edge suppression (3)

Step I is completed when no more edge can be removed (no significant outlier) from any of the subgraphs. Each subgraph is thus a topologically homogeneous cluster of buildings.

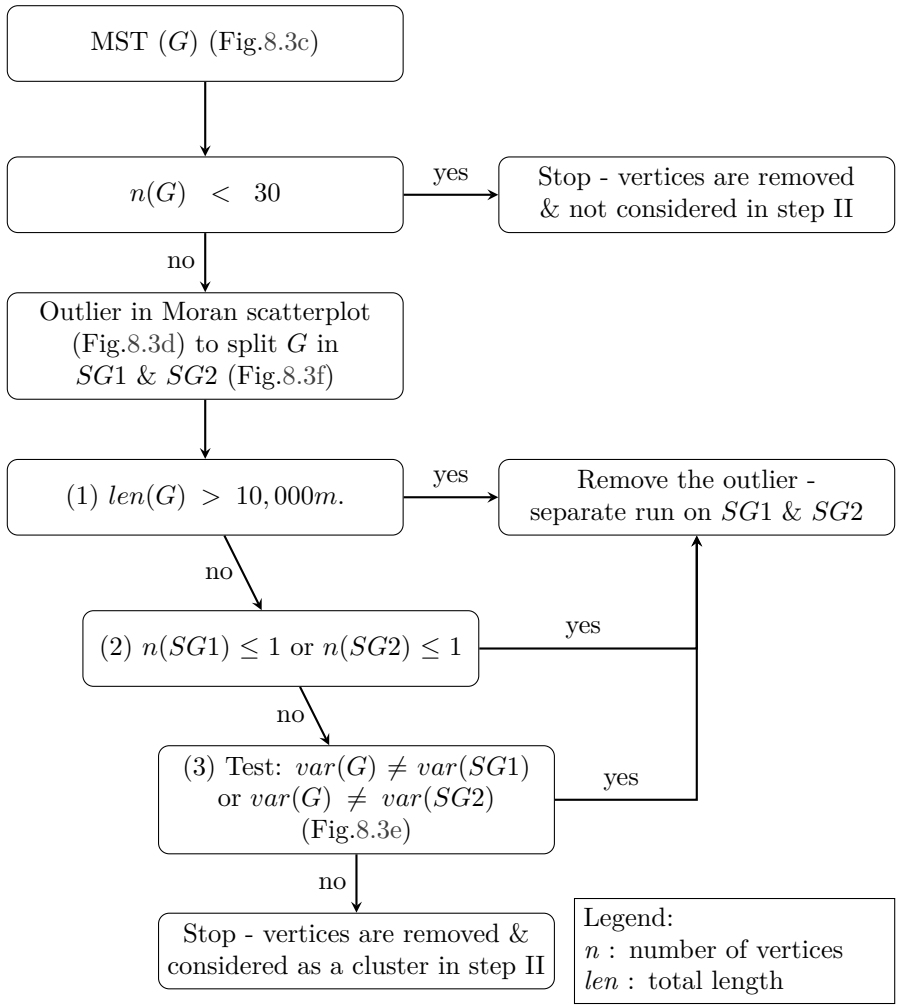


Figure 8.4. Step I: Flow chart

Step II - The topology-based density index

Let i be a cluster of buildings resulting from the SDHC above. D_i^*, the topology-based density of the cluster i, is then defined as the ratio of N_i, the number of buildings and L_i (Fig.8.3g) the total length of the edges of the MST of i:

$$D^*{}_i = \frac{N_i}{L_i} \tag{8.1}$$

(Fig.8.3g)

D_i^* is then a linear density (buildings per linear distance). It would be possible to use the square of L_i to obtain a surface version of the density. However, this transformation brings new biases (due to the different lengths of L_i) without bringing any new information.

D_i^* can then be mapped onto each building of the corresponding graph (see Fig.8.3h). Unlike common practice, the denominator is no longer an *a priori* chosen surface but the length of the shortest line connecting all buildings in a cluster. Contrary to grid based approaches, it is not constant over space in order to match the local spatial structure.

8.4 Results

8.4.1 The Minimum Spanning Tree

The Minimum Spanning Tree computed on all buildings located in Belgium describes the global topological structure of the built-up Belgian reality. As expected by the level of urbanisation of the country, the lengths between buildings in the MST are shorter than 50 m. for a very large majority (95%) (see Tab.8.1). The histogram (Fig.8.5) is right-skewed and has a bimodal distribution with a very strong peak around 6 metres, and a secondary peak around 20 metres. The left part of the histogram shows short edges which correspond to attached or very close buildings of small size. Note that the very small peak around 3 metres corresponds mostly to sets of contiguous building extensions. In the database, each extension is a small polygon (although larger than 12m²). A set of small extensions joined together then leads to the creation of small edges in the MST. The right part corresponds to edges between 15 and 50 metres characterizing more detached buildings. Edges longer than 50 metres (not illustrated in the histogram) are typical of more isolated buildings. The presence of a large peak at short distances combined with a second peak at medium distances and the absence of a peak at longer distances fully reflects Belgian urbanisation. This urbanisation is indeed characterised by numerous centres (towns, villages) connected with a strong level of suburbanisation and sprawl (Vanneste et al., 2008; Vandermotten et al., 2008; Vanderstraeten and Van Hecke, 2019). One would expect a much smaller second peak and a rural peak (long distance) for a less continuous urbanisation, as is the case for example in the Netherlands.

Mode	q0.25	Median	Mean	q0.75	q0.95	Max
6.00	8.20	15.24	20.85	23.54	49.29	4378.99

Table 8.1. Descriptive statistics: length of the edges of the Minimum Spanning Tree

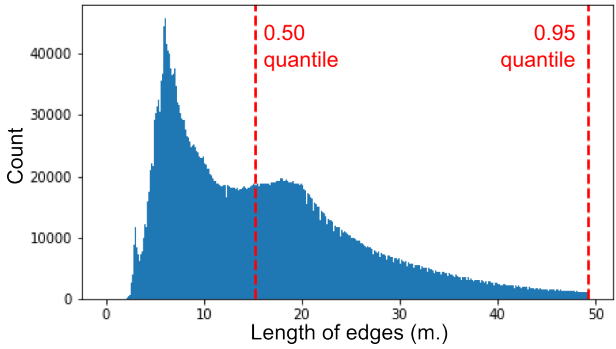
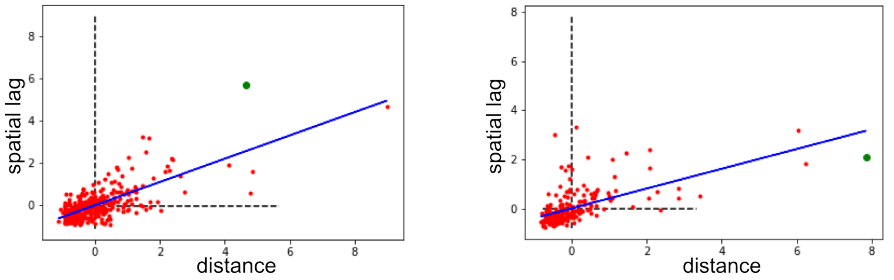


Figure 8.5. Distribution of the 95% smallest edges of the Minimum Spanning Tree

8.4.2 Moran scatterplots and their outliers

Out of the 260 359 Moran scatterplot regressions performed during the whole process, 93% show a positive and significant slope (Moran's I). This demonstrates a high degree of homogeneity in the spatial distribution of buildings within the Belgian landscape. Indeed, at each iteration of the method and therefore at all scales, there would be very few abrupt discontinuities in the spatial distribution of the buildings. This confirms the observations made earlier. Furthermore, the high level of significance of the OLS confirms the relevance of using Moran scatterplot to identify outliers.

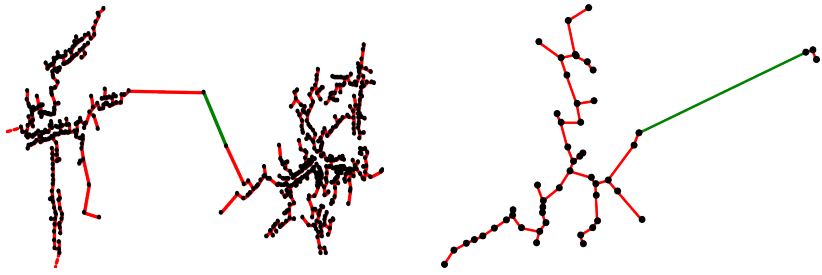
We observe two types of outliers within the scatterplots: first, those corresponding to an edge surrounded by edges longer than expected by the global spatial autocorrelation (Fig.8.6a) and second, those corresponding to an edge surrounded by edges much smaller than expected (Fig.8.6b). An outlier separates two distinct topological forms (Fig.8.7a) but in some cases it can simply isolate some remote buildings from the rest (Fig.8.7b). An outlier therefore separates settlements, towns, city districts, villages, etc. or separates isolated and heterogeneous housing from a homogeneous structure of buildings (a farm on the outskirts of a village, a church in a city centre, etc.).



(a) Outlier is an edge surrounded by edges longer than expected by the OLS

(b) Outlier is an edge surrounded by edges smaller than expected by the OLS

Figure 8.6. Two examples of typical Moran scatterplot (outlier in green)



(a) Outlier separates two distinct morphological spaces

(b) Outlier isolate some remote vertices from the rest

Figure 8.7. Two examples of graphs and related sub-graphs according to the outlier (in green)

75% of the removed edges have a length between 30 and 80 metres while the median length takes the value of 44 metres (Tab.8.2). The large observed range of removed edges shows that the iterative process implemented in the method allows the identification of clusters of different patterns at different scales and for different realities. Indeed, from the first removed edge up to the last one, the iterative process progressively splits the initial graph (all of Belgium) into a series of smaller graphs that outline nested clusters, each with its specific characteristics. The use of the Moran scatterplot allows the selection of the edge to be deleted taking into account these specificities. The method identifies the removed edge at each step. It is therefore possible to go back up the clustering tree to observe these different clusters at different scales. This would not have been possible when using a method based on an *a priori* defined threshold (Zahn, 1971; Yu et al., 2014; Caruso et al., 2017). Moreover, the median value (43 m.) shows that the discontinuity in buildings is in the majority of cases much lower than the one generally used by those studies (between 100 and 200 m.).

Min.	q0.25	Median	Mean	q0.75	Max.
0.71	30.66	43.91	78.09	80.95	4378.99

Table 8.2. Descriptive statistics: length (in meters) of the removed edges

8.4.3 Clusters of buildings

At the end of step I, the method discriminates 26,462 subgraphs (see Fig.C.6 in Appendix C.4). Over 95% of the subgraphs have a coefficient of variation of the length of the edges smaller than 1, which means that within a subgraph, the lengths of the edges are homogeneous (Fig.8.8). Each subgraph can thus be considered as a topologically homogeneous cluster of buildings.

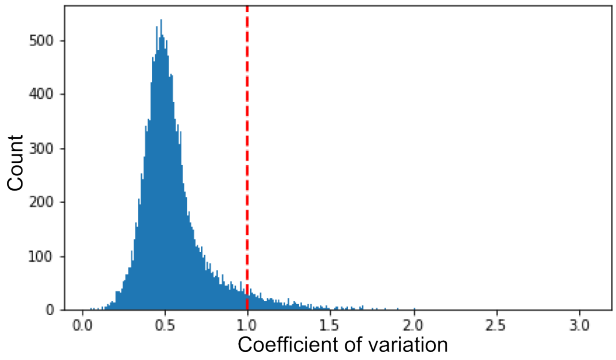


Figure 8.8. Histogram of the coefficients of variation of the edge lengths for all 26,462 subgraphs

Within each cluster, the variance of the inter-building distances is small, which results in the detection of built-up footprints characterised by a regular pattern of buildings (homogeneous topology). Let us now consider a first example illustrated in Fig.8.9. It includes two regular neighbourhoods (A and B). A is a compact village with a radial morphological structure; B is made up of a regular alignment of buildings that forms a linear ribbon development. A second example is reported in Fig.8.10 already used in Section 8.2, south of Brussels, composed of eight homogeneous neighbourhoods (A:H). Each neighbourhood corresponds to a particular pattern of buildings, with a historical centre around the church (A), classical planned housing estates (B:F), and two more linear developments (G:H). Isolated buildings or heterogeneous groups of less than 30 buildings are left out (mainly isolated farms typical of the area).

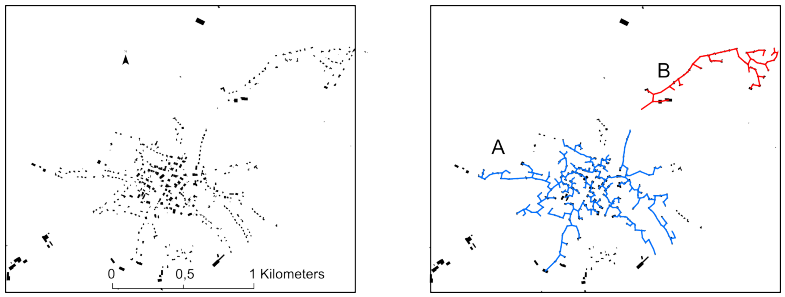


Figure 8.9. Example 1: Built-up footprints (left) and detected clusters (right) of a Belgian village (Ochamps)

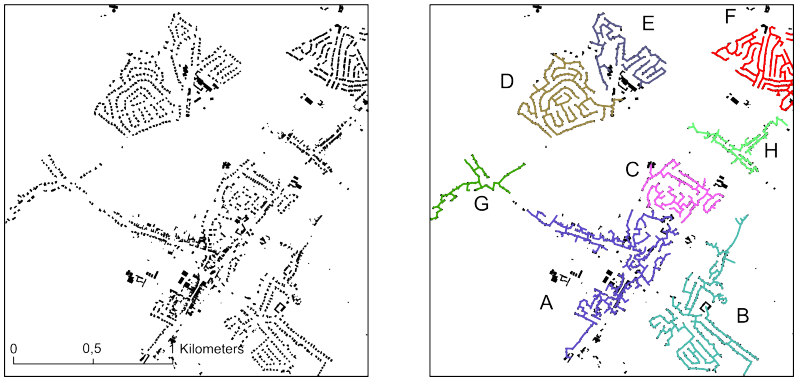


Figure 8.10. Example 2: Built-up footprints (left) and detected clusters (right) of a suburban settlement (Ophain-Bois-Seigneur-Isaac)

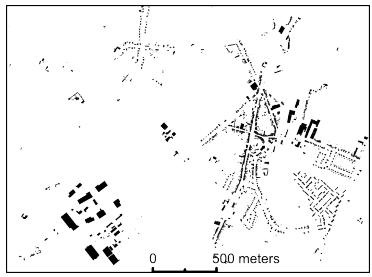
Our 26 462 clusters can now be considered to be topologically relevant Basic Spatial Units (BSU). Since each BSU are internally homogeneous in terms of distance between centroids, we can confidently use those units to compute index such as density, which characterise the spatial distribution of buildings centroids within each cluster.

8.4.4 Density of buildings

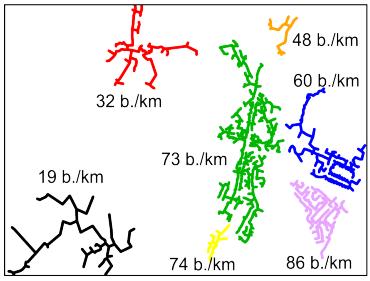
The topology-based density is now computed for each cluster. Each group of buildings with a homogeneous topology, has a specific density value. To explain this specificity of our method, we have compared our results with those obtained by a simple grid-based density smoothed by a kernel function with two examples (Fig.8.11 and Fig.8.12).

With our method, only one density value is computed by cluster when several values are needed with a grid. For example, in the case of Seneffe (Fig.8.11), we identify seven clusters, with seven density values ranging from 19 to 86 buildings per km (Fig.8.11b). The smoothed grid approach covers the area and compute density ranging from 0 to 31 buildings per hectare (Fig.8.11c). While our method detects seven homogeneous buildings patterns with precise contours, the grid method suggests two or three main centres surrounded by less dense periphery located in the east and some shadows in the west. Similarly in the case of Genappe (Fig.8.12), our method detects three distinct homogeneous buildings patterns (53, 58 and 100 b./km) while the grid method delivers different density values ranging from 0 to 57 b./ha) and showing a large centre in the west surrounded by a periphery that develops in a ribbon towards the east.

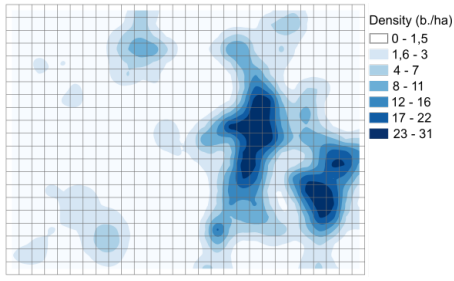
In each example, the different density values, associated with the different clusters, allow the identification of particular urban structures. In the case of Seneffe (Fig.8.11b) the two clusters with a density of 73 b./km and 74 b./km include the buildings of the centre, consisting mainly of semi-detached buildings along a main axis from North to South. In the periphery of the centre, the cluster with a density of 60 b./km is formed of detached buildings while the cluster of 86 b./km includes much more attached buildings (public housing). The clusters around 30-50 b./km are associated with housing estates well separated from the centre with exclusively detached buildings. Large inter-building distances characterise the cluster with the lowest density (19 b./km) (industrial zone). In the case of Genappe (Fig.8.12b), the cluster with a density of 100 b./km includes the buildings of the centre. In an extension of the centre, two clusters are identified with a density about 50-60 b./km consisting of well separated buildings. One of these clusters is a ribbons extension along a road from the centre (53 b./km) and the other is a more widely spread cluster assimilated to a district in the periphery of the centre (58 b./km).



(a) Built-up footprints

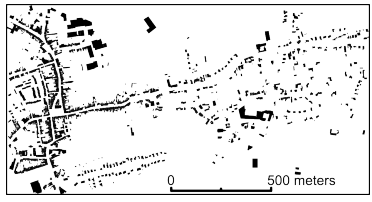


(b) topology-based densities

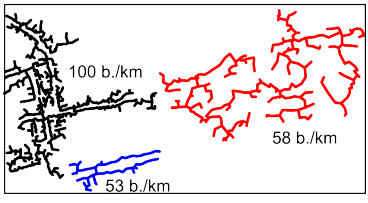


(c) grid-based density (resolution : 100m x 100m ; kernel function (bandwith = 250 m.))

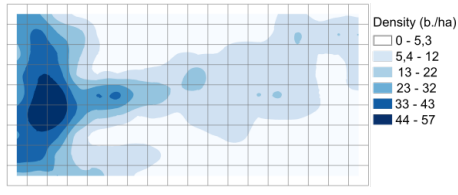
Figure 8.11. The case of Seneffe



(a) Built-up footprints



(b) topology-based densities



(c) grid-based density (resolution : 100m x 100m ; kernel function (bandwith = 250 m.))

Figure 8.12. The case of Genappe

Given the definition of the topology-based density (the number of buildings divided by the length of the MST), there is a direct relation between the density value and the average length within a cluster ($\overline{l_i}$). In fact, inverting the equation 8.1,

$$D_i^* = \left(\frac{L_i}{N_i}\right)^{-1} \tag{8.2}$$

and because in a MST the number of edges is always equal to the number of points (N_i) minus 1,

$$\overline{l_i} = \frac{L_i}{N_i - 1} \approx \frac{L_i}{N_i} \tag{8.3}$$

Then D_i^* becomes:

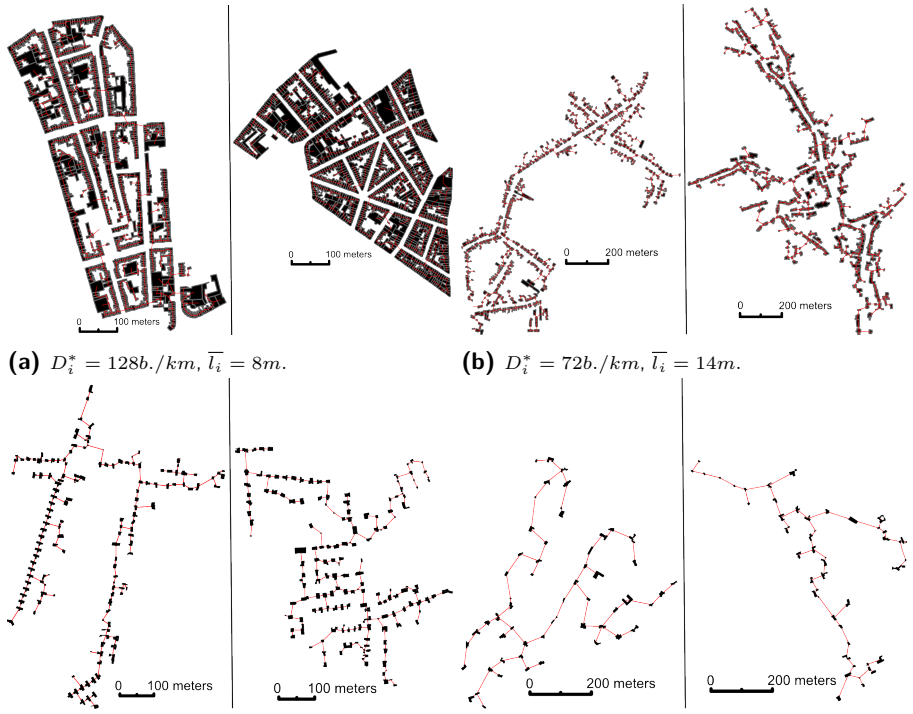
$$D_i^* = \overline{l_i}^{\beta} \tag{8.4}$$

with $\beta \approx -1$ when N_i is large ($N_i \approx (N_i - 1)$). For all 26 462 clusters obtained in Belgium, the value of β can be estimated (OLS after logging both sides of equation 8.4). We obtain a value of β equal to -0.996. This is indeed very close to 1 but shows a slight under(over) estimation for graphs of longer (smaller) average length. In the words of Beguin and Thisse (1979), we show that density (D_i^*) (of buildings in this case), cannot be separated from the metric of the relative location of places ($\overline{l_i}$) and that this relationship follows a simple power law.

According to equation 8.4, the topology-based density only depends on the inter-building distances unlike a surface-based approach where the density can vary according to the area of the BSU without considering the distance between buildings. This might sound like a trivial result, but the use of surface-based densities cannot differentiate between two BSUs where the same number of buildings are located but where once is concentrated and once is dispersed. While others researchers would add additional metrics to capture this (e.g. Galster et al., 2001; Berghauser Pont and Haupt, 2005), our density measure suffices.

Practically, if we know the value of the topology-based density, we can work out the relative spatial organisation of the buildings. Fig.8.13 illustrates four such cases. (a) A built density value higher than 100 b./km is computed on adjoining buildings or very close to each other (mean distances of less than 10 metres) as is the case in city centres (Fig.8.13a). (b) A value between 50 b./km and 100 b./km is related to a topology of buildings relatively close to each other (mean distances between 10 m. and 20 m.) as may be the case, for example,

in the periphery of cities or in smaller centres (Fig.8.13b). (c) Relatively well-separated buildings such as in a peri-urban housing estate (mean distances between 20 m. and 50 m.) have densities in a range between 20 b./km and 50 b./km (Fig.8.13c). Last but not least, (d) densities lower than 20 b./km reflect clusters of buildings with average distances greater than 50 m. (Fig.8.13d).



(a) $D_i^* = 128b./km$, $\overline{l_i} = 8m.$ **(b)** $D_i^* = 72b./km$, $\overline{l_i} = 14m.$

(c) $D_i^* = 41b./km$, $\overline{l_i} = 24m.$ **(d)** $D_i^* = 18b./km$, $\overline{l_i} = 55m.$

Figure 8.13. Typical cases of topology-based density according to inter-building distances

At the scale of the entire country, the newly computed topology-based density (see Fig.C.7 in Appendix C.5) shows a spatial structure that expresses urbanisation in Belgium (see Section 8.4.1).

8.5 Discussion

8.5.1 Thematic contribution

We have proposed a Spatial Descending Hierarchical Clustering method that delineates clusters of buildings with homogeneous inter-building distances. Based on these clusters, we compute a topology-based density index where the denominator preserve the relative positions of buildings. We show (Section 8.4.4) that the index eventually only depends on the average distance between buildings in each cluster. It is a strong advantage compared to standard surface-based densities where density depends on the delineation and definition of a reference area (BSU).

The numerator considered here is simply the number of buildings. Depending on the final objective of the measure, other numerators could equally be used such as the surface of buildings, their volume or height (see e.g. Yu et al., 2010). In our case, we could imagine using the surface area of the buildings in each cluster or the total surface area of their floors as the numerator. These indices would remain topology-based as the denominator does not change. The diversity of indices is therefore a function of the diversity of possible numerators. As shown by Wu et al. (2018), a large number of characteristics (distance, orientation, height, size, etc.) of buildings can easily be integrated into a graph. It is up to the planners to develop and use them according to their needs. In the same way, other parameters can be used as a basis for the SDHC. The association of a Cooks distance with a Moran Scatterplot can be used to distinguish the most different object in a cluster. Rather than having clusters with identical patterns in terms of distance between buildings, some could, for example, look at groups with similar building heights.

The only data input used here are the buildings. This can appear counterproductive since streets, squares, parks, and gardens are traditionally identified as important places in built-up realities (Gehl, 1987). However, it is possible to link a large number of urban issues to the structure/proximity of buildings (energy consumption (Rinkinen et al., 2021), mental health (Sullivan and Chang, 2011), population estimates (Tomás et al., 2016), etc.. This is why we believe that focusing on the density of buildings can be relevant for urban space issues.

We are aware that the use of our graph-based index may appear difficult for urban policy makers as they are often used to work on a externally determined surface basis. However, we believe that it is sometimes necessary to change the approach because of biases and errors in measurement and

interpretation induced by these surfaces. Moreover, measuring the density along a topological network (as proposed here) has also two practical advantages compared to a more classical surface approach. First, it is a more operational way to study the relationships/interactions between buildings and linear infrastructures. Linear infrastructures (electricity, gas, water) do not always follow roads and their planning could also benefit from our measure. Our method enables to determine which buildings are spatially connected and at which specific distance. Second, by using a distance-weighted graph, we find that our measure is more likely to lead to the study of the relationships and interactions that can exist between points (buildings). Indeed, the network approach allows us to distinguish points that are connected to each other while measuring the characteristics of each group. A parallel can be drawn with ecological research where graphs are used. For example, in the same way that Foltête and Vuidel (2017) delineates functional ecological zones with by means of landscape graphs, we should be able to better measure and therefore better understand the relationships between people living in different spaces of a city, or to understand how these different spaces are organized.

Our topology-based index is a contribution to increase the quality of the measurement and understanding of the morphology of the built space. We know that the topology of the buildings is only one aspect of the complexity of such space. Taking a multi-factorial approach, it would certainly be possible to develop and combine other indices with the topology-based density index as presented here. We already mentioned the addition of the third dimension (height of the buildings). But it is certainly also possible to adapt other indicators. The concentration index developed by Galster et al. (2001) could for example be adapted to measure whether or not, within a cluster, the buildings are rather aligned or form a block. Complementing the topology-based density with other indices could be the next step in this research.

8.5.2 Methodological limits

A first methodological limit is the use of the centroid of each building instead of the building footprint in the creation of the graph. We have seen that in our case, it is the distance between building centroids that must be considered in order to obtain a calculable and more easily interpretable measure. We note, however, that in some cases the interpretation of the measurement may lead to a poorer perception of the built environment. For example, a given spatial organization of centroids may reflect the location of small buildings that are far apart (i.e. isolated farms) or very large buildings

that almost join (i.e. industrial zoning).

The use of a Cook's distance for identifying the outlier in the Moran scatterplot can also be discussed. There are many alternative graphical (scatterplot, boxplot, etc.) or analytical (standardised residuals, hat matrix, etc.) methods of detection of outliers in regression (Ampanthong and Suwattee, 2009). Analytical methods have the advantage that they do not require human visual interpretation. Cook's distance is pointed out by Ampanthong and Suwattee (2009) as one of the best indices for the detection of outliers in multiple regression. In our paper, we find it interesting because it combines both residual information (is a point far from the line?) and information on the influence of each point in the regression. It should be noted that the method identifies a single outlier (the observation for which the Cook's distance is maximum). If the outlier is clearly identifiable, the different indices will converge. If several outliers are present and if they are not clearly identifiable, results might not converge. However, we did not encounter this figure in our empirical analyses, but are aware that it could happen. Another limitation that should be investigated in the future with the help of statisticians concerns the use of outliers in a regression whose slope is not significant. This happens very little in our case (5%) but could happen in a more important way if someone wants to work on a less homogeneous variable than the distance between centroid buildings.

Another methodological limit concerns the use of the 10 000 metres threshold as a constraint on the removal of an outlier. On the one hand, the objective of the method is not impacted by the value of the threshold. Whatever the threshold, the method always creates clusters whose topologies tend to be more and more homogeneous during the iterations. On the other hand, the threshold can modify the scale at which the method will stop. A high value will lead to the creation of large clusters (large length of graph) whereas a small value leads to the creation of very fine scale clusters. The threshold of 10 000 seems to be the most relevant for density measurements in Belgium but remains debatable.

Last but not least, the choice of variance test can also be a source of discussion. Indeed, it is important *a priori* to control the distribution of the populations tested when carrying out a test of variance (Box, 1953). We do not carry out this control systematically. However, we have noticed in a large majority of cases that the distribution of the length of the edges in the (sub)graphs was heavy-tailed (as shown in Fig.8.5). Therefore, we opted for the Brown & Forsythe variance test. This test is the most appropriate for this type of distribution (it does not consider the most extreme 10% of the

distribution) (Brown and Forsythe, 1974). In comparison with other tests, the Brown & Forsythe test gave the most visually appropriate results. One way to further improve the method may be to systematically assess the shape of the distributions to be tested and the application of the most appropriate test to each case.

8.6 Conclusion

Following Caruso et al. (2017) who sought to identify urban form patterns using methods based on graphs, and following Berghauser Pont and Haupt (e.g. 2005) who sought to measure buildings' density with distinct indices depending on spatial units, we develop here a method to obtain a built density index that preserves the topology of buildings. This means that we can now identify clusters of buildings with homogeneous inter-building distances, and we can further measure, for each cluster, the density of buildings while preserving information about their relative positions.

Our method works in two steps. After retrieving the centroids of the buildings, the first step in the method consists in a Spatial Descending Hierarchical Clustering (iterative approach). Based on a Minimum Spanning Tree weighted by inter-building distances, a Moran scatterplot combined with a maximum Cook's distance are used to identify the edge of the MST that diverges most from its neighbours (outlier). This edge is removed if it meets several criteria; one of these is the inequality of the variances of the lengths of the edges with and without outlier. At the end of step 1, clusters of buildings with homogeneous inter-building distances are delineated. In the second step, the topology-based density is computed by dividing the number of buildings in a cluster by the total length of the MST connecting all buildings in that cluster.

The method is applied to all buildings located in Belgium. Clusters with homogeneous inter-building distances are clearly identified. For example, some clusters refer to the organization in a compact village, others to a linear development along a road, or to a housing estate organization, etc. For each cluster, the value of the newly developed density index reflects the topology i.e. the relative position of buildings. A high (low) density will be measured when the distance between buildings is small (large). The topology-based density index is then only influenced by the relative position of the buildings (average inter-building distance). This is not the case for standard density measures, using an *a priori* fixed surface. Topology-based density is therefore a quite useful index for measuring and understanding built-up patterns.

9

Conclusion

This chapter concludes our work. It first looks back at our main findings in terms of delineation and characterisation of urban spaces. Can we understand our measurements in a meaningful way? This chapter then discusses the limitations of our work. We finish by taking the opportunity to put forward ideas for future research.

9.1 Summary of our findings

9.1.1 General objectives of this thesis

A city is a complex space. It is in fact the result of the addition of several spaces (physical, functional, etc). There are a number of possible approaches to understand the organisation of such spaces. Chapter 2 examined four methods that make it possible to understand a region (the Brussels-Capital Region) and its spatial structure. It is clear that, depending on the method used, the spatial structure revealed varies. Our understanding of an urban space is therefore dependent on the measurements we use to describe it.

This thesis has therefore focused on some measurements that help us to comprehend the spatial organisation of urban spaces. The aim was to understand what the different methods and indices actually measure in order to interpret their results as accurately as possible. We worked on two essential aspects of understanding an urban space: delineation and characterisation. In order to have a realistic research field, we limited our research to a morphological approach only. Only physical elements (buildings and road intersections) were considered here. We are fully aware that this choice only captures part of urban complexity. However, the physical elements are easily collected and allow comparisons between study areas and with populations.

We have seen throughout this thesis that it is important to grasp the mechanisms and biases behind the methods and measurements used. Comprehending the way in which the methods and measurements work means that we can better understand the results they provide. In any case, we do not claim that one method or index is better than another for delineating and characterising urban spaces. Rather, we emphasise that the use of one method over another must be motivated by a specific research objective.

9.1.2 Delineation and Characterisation

Chapter 5 provided further evidence that, in addition to the extent of the study area, the use of one method to delineate an urban area rather than another leads to different conclusions. It is therefore crucial to link method and objective. Indeed, Belgium sometimes appears as one large conurbation and sometimes as having well-separated urban entities. For example, researchers or urban planners wishing to delimit spaces that are morphologically similar may prefer to use *MorphoLim*. On the other hand, if they are looking to delimit the largest continuous morphological space, they

will give priority to the *Hierarchical Percolation method*.

It is impossible to delimit an urban area in a single and unambiguous way. Chapter 5 showed that even when using methods that do not require data aggregation or a priori threshold definition, the results can vary greatly. The data used, the extent of the study area and the delineation method all have an impact on the results. A single morphological delineation does not exist, just as a single functional delineation is impossible (also depending on data and methods). In order to understand the results of a delineation with *MorphoLim*, chapter 6 showed that three categories of variables were relevant. These are variables describing (1) the size of urban agglomerations, (2) the spatial organisation of their built-up areas and (3) the morphological characteristics of their boundaries. Once again, each category of variable makes it possible to highlight specific characteristics for the same agglomeration. They are complementary, not competing.

Chapter 7 showed that the use of the third dimension (number of floors in buildings) gives us a deeper insight into the city through different delimitations. It is indeed possible to delimit the urban space from ground level. But it is also possible to delimit specific urban spaces that have a particular building height. Again, it is not a question of preferring one delimitation over another, but of using all the delimitations to better understand the space. In the case of Valencia, we saw that the morphological agglomeration at ground level was very difficult to delimit because of very diffuse urbanisation. The delineation of the agglomeration of Valencia looks different when buildings of a greater height are considered. We find that the urban area of Valencia has developed very strongly through buildings with few floors, probably intended for residential use. The upper city appears thanks to the application of *MorphoLim* but for buildings with a higher number of floors. It would then be possible to link some specific functional delimitations to some morphological delimitations at different levels (number of floors).

Without questioning the usefulness of the traditional density indices, we found that they were often computed using a basic spatial unit fixed a priori and that the distance within these units was not taken into account. We focused on a morphological element (buildings) and developed a method to obtain spatial units (graphs) in which the spatial organisation of these buildings was homogeneous (same distance between them). The topology-based density index therefore retains the notion of distance between elements. Consequently, this index reflects the spatial organisation of the buildings because it adapts to it. Whatever the study area, the density index will give an idea of the average distance between the elements for which it is

measured.

Beyond the morphological meaning that can be assigned to the topological density index, this index facilitates easier comparison between different study areas. Indeed, there is no need to choose a common basic spatial unit that might not make sense in all areas.

As a complement to Euclidean geometry, it has already been shown that fractal geometry can help to characterise complex urban spaces. However, it became clear in Chapter 4 that the fractal dimension is, like most indices, dependent on the computation method used. The interpretation and comparison of fractal dimensions for different spaces therefore requires a careful use of the computation methods (choice of parameters). We therefore established a coherent protocol to measure the fractal dimension (box-counting and correlation). We also provided some keys to understand the morphological meaning of the values taken by these dimensions. In particular, we observed that the presence of unbuilt spaces of irregular sizes and spatial distribution could cause differences between the box-counting dimension and the correlation dimension.

9.2 Critical assessment and avenues for further research

In this section, we will attempt to address some of the limitations and criticisms of our work and will endeavour to look to the future. We will acknowledge that our contribution to the understanding of urban spaces is modest, as this is a vast field of research. We will then explain why it is difficult at present to fully exploit our results. Finally, we will see that understanding is, unfortunately, not predicting when it comes to spatial planning.

9.2.1 A vast field of research

The aim of this thesis is to contribute to the understanding of the measurements that enable researchers and urban planners to understand urban spaces. Of course, we have to admit that our work only touches on a very small part of the existing themes, methods and indices.

The morphological approach is inevitably reductive. In the introduction, we noted that in order to understand a city, its many facets need to be measured. We have deliberately confinded ourselves to the morphological aspects, but clearly it is at least as important to have a proper understanding of the indices that use a functional approach. The representations

(delimitation or characterisation) that we have made of urban areas (such as Brussels) are therefore just one of several possible representations. However, we have tried to associate each of these with the methodological mechanisms that lead to them (and therefore sometimes bias them).

The data we used is not very diverse. Since our aim was to shed light on the meaning of some methods and measurements, we attached little importance to the data itself. The easy access to good quality data in many countries is a major advantage of the data we used, making it possible to draw comparisons between different study areas. However, the importance of the data must be mentioned, as this influences the results we analyse. For the sake of simplicity and in order to keep a readable thread throughout the thesis, we have limited ourselves in our choice of data by working almost exclusively on the basis of buildings (and a little on the basis of road intersections). For example, we do not analyse public transport networks (train, metro) or other infrastructures that help structure urban spaces. Our representations of urban spaces, and therefore what we understand about them, are therefore only some of the possible morphological representations. We should also note that we have always considered urban spaces through their buildings and by regarding the rest of the space as empty. This approach is rather incomplete since it makes no distinction between a tarmac space (road, car park, etc.) and a green (park, garden, etc.) or a blue space (river, pond, etc.). Except in Chapter 7 where we included the number of floors of the buildings, we have only used their footprint and did not concider their function (residential, office, commercial, etc.) or their condition (old, renovated, new, etc.). It would be interesting to take these variables into account to gain a better understanding of morphological spaces.

In addition to the single (morphological) approach and the limited diversity of data, we recognise that we have been interested only in forms of spatial organisation and not in the processes that structure them. In her definition of urban form, Moudon (1997) points out the importance of a historical approach, as physical elements are constantly changing and being replaced. For example, it would have been interesting to analyse the boundaries of the same urban space at different time periods. This would perhaps have made it possible to highlight morphological singularities measured by the methods we have used. We have not addressed this in this manuscript, but we have reproduced for Montero et al. (2199) a delimitation of the Brussels conurbation using *MorphoLim* as carried out in Chapter 5. The data used for Montero et al. (2199) is from 2018, compared 2009 for the data in Chapter 5. While only 10 years separate the two databases, it can be observed that the agglomeration of Ghent merges with that of Brussels with

the 2018 data (they were only close with the 2009 data). This result tends to confirm, for example, that ribbon urbanisation (detected by *MorphoLim*) is still developing in Belgium and that the Brussels conurbation is still expanding via this process.

We confined ourselves to a very limited number of measurements and methods. Our main objective is to contribute to urban form research by studying the meaning of some measurements. In the end, we worked on very few of these measurements. By way of comparison, Schwarz (2010) uses up to 27 indices to revisit the urban forms of European cities. The scholarly literature has innumerable indices to morphologically describe an urban space, and some confusion exists regarding all these indices and their uses. In a review of 72 quantitative studies, Fleischmann et al. (2021b) find 465 individual measurable urban form characters (some measuring the same thing but with different names or having identical names but measuring different things). Although our work focuses on only a few indices, it seems relevant because it focuses on the meaning that can be given to these indices. Anyone who uses one of the available indices should question the way in which the index is constructed and thus influences the results.

Finally, we are aware that our work still lacks external validation (Ramspek et al., 2021). We have limited ourselves to internal validations often based on our knowledge of the study areas. Tests on other cities would be necessary to support our conclusions. For example, we observe discrepancies between our covariations of indices characterising an agglomeration (see Chapter 6) and those observed by Tannier and Thomas (2013). We attribute these differences to methodological aspects but also question the spatial structures (they would be different in France and Belgium). It would therefore be relevant to carry out a new study, retaining our methodological choices but applying them to other areas. We think that Belgium is a good example of urban/periurban/rural built up complexities, leading to strong methodological and geographical results. However, some external validations in other parts of the World would be more than welcome.

9.2.2 Form to function?

In our introduction, we noted that "the physical part of the city, essentially buildings, roads and plots, reveals the presence of inhabitants, activities and hence movements (flows), which themselves relate to planning and normative issues" (such as what is a good city?). In order to exploit our results (indices and delimitations), they should therefore be linked to the presence of inhabitants, activities and hence movements. But how can we achieve this with our measurements? We have criticised measurements that use predefined

basic spatial units such as administrative divisions, but it must be recognised that with such methods, the link with inhabitants and their ways of occupying the space are more direct than with our results. Indeed, the results obtained with traditional methods using administrative divisions can be linked to a series of data aggregated at a specific administrative level (census, surveys, etc.) made available by the authorities (federal, regional, municipal, etc.). Unfortunately, our results often do not correspond to a territorial breakdown that provides easy access to data. On a local scale, our delimitations of urban spaces or clusters (for the calculation of topological density) divide buildings belonging to different municipalities or statistical sectors or bring together buildings belonging to different entities. Even on a more global scale, we can see, for example, that the Brussels conurbation can extend far beyond the Brussels-Capital Region into Flanders and Wallonia according to *MorphoLim*. So which organisation should be contacted for data collection or planning within the Brussels conurbation? This is a crucial question, considering that in Belgium it is the Regions that are the competent bodies in terms of, for example, spatial planning and transport.

To study the impact of our results (delimitation and characterisation of urban spaces) on the processes that take place in them, we see two possible avenues of research: (1) to obtain from data suppliers data aggregated at the scale of our partition and (2) to use individual geolocalised data (ICT, GPS, etc.). The first option requires negotiation with traditional data providers. For obvious reasons of confidentiality, it is impossible to obtain individual data at building level. However, it is possible to obtain some data (socio-economic, etc.) at a specific level of aggregation. Obtaining the data is of course more tedious and delicate than when using traditional administrative breakdowns (which inevitably slows down the work). However, we believe that it may be worthwhile.

Individual geolocalised data is becoming increasingly available and accessible. For example, Jones et al. (2016) presented a partition of Belgium based on user requests for train journeys via a dedicated application (*iRail*). It would be conceivable to use such a database to observe possible divergence or convergence between users identified as living in our morphological agglomerations and those living outside them. Similarly, using, for example, data from the *TEC* (Transport En Commun en Wallonie) application, we could look for a link between topological density and bus use in Wallonia. It is important to note here that this type of data requires a great deal of collection, sorting and pre-processing work beforehand (Adam, 2019). Furthermore, it should be borne in mind that the user populations are not specifically a representative sample of the general population.

So we see that in order to help the United Nations meet its commitment to sustainable cities, there is still a lot of research work to be done.

9.2.3 Understanding is not predicting

In his book *"L'espace, un objet central de la sociologie"* sociologist Jean Remy warns that there is no mechanical relationship between spatial form and social practices (Remy, 2015). It is not, for example, because a neighbourhood is built in such a way as to favour interactions between neighbours that these will take place and that they will be good. We need to be aware that there is a difference between the actual morphological space as we measure it and the space as it is perceived and experienced by the people who frequent (live, work, pass through, etc.) these spaces.

Understanding physical (real) space via measurements that we understand does not mean that we are able to give infallible recommendations in terms of urban planning. For example, our work makes it possible to measure topological densities, but it does not currently enable us to define an optimal density towards which different spaces should aim.

We are therefore aware that measuring and understanding physical space is not automatically linked with how that space is used. However, we believe that it is relevant and important to continue research into the measurement of physical space in order to better measure its relationship with lived space. Indeed, it seems difficult to understand how lived space overlaps with real space if the latter is poorly measured. Most studies of neighbourhood effects use the neighbourhood as an amorphous concept and compare neighbourhoods that are generally of the same standard size (Sastry et al., 2002). However, a more detailed study of neighbourhoods, presenting them in their morphological singularity, could allow a more accurate representation of the space and would then lead to a better perception of the relationships between these real and experienced spaces (Coulton et al., 2013; Vallée et al., 2016). We are therefore once again aware that our work is only a drop in the ocean of research and that there is still much to explore.

9.3 Concluding remarks

In this thesis, our main objective was to contribute to the understanding of measurements that, in turn, provide a better understanding of morphological urban spaces. We looked at the issues of agglomeration delineation by comparing methods and questioning the meaning of the results obtained. We then considered how to characterise urban spaces. We developed a method to calculate topological densities. We also tried to understand how to measure fractal dimensions and how to interpret their values.

The complexity of cities means that they are difficult to measure. In our work, we emphasise that it is impossible to use methods and indices that are free of bias. Researchers and urban planners must bear in mind that the results and conclusions they can present following a study are always affected by the choices they have made. The choice of data, methods and study areas are all factors influencing their results and therefore their conclusions and recommendations.

We will end this thesis with a call for multidisciplinarity. We have worked and reasoned as geographers. Geography is a science at the crossroads of many disciplines. We are therefore perfectly aware that we need many other disciplines, each holding a key to the understanding of space. As a sociologist, Jean Remy sees the city as the result of four dimensions. *A morphological dimension specific to the built environment*, a dimension that refers to *social morphology and demography*, a *socio-affective* dimension and a *socio-political* dimension (Remy, 2015, p.12). Each dimension must be studied separately. We situate our research easily in the first of these dimensions. But it is also necessary to try to understand the interactions between the different dimensions. To do this, it is absolutely essential to cultivate links between disciplines and encourage geographers, architects, urban planners, sociologists, political scientists, etc. to work together.

A

Appendices of chapter 2

A.1 48 variables used in the factorial ecology

- % children in the neighbourhood and around, enrolled in kindergartens in the neighbourhood.
- % children in the neighbourhood and around, enrolled in a primary school in the neighbourhood.
- Population density (inhabitants/sq km).
- Offices density (sq meters/sq km).
- % of buildings with 5 floors and more.
- % of impervious surfaces.
- % dwellings built before 1961.
- % households unsatisfied with the cleanness of the environment around their housing.
- % of population aged 18–29.
- % of population aged 65–79.
- % population aged 80 and more.
- Average age (years).

- Mobility index (sum of the immigrants and emigrants divided by total population).
- Sedentariness index (non-migrants divided by the total population).
- % of couples with children in total number of private households.
- % of EU population (EU15) in the total population.
- % of North Africa population in the total population.
- % of Latin America population in the total population.
- % of Sub-Saharan Africa population in the total population.
- % of foreigners in the total number of inhabitants.
- % of inhabitants with French nationality.
- % people living alone within the 18–29y old.
- % people living alone within the 65 and more.
- Mean size of private household.
- Index of masculinity (number of men *100 divided by the number of women).
- % of the 0–17y old. in the total population.
- % population aged 65 and more.
- Economic dependence coefficient (number of 0–17 years and 65+ years divided by the number of 18–64 years).
- Ageing coefficient (%) (population of 65+yaers divided by the population of 0–17years).
- Activity rate (%) (number of actives divided by population of 18–64 years).
- Application rate (%) (total number of unemployment people divided by population of 18–64 yaers)
- Unemployment rate (%) (total number of unemployment people divided by the active population)
- Medium income of the tax report (€).
- % of the salaried within the labor force.Employment rate (

- Absolute difference between male and female activity rates.
- % households residing in an apartment.
- % dwellings under 55 sq meters.
- % dwellings larger than 104 sq meters.
- Average area of a dwelling (sq m).
- Average area of a dwelling by inhabitant.
- % of the social housing (number of social housing for 100 households).
- % dwellings occupied by the owner.
- Number of renovation subsidies (for 1000 households) (‰).
- % of inhabitants living close from a public transportation stop (250 m from the bus, 400m from a tram and 500m from the metro).
- Density of private households (number of private households divided by the area of the district)
- % of children (less than 18y) living in a household without labour income.
- % inhabitants estimating themselves not to be in good health.

B

Appendices of chapter 5

B.1 Natural cities: head/tail breaks tables

Head/tail breaks method applied to the rank-size distribution of morphological clusters, i.e. polygons formed by the continuous edges of the TIN whose length is shorter than the mean edge length. Grey line in the Table: end of the head/tail breaks process. After this step, the percentage of clusters in the head of the distribution is equal to 1 or higher than 40%.

Step of the head/tail breaks	Total number of morphological clusters	Mean area of the clusters (in m^2)	Number of clusters in the head of the distribution	Number of clusters in the tail of the distribution	% of clusters in the head of the distribution	% of clusters in the tail of the distribution
1	4153	0.0854	375	3778	9%	91%
2	375	0.8105	48	327	13%	87%
3	48	4.7974	7	41	15%	85%
4	7	22.6925	1	6	14%	86%

Table B.1. Street nodes of the province of Brabant

Step of the head/tail breaks	Total number of morphological clusters	Mean area of the clusters (in m^2)	Number of clusters in the head of the distribution	Number of clusters in the tail of the distribution	% of clusters in the head of the distribution	% of clusters in the tail of the distribution
1	23770	0.1082	2435	21335	10%	90%
2	2435	0.8980	402	2033	17%	83%
3	402	4.0189	75	327	19%	81%
4	75	13.8951	17	58	23%	77%
5	17	39.2771	5	12	29%	71%
6	5	81.3112	2	3	40%	60%
7	2	118.8025	1	1	50%	50%

Table B.2. Street nodes of the entire country (Belgium)

Step of the head/tail breaks	Total number of morphological clusters	Mean area of the clusters (in km^2)	Number of clusters in the head of the distribution	Number of clusters in the tail of the distribution	% of clusters in the head of the distribution	% of clusters in the tail of the distribution
1	22629	0.0191	1517	21112	7%	93%
2	1517	0.2575	250	1267	16%	84%
3	250	1.2446	63	187	25%	75%
4	63	3.3218	14	49	22%	78%
5	14	8.2630	4	10	29%	71%
6	4	16.7524	3	1	75%	25%

Table B.3. Building centroids of the province of Brabant

Step of the head/tail breaks	Total number of morphological clusters	Mean area of the clusters (in km^2)	Number of clusters in the head of the distribution	Number of clusters in the tail of the distribution	% of clusters in the head of the distribution	% of clusters in the tail of the distribution
1	137445	0.0199	8815	128630	6%	94%
2	8815	0.2793	1372	7443	16%	84%
3	1372	1.4354	369	1003	27%	73%
4	369	3.6304	104	265	28%	72%
5	104	7.2101	36	68	35%	65%
6	36	11.6104	9	27	25%	75%
7	9	19.9756	2	7	22%	78%
8	2	35.9760	1	1	50%	50%

Table B.4. Building centroids of the entire country (Belgium)

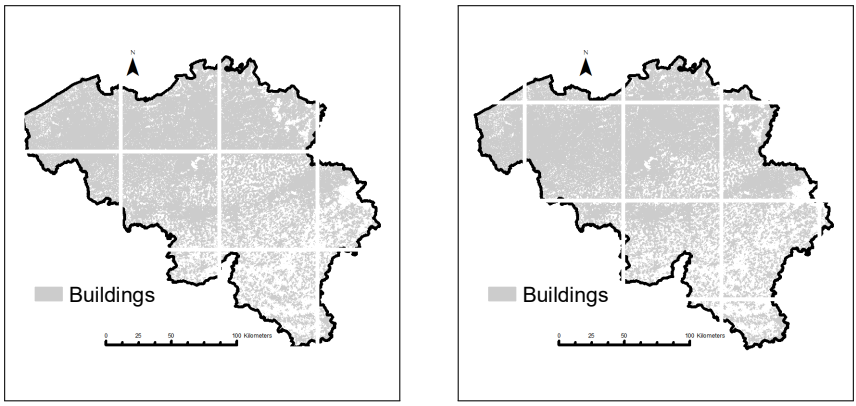
C

Appendices of chapter 8

C.1 Chunking of database

To limit computation time, the database is divided into different chunks spatially adjacent to each other. However, this chunking produces border effects that might bias the results. Indeed, two buildings belonging potentially to the same morphological structure can be separated if they belong to two different chunks. To avoid these border effects, a second chunking process is applied to the same dataset (figure C.1). The method runs for each chunk separately. The subgraphs close (1,000 metres) to the borders of chunking 1 are replaced by subgraphs from chunking 2 to give the buildings' density of Belgium (section 8.4).

The final results of the method are not dependent on the initial chunking. Indeed, if each building is assigned the density value of the subgraph to which it belongs, then more than 92% of the buildings have exactly the same density with chunking 1 and chunking 2. If buildings within 1,000 metres of the chunking boundaries are not taken into account (boundary effect), the percentage of identical density between the two chunkings increases even more to 97%.

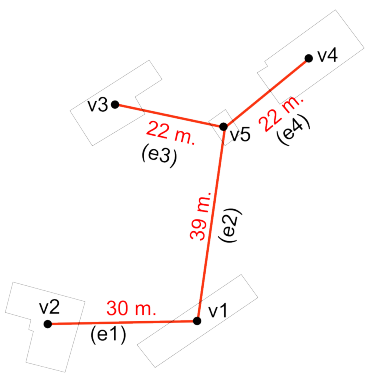


(a) Chunking 1 **(b)** Chunking 2

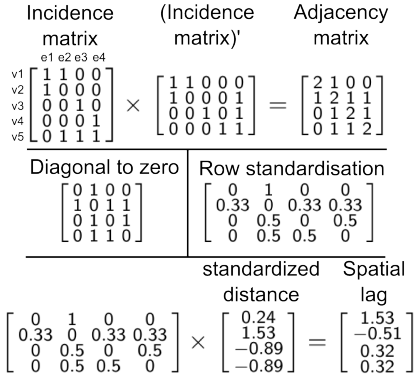
Figure C.1. The two chunkings

C.2 Calculation of the Spatial Lag

The spatial lag is the product of the row standardised matrix of adjacency with the standardised vector of distance between the centroids of the buildings. The adjacency matrix (diagonal equal to zero) is obtained by the product of the incidence matrix with its transpose.



(a) a simple example



(b) mathematical process

Figure C.2. Calculation of the spatial lag on a simple spatial toy structure of buildings

C.3 Sensitivity to condition (1)

Condition (1) (8.4) expresses the fact that if the total length of the graph is higher than 10 000 metres, the outlier is removed. With this condition, we ensure that the method detects fine-scale morphological features. We assume that a graph longer than 10 000 m. is too heterogeneous with regard to its inter-building distances. It is therefore necessary to remove the outlier detected in the Moran scatterplot to allow the method to search for more homogeneous topological clusters as it does in condition (3). Note that the 10 000 m. threshold was not selected at random: several values were tested (see examples in Figures C.3,C.4 and C.5). With values larger than 10 000, the algorithm cannot distinguish different topological patterns, especially in the countryside. Conversely, with values lower than 10 000 the algorithm detects small local topological differences that are not relevant in terms of density measure.

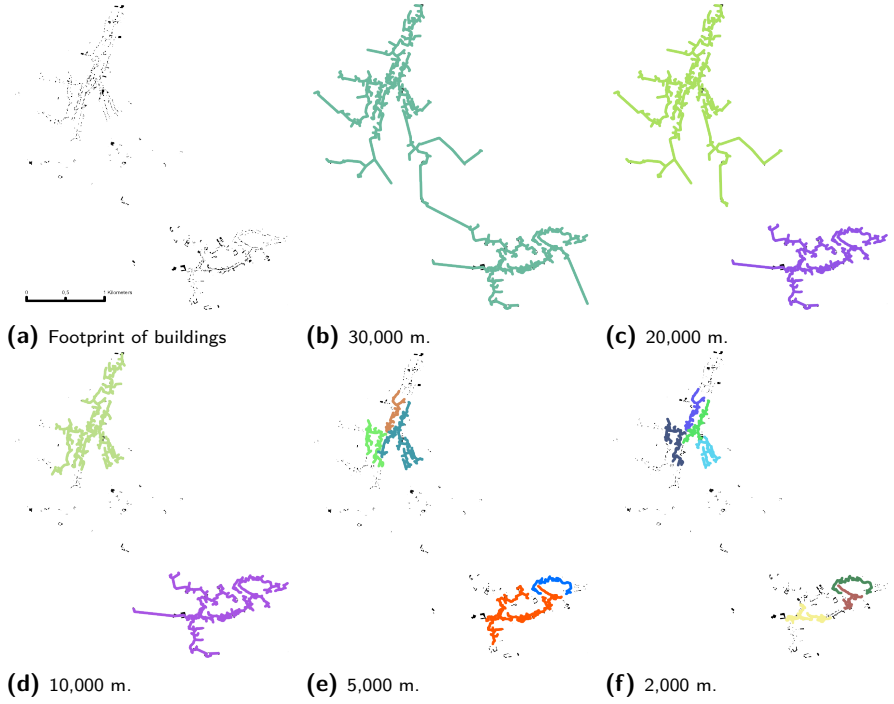
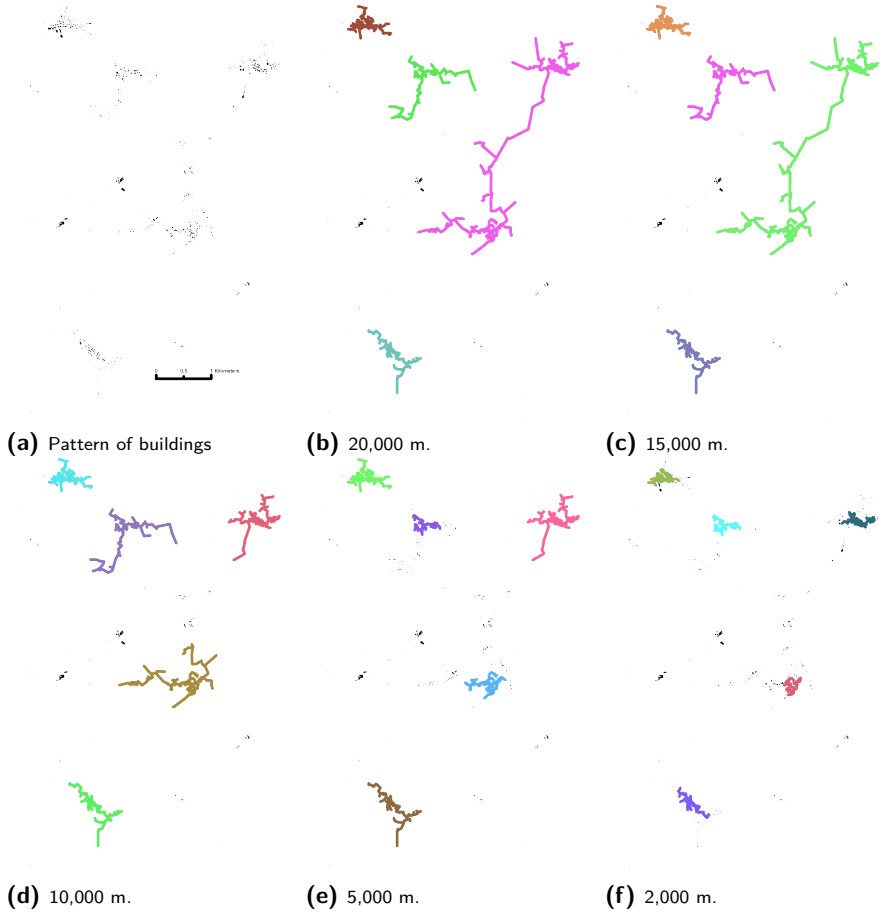


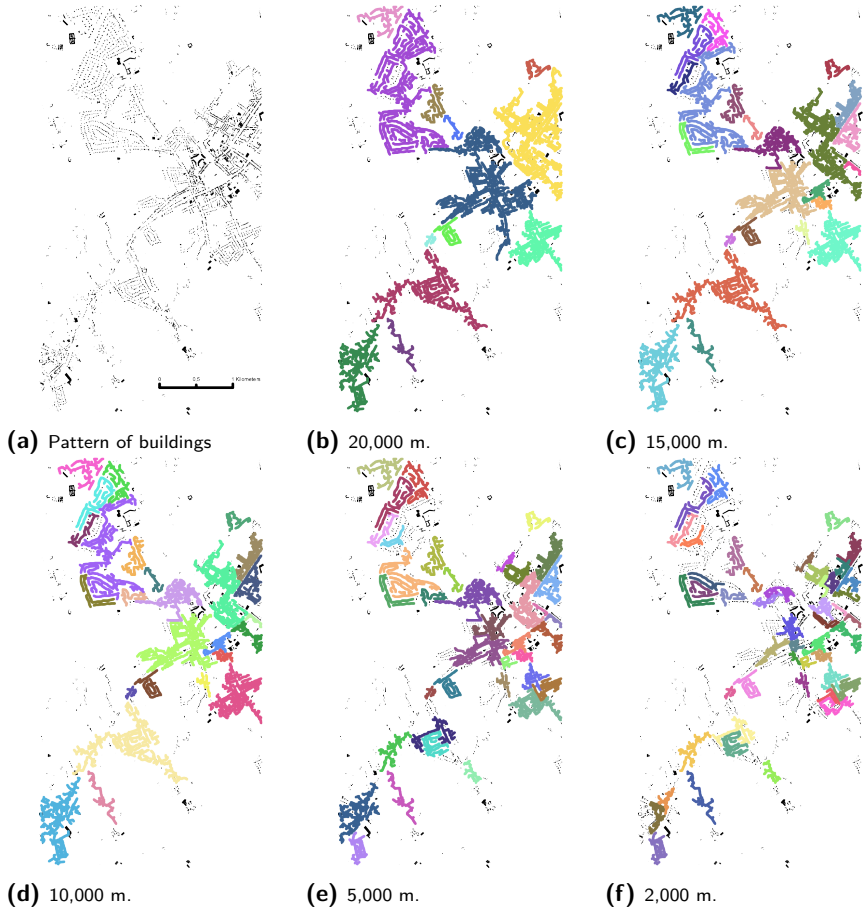
Figure C.3. Comparing different thresholds in condition (1) (see 8.4) on a given pattern of buildings (a)



(a) Pattern of buildings **(b)** 20,000 m. **(c)** 15,000 m.

(d) 10,000 m. **(e)** 5,000 m. **(f)** 2,000 m.

Figure C.4. Comparing different thresholds in condition (1) (see 8.4) on a given pattern of buildings (a)



(a) Pattern of buildings **(b)** 20,000 m. **(c)** 15,000 m.

(d) 10,000 m. **(e)** 5,000 m. **(f)** 2,000 m.

Figure C.5. Comparing different thresholds in condition (1) (see 8.4) on a given pattern of buildings (a)

C.4 The 26 462 clusters for Belgium according to the SDHC

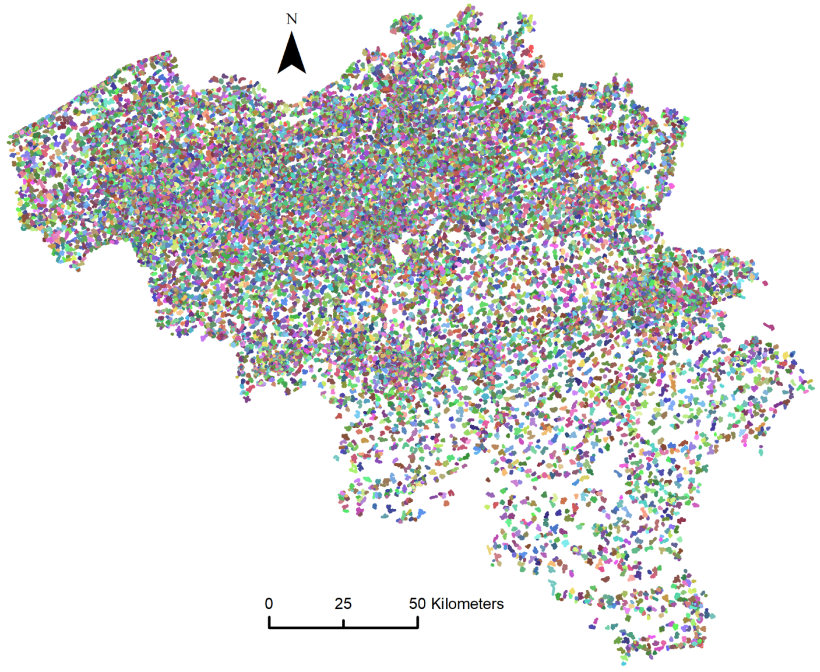


Figure C.6. The 26 462 clusters following spatial clustering

C.5 The topology-based built density index computed for Belgium

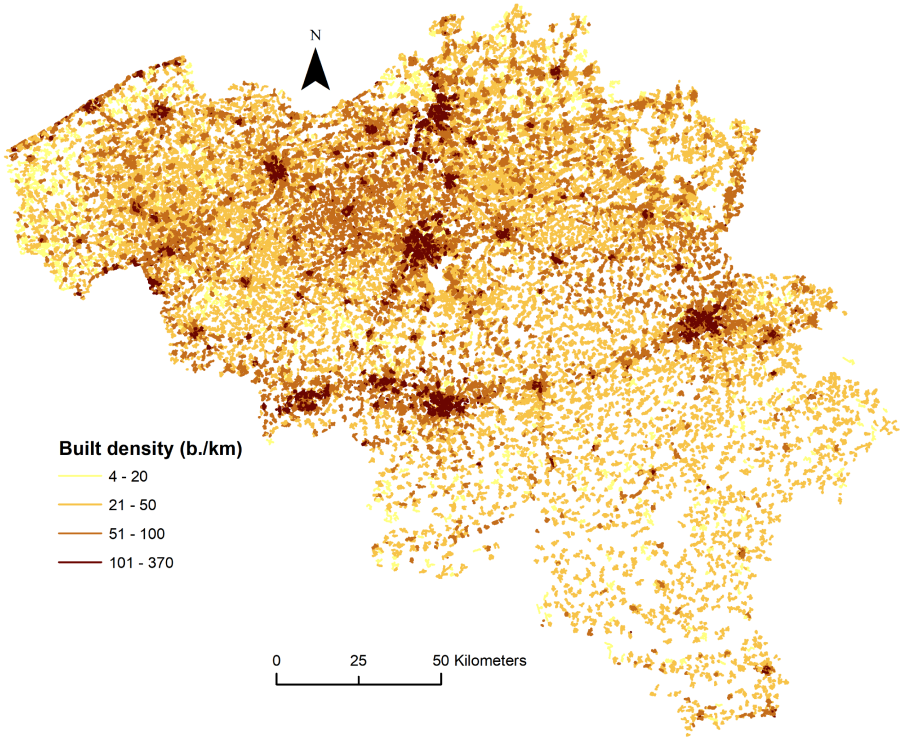


Figure C.7. The topology-based built density index computed for Belgium

Bibliography

Adam, Arnaud (2019) "Exploring new geographies of interactions in and around the metropolitan area of Brussels," Ph.D. dissertation, Université catholique de Louvain (UCLouvain).

Adam, Arnaud, Julien Charlier, Marc Debuisson, Jean-Paul Duprez, Isabelle Reginster, and Isabelle Thomas (2018a) "Bassins résidentiels en Belgique : deux méthodes, une réalité ?" *Espace géographique*, Vol. 47, p. 35, Number: 1.

Adam, Arnaud, Jean-Charles Delvenne, and Isabelle Thomas (2017) "Cartographies des champs d'interaction dans et autour de Bruxelles : navettes, déménagements et appels téléphoniques," *Brussels Studies*, Vol. 1, pp. 1–18.

——— (2018b) "Detecting communities with the multi-scale Louvain method: robustness test on the metropolitan area of Brussels," *Journal of Geographical Systems*, Vol. 20, pp. 363–386.

Agterberg, Frederik P. (2013) "Fractals and spatial statistics of point patterns," *Journal of Earth Science*, Vol. 24, pp. 1–11.

Ahas, R., A. Aasa, Y. Yuan, M. Raubal, Z. Smoreda, Y. Liu, C. Ziemlicki, M. Tiru, and M. Zook (2015) "Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn," *International Journal of Geographical Information Science*, Vol. 29, pp. 2017–2039, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13658816.2015.1063151.

Ampanthong, Pimpan and Prachoom Suwattee (2009) "A Comparative Study of Outlier Detection Procedures in Multiple Linear Regression," *Proceedings of the International MultiConference of Engineers and Computer Scientists*, p. 7.

Anders, Karl-Heinrich, Monika Sester, and Dieter Fritsch (2001) "Analysis of settlement structures by graph-based clustering," p. 10.

Andrienko, Gennady, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski (2010) "Space, time and visual analytics," *International Journal of Geographical Information Science*, Vol. 24, pp. 1577–1600, Number: 10.

Angel, Shlomo, Patrick Lamson-Hall, and Zeltia Gonzalez Blanco (2021) "Anatomy of density: measurable factors that constitute urban density," *Buildings and Cities*, Vol. 2, pp. 264–282, Number: 1.

Anselin, Luc (1995) "Local indicators of spatial association—LISA," *Geographical analysis*, Vol. 27, pp. 93–115, Number: 2.

——— (1996) "The Moran scatterplot as an ESDA tool to assess local instability in spatial association," in *Spatial Analytical Perspectives on GIS*, Vol. 4.

Araldi, Alessandro and Giovanni Fusco (2019) "From the street to the metropolitan region: Pedestrian perspective in urban fabric analysis," *Environment and Planning B: Urban Analytics and City Science*, Vol. 46, pp. 1243–1263.

Arcaute, E., E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty (2014) "Constructing cities, deconstructing scaling laws," *Journal of The Royal Society Interface*, Vol. 12, pp. 20140745–20140745, Number: 102.

Arcaute, Elsa, Carlos Molinero, Erez Hatna, Roberto Murcio, Camilo Vargas-Ruiz, A. Paolo Masucci, and Michael Batty (2016) "Cities and regions in Britain through hierarchical percolation," *Royal Society Open Science*, Vol. 3, p. 150691, Number: 4.

Arlinghaus, Sandra Lach (1985) "Fractals Take a Central Place," *Geografiska Annaler. Series B, Human Geography*, Vol. 67, pp. 83–88, Publisher: [Swedish Society for Anthropology and Geography, Wiley].

——— (1987) *Essays on Mathematical Geography–II*, Ann Arbor: Institute of Mathematical Geography, Monograph Series, Monograph #5,: Institute of Mathematical Geography, Accepted: 2008-04-24T15:19:08Z.

Arribas-Bel, Daniel, M.-A. Garcia-López, and Elisabet Viladecans-Marsal (2019) "Building(s and) cities: Delineating urban areas with a machine learning algorithm," *Journal of Urban Economics*, p. 103217.

Assunção, R. M., M. C. Neves, G. Câmara, and C. Da Costa Freitas (2006) "Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees," *International Journal of Geographical Information Science*, Vol. 20, pp. 797–811, Number: 7.

Ban, Pengfei, Wei Zhan, Qifeng Yuan, and Xiaojian Li (2021) "Delineating the Urban Areas of a Cross-Boundary City with Open-Access Data: Guangzhou–Foshan, South China," *Sustainability*, Vol. 13, p. 2930.

Barnsley, Michael F. (1988) *Fractals Everywhere*, academic press, new york edition.

Batty, M., K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali (2012) "Smart cities of the future," *The European Physical Journal Special Topics*, Vol. 214, pp. 481–518.

Batty, M and P A Longley (1986) "The Fractal Simulation of Urban Structure," *Environment and Planning A: Economy and Space*, Vol. 18, pp. 1143–1179, Publisher: SAGE Publications Ltd.

Batty, Michael (2013) *The New Science of Cities*: MIT Press.

Batty, Michael and Paul Longley (1987) "Urban Shapes as Fractals," *Area*, Vol. 19, pp. 215–221, Number: 3.

——— (1994) *Fractal Cities: A Geometry Of Form And Function*: Academic Press.

Beguin, Hubert and Jacques-François Thisse (1979) "An Axiomatic Approach to Geographical Space," *Geographical Analysis*, Vol. 11, pp. 325–341, Number: 4.

Behnisch, Martin, Martin Schorcht, Steffen Kriewald, and Diego Rybski (2019) "Settlement percolation: A study of building connectivity and poles of inaccessibility," *Landscape and Urban Planning*, Vol. 191, p. 103631.

de Bellefon, Marie-Pierre, Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Clément Gorin (2019) "Delineating urban areas using building density," *Journal of Urban Economics*, p. 103226.

Benguigui, Lucien, Daniel Czamanski, Maria Marinov, and Yuval Portugali (2000) "When and Where is a City Fractal?" *Environment and Planning B: Planning and Design*, Vol. 27, pp. 507–519, Number: 4.

Berghauser Pont, Meta (2021) "Measuring Urban Form," *28th International Seminar on Urban Form - ISUF 2021 : Urban Form For Sustainable and Prosperus Cities, Glasgow*.

Berghauser Pont, Meta and P. Haupt (2005) "The spacemate: Density and the Typomorphology of the Urban Fabric," *Nordisk Arkitekturforskning*, Vol. 18(4), pp. 55–68.

Berghauser Pont, Meta, Per Haupt, Per Berg, Victoria Alstäde, and Axel Heyman (2021) "Systematic review and comparison of densification effects and planning motivations," *Buildings and Cities*, Vol. 2(1), pp. 378–401.

Berghauser Pont, Meta, Gianna Stavroulaki, and Lars Marcus (2019) "Development of urban types based on network centrality, built density and their impact on pedestrian movement," *Environment and Planning B: Urban Analytics and City Science*, Vol. 46, pp. 1549–1564.

Berry, Brian J.L. and Philip H. Ress (1969) "The Factorial Ecology of Calcutta," *American Journal of Sociology*, Vol. 74, pp. 445–491, Publisher: The University of Chicago Press.

Blondel, Vincent, Gautier Krings, and Isabelle Thomas (2010) "Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone," *Brussels Studies*, Vol. 42.

Bobkova, Evgeniya, Lars Marcus, and Berghauser Pont (2017) "Multivariable measures of plot systems: describing the potential link between urban diversity and spatial form based on the spatial capacity concept.," *Proceedings of the 11th Space Syntax Symposium, Lisbon, Portugal*, p. 16.

Bode, Eckhardt (2008) "Delineating Metropolitan Areas Using Land Prices," *Journal of Regional Science*, Vol. 48, pp. 131–163.

Borruso, Giuseppe (2003) "Network Density and the Delimitation of Urban Areas," *Transactions in GIS*, Vol. 7, pp. 177–191, Number: 2.

Bouligand, G (1929) "Sur la notion d'ordre de mesure d'un ensemble plan," *Bulletin des Sciences Mathématiques*, Vol. II-53, pp. 185–192.

Boussauw, Kobe, Tijs Neutens, and Frank Witlox (2012) "Relationship between Spatial Proximity and Travel-to-Work Distance: The Effect of the Compact City," *Regional Studies*, Vol. 46, pp. 687–706, Number: 6.

Box, George (1953) "Non-normality and tests on variances," *Biometrika*, Vol. 40, pp. 318–335, Number: 3-4.

Boyko, Christopher T. and Rachel Cooper (2011) "Clarifying and re-conceptualising density," *Progress in Planning*, Vol. 76, pp. 1–61, Number: 1.

Briant, A., P-P. Combes, and M. Lafoucarde (2010) "Dots to boxes: do the size and shape of spatial units jeopardize economic geography estimations?" *Journal of Urban Economics*, Vol. 67(3), pp. 287–302.

Brown, Morton B. and Alan B. Forsythe (1974) "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, Vol. 69:346, pp. 364–367.

Brunsdon, Chris, Stewart Fotheringham, and Martin E Charlton (1996) "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity," *Geographical Analysis*, pp. 281–298.

Burriel, E (2009) "La planificación territorial en la Comunidad Valenciana (1986-2009)," *Scripta Nova. Revista electrónica de Geografía y Ciences Sociales*, Vol. 13.

Calabrese, Francesco, Laura Ferrari, and Vincent D. Blondel (2014) "Urban Sensing Using Mobile Phone Network Data: A Survey of Research," *ACM Computing Surveys*, Vol. 47, pp. 25:1–25:20.

Carlson, Carl A. (1991) "Spatial distribution of ore deposits," *Geology*, Vol. 19, pp. 111–114.

Caruso, Geoffrey (2001) "Peri-urbanisation : the situation in Europe —A Bibliographical Note and Survey of Studies in the Netherlands, Belgium, Great Britain, Germany, Italy and the Nordic Countries.," *Report prepared for the DATAR (with contributions from J. Cavailhe‘ s, D. Peeters, M. Rounsevell, I. Thomas), Ministe‘re de l'Ame´nagement du Territoire et de l'Environnement, France.*

Caruso, Geoffrey, Mohamed Hilal, and Isabelle Thomas (2017) "Measuring urban forms from inter-building distances: Combining MST graphs with a Local Index of Spatial Association," *Landscape and Urban Planning*, Vol. 163, pp. 80–89.

Cetinkaya, Sinan, Melih Basaraner, and Dirk Burghardt (2015) "Proximity-based grouping of buildings in urban blocks: a comparison of four algorithms," *Geocarto International*, Vol. 30, pp. 618–632, Number: 6.

Chakraborti, Suman, Dipendra Nath Das, Biswajit Mondal, Hossein Shafizadeh-Moghadam, and Yongjiu Feng (2018) "A neural network and landscape metrics to propose a flexible urban growth boundary: A case study," *Ecological Indicators*, Vol. 93, pp. 952–965.

Chaudhry, Omair and William A. Mackaness (2008) "Automatic identification of urban settlement boundaries for multiple representation databases,"

Computers, Environment and Urban Systems, Vol. 32, pp. 95–109, Number: 2.

Chen, Yanguang (2011) "Derivation of the functional relations between fractal dimension of and shape indices of urban form," *Computers, Environment and Urban Systems*, Vol. 35, pp. 442–451, Number: 6.

——— (2016) "Defining urban and rural regions by multifractal spectrums of urbanization," *Fractals*, Vol. 24, p. 1650004, Number: 01.

Chen, Zifeng and Anthony Gar-On Yeh (2022) "Delineating functional urban areas in Chinese mega city regions using fine-grained population data and cellphone location data: A case of Pearl River Delta," *Computers, Environment and Urban Systems*, Vol. 93, p. 101771.

Cheshire, P and G Gornostaeva (2002) "Cities and Regions: comparable measures require comparable territories," *Les Cahiers de l'Institut d'Amenagement et d'Urbanisme de la Region d'Ile de Frnace*, Vol. 135, pp. 13–32.

Churchman, Arza (1999) "Disentangling the Concept of Density," *Journal of Planning Literature*, Vol. 13, pp. 389–411, Number: 4.

Cottineau, Clementine, Olivier Finance, Erez Hatna, Elsa Arcaute, and Michael Batty (2019) "Defining urban agglomerations to detect agglomeration economies," *Environment and Planning B: Urban Analytics and City Science*, Vol. 46, pp. 1611–1626, Number: 9 arXiv: 1601.05664.

Coulton, Claudia J., M. Zane Jennings, and Tsui Chan (2013) "How Big is My Neighborhood? Individual and Contextual Effects on Perceptions of Neighborhood Scale," *American Journal of Community Psychology*, Vol. 51, pp. 140–150.

Cörvers, Frank, Maud Hensen, and Dion Bongaerts (2009) "Delimitation and Coherence of Functional and Administrative Regions," *Regional Studies*, Vol. 43, pp. 19–31.

De Keersmaecker, Marie-Laurence de, Pierre Frankhauser, and Isabelle Thomas (2004) "Dimensions fractales et réalités périurbaines. L'exemple du Sud de Bruxelles," *L'Espace géographique*, Vol. tome 33, pp. 219–240, Number: 3.

De Keersmaecker, Marie-Laurence, Pierre Frankhauser, and Isabelle Thomas (2003) "Using Fractal Dimensions for Characterizing Intra-urban Diversity: The Example of Brussels," *Geographical Analysis*, Vol. 35, pp. 310–328, Number: 4.

Dempsey, Nicola, Caroline Brown, Shibu Raman, Sergio Porta, Mike Jenks, Colin Jones, and Glen Bramley (2009) "Elements of Urban Form," in *in Jenks M, Jones C, Dimensions of the Sustainable Cities*, Vol. 2 of Springer, London, pp. 21–51.

Deng, Min, Qiliang Liu, Tao Cheng, and Yan Shi (2011) "An adaptive spatial clustering algorithm based on delaunay triangulation," *Computers, Environment and Urban Systems*, Vol. 35, pp. 320–332, Number: 4.

Deville, Pierre, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R. Stevens, Andrea E. Gaughan, Vincent D. Blondel, and Andrew J. Tatem (2014) "Dynamic population mapping using mobile phone data," *Proceedings of the National Academy of Sciences*, Vol. 111, pp. 15888–15893, Publisher: Proceedings of the National Academy of Sciences.

Ding, Chengri, Gerrit J. Knaap, and Lewis D. Hopkins (1999) "Managing Urban Growth with Urban Growth Boundaries: A Theoretical Analysis," *Journal of Urban Economics*, Vol. 46, pp. 53–68.

Dong, Qi, Shuxue Qu, Jiahui Qin, Disheng Yi, Yusi Liu, and Jing Zhang (2022) "A Method to Identify Urban Fringe Area Based on the Industry Density of POI," *ISPRS International Journal of Geo-Information*, Vol. 11, p. 128.

Dujardin, Claire, Harris Selod, and Isabelle Thomas (2008) "Residential Segregation and Unemployment: The Case of Brussels," *Urban Studies*, Vol. 45, pp. 89–113, Publisher: Temporary Publisher.

Dujardin, Claire, Isabelle Thomas, and Henry Tulkens (2007) "Quelles frontières pour Bruxelles ? Une mise à jour," *Reflets et perspectives de la vie économique*, Vol. XLVI, p. 155, Number: 2.

Ester, Martin, Hans-Peter Kriegel, and Xiaowei Xu (1996) "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Vol. 240, p. 6.

Ewing, Reid, Shima Hamidi, Guang Tian, David Proffitt, Stefania Tonin, and Laura Fregolent (2018) "Testing Newman and Kenworthy's Theory of Density and Automobile Dependence," *Journal of Planning Education and Research*, Vol. 38, pp. 167–182, Number: 2.

Falconer, Kenneth (1990) *Fractal Geometry - Mathematical foundations and applications*, John Wiley & Sons.

Fang, Chuanglin and Danlin Yu (2017) "Urban agglomeration: An evolving concept of an emerging phenomenon," *Landscape and Urban Planning*, Vol. 162, pp. 126–136.

Farinós Dasí, J, E Peiró Sánchez-Manjavacas, and C Zornoza Gallego (2019) "Análisis del suelo y del planeamiento urbano y territorial en la Comunitat Valenciana en el periodo 2006-2017," *en: Informe sobre la evolución y la situación territorial de la Comunitat Valenciana, Universitat de València*, pp. 269–354.

Feng, Jian and Yanguang Chen (2010) "Spatiotemporal evolution of urban form and land-use structure in Hangzhou, China: evidence from fractals," *Environment and Planning B: Planning and Design*, Vol. 37(5), pp. 838–556.

Feria, José M., José M. Casado-Díaz, and Lucas Martínez-Bernabéu (2015) "Inside the metropolis: the articulation of Spanish metropolitan areas into local labor markets," *Urban Geography*, Vol. 36, pp. 1018–1041.

Feria Toribio, J.M and L Martínez Bernabeu (2016) "La definición y delimitación del sistema metropolitano español: permanencias y cambios entre 2001 y 2011," *Ciudad y territorio: Estudios territoriales*, Vol. 48, pp. 9–24.

Feria Toribio, Jose Maria, Gwendoline De Oliveira Neves, and Claudia Hurtado Rodriguez (2018) "Une méthode pour la délimitation des aires métropolitaines. Application au système urbain espagnol," *Cybergéo: European Journal of Geography, Environnement, nature, Paysage, artículo 852*.

Ferreira, José Antunes, Beatriz Condessa, Joana Castro e Almeida, and Pedro Pinto (2010) "Urban settlements delimitation in low-density areas—An application to the municipality of Tomar (Portugal)," *Landscape and Urban Planning*, Vol. 97, pp. 156–167, Number: 3.

Fleischmann, Martin (2019) "momepy: Urban Morphology Measuring Toolkit," *Journal of Open Source Software*, Vol. 4, p. 1807.

——— (2021) "The Urban Atlas: Methodological Foundation of a Morphometric Taxonomy of Urban Form," *Urban Design Studies Unit Department of Architecture University of Strathclyde, UK*, p. 511.

Fleischmann, Martin, Alessandra Feliciotti, Ombretta Romice, and Sergio Porta (2020) "Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale," *Computers, Environment and Urban Systems*, Vol. 80, p. 101441.

——— (2021a) "Methodological foundation of a numerical taxonomy of urban form," *Environment and Planning B: Urban Analytics and City Science*, Vol. 0, pp. 1–17.

Fleischmann, Martin, Ombretta Romice, and Sergio Porta (2021b) "Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities," *Environment and Planning B: Urban Analytics and City Science*, Vol. 48, pp. 2133–2150.

Floridi, Luciano (2012) "Big Data and Their Epistemological Challenge," *Philosophy & Technology*, Vol. 25, pp. 435–437.

Foltête, Jean-Christophe and Gilles Vuidel (2017) "Using landscape graphs to delineate ecologically functional areas," *Landscape Ecology*, Vol. 32, pp. 249–263.

de Fomento, Ministerio (2018) "Ministerio de Fomento, DG de Arquitectura, Vivienda y Suelo. SG de Suelo, Información y Evaluación, 2018, Áreas urbanas en España 2018. Constitución, Cuarenta años de las ciudades españolas, Madrid, Ministerio de Fomento, Secretaría General Técnica, Centro de Publicaciones. http:// atlasau.fomento.gob.es/."

Frankhauser, P. (1988) "Fractal Aspects of Urban Systems," *Sonderforschungsbereich 230 « Natürliche Konstruktionen »,*, Vol. 1, pp. 67–76.

Frankhauser, Pierre (1994) *La fractalité des structures urbaines*: Anthropos.

——— (2003) "Morphologie des "villes émergentes" en Europe à travers les analyses fractales,"Technical report.

Fujishima, Shota, Naoya Fujiwara, Yuki Akiyama, Ryosuke Shibasaki, and Ritsu Sakuramachi (2020) "The size distribution of 'cities' delineated with a network theory-based method and mobile phone GPS data," *International Journal of Economic Theory*, Vol. 16, pp. 38–50.

Galster, George, Royce Hanson, Michael R. Ratcliffe, Harold Wolman, Stephen Coleman, and Jason Freihage (2001) "Wrestling Sprawl to the Ground: Defining and measuring an elusive concept," *Housing Policy Debate*, Vol. 12, pp. 681–717, Number: 4.

Ge, Meiling and Qizhong Lin (2009) "Realizing the box-counting method for calculating fractal dimension of urban form based on remote sensing image," *Geo-spatial Information Science*, Vol. 12, pp. 265–270, Publisher: Taylor & Francis _eprint: https://doi.org/10.1007/s11806-009-0096-1.

Gehl, Jan (1987) *Life between buildings: using public space*, island press edition.

Godoy-Shimizu, Daniel, Philip Steadman, and Stephen Evans (2021) "Density and morphology: from the building scale to the city scale," *Buildings and Cities*, Vol. 2, pp. 92–113, Number: 1.

Goerlich Gisbert, Francisco J., Isidro Cantarino Martí, and Eric Gielen (2017) "Clustering cities through urban metrics analysis," *Journal of Urban Design*, Vol. 22, pp. 689–708, Number: 5.

González, Romero, Albertos Puebla, and Melo Escrihuela (2018) "Trayectoria y recomposición metropolitana post-crisis en Valencia: ¿a la búsqueda de nuevos modelos?" *Ciudad y Territorio Estudios Territoriales*, Vol. L, pp. 679–696.

Goodchild, M (2016) "Laws in Geography. Conférence au Collège International des Sciences du Territoire, Paris."

Goodchild, Michael F. (1980) "Fractals and the accuracy of geographical measures," *Journal of the International Association for Mathematical Geology*, Vol. 12, pp. 85–98.

Goodchild, Michael F. and David M. Mark (1987) "The Fractal Nature of Geographic Phenomena," *Annals of the Association of American Geographers*, Vol. 77, pp. 265–278, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8306.1987.tb00158.x.

Griffiths, Patrick, Patrick Hostert, Oliver Gruebner, and Sebastian van der Linden (2010) "Mapping megacity growth with multi-sensor data," *Remote Sensing of Environment*, Vol. 114, pp. 426–439.

Guérois, Marianne, Anne Bretagnolle, Hélène Mathian, and Antonin Pavard (2014) "Functional Urban Areas (FUA) and European harmonization. A feasibility study from the comparison of two approaches: commuting flows and accessibility isochrones (Technical Report, Espon 2013 Database)," *Paris: Union Européenne.*, p. (p. 35).

Gyenizse, Péter, Zita Bognár, Szabolcs Czigány, and Tibor Elekes (2014) "Landscape shape index, as a potential indicator of urban development in Hungary," *Landscape & Environment*, Vol. 8, pp. 78–88.

Haggett, P (1972) *Geography: a modern synthesis*, harper & row edition.

Haisch, Tina and Urs Müller (2013) "Defining and measuring urban regions: A sensitivity analysis," *Papers in Regional Science*, Vol. 94, p. 9.

Han, Hao-ying, Shih-Kung Lai, An-rong Dang, Zong-bo Tan, and Ci-fang Wu (2009) "Effectiveness of urban construction boundaries in Beijing: an assessment," *Journal of Zhejiang University-SCIENCE A*, Vol. 10, pp. 1285–1295.

Harig, Oliver, Robert Hecht, Dirk Burghardt, and Gotthard Meinel (2021) "Automatic Delineation of Urban Growth Boundaries Based on Topographic Data Using Germany as a Case Study," *ISPRS International Journal of Geo-Information*, Vol. 10, p. 353.

Henderson, M., E. T. Yeh, P. Gong, C. Elvidge, and K. Baugh (2003) "Validation of urban boundaries derived from global night-time satellite imagery," *International Journal of Remote Sensing*, Vol. 24, pp. 595–609.

Hillier, Bill (1996) *Space is the machine*, Space Syntax : A configurational theory of architecture.

Hoyt, H (1939) "The structure and growth of residential neighborhoods in American cities.," *Washington, DC: Government Printing Office*.

Hu, Shougeng, Luyi Tong, Amy E. Frazier, and Yansui Liu (2015) "Urban boundary extraction and sprawl analysis using Landsat images: A case study in Wuhan, China," *Habitat International*, Vol. 47, pp. 183–195.

Hunter, Alfred A. (1972) "Factorial Ecology: A Critique and Some Suggestions," *Demography*, Vol. 9, pp. 107–117, Publisher: Springer.

Hurvitz, Philip M., Anne Vernez Moudon, Bumjoon Kang, Brian E. Saelens, and Glen E. Duncan (2014) "Emerging Technologies for Assessing Physical Activity Behaviors in Space and Time," *Frontiers in Public Health*, Vol. 2.

Indovina, F (1998) "Algunes consideracions sobre la 'ciutat difusa," *Documents d'Analisi Geogràfica*, Vol. 33, pp. 21–32.

Indovina, F and O Nel.Io (2012) "Francesco Indovina: del análisis del territorio al gobierno de la ciudad," *Barcelona, Icaria*.

Isard, Walter (1956) "Regional Science, the Concept of Region, and Regional Structure," *Papers in Regional Science*, Vol. 2, pp. 13–26, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1435-5597.1956.tb01542.x.

——— (1975) *Introduction to Regional Science*: Prentice-Hall.

Jia, Tao and Bin Jiang (2010) "Measuring Urban Sprawl Based on Massive Street Nodes and the Novel Concept of Natural Cities," *arXiv:1010.0541 [physics]*, arXiv: 1010.0541.

Jiang, Bin (2013) "Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution," *The Professional Geographer*, Vol. 65, pp. 482–494, Number: 3.

——— (2015) "Head/tail breaks for visualization of city structure and dynamics," *Cities*, Vol. 43, pp. 69–77.

——— (2018) "A Topological Representation for Taking Cities as a Coherent Whole: Topological Representation," *Geographical Analysis*, Vol. 50, pp. 298–313, Number: 3.

——— (2019) "A Recursive Definition of Goodness of Space for Bridging the Concepts of Space and Place for Sustainability," *Sustainability*, Vol. 11, p. 4091, Number: 15.

Jiang, Bin and Tao Jia (2011) "Zipf's law for all the natural cities in the United States: a geospatial perspective," *International Journal of Geographical Information Science*, Vol. 25, pp. 1269–1281, Number: 8.

Jiang, Bin and Xintao Liu (2012) "The Principle of Scaling of Geographic Space and its Application in Urban Studies."

Jiang, Bin and Yufan Miao (2015) "The evolution of natural cities from the perspective of location-based social media," *The Professional Geographer*, Vol. 67, pp. 295–306, Number: 2.

Jiang, Penghui, Qianwen Cheng, Yuan Gong, Liyan Wang, Yunqian Zhang, Liang Cheng, Manchun Li, Jiancheng Lu, Yuewei Duan, Qiuhao Huang, and Dong Chen (2016) "Using Urban Development Boundaries to Constrain Uncontrolled Urban Sprawl in China," *Annals of the American Association of Geographers*, Vol. 106, pp. 1321–1343.

Johnston, R (1978) "Residential area characteristics: Research methods for identifying urban sub-areas analysis and factorial ecology," in *D. T. Herbert & R. J. Johnston (Eds.), Social areas in cities: Processes, patterns and problems*, chichester, west sussex: john wiley and sons. edition, pp. 175–217.

Jones, Jonathan, Christophe Cloquet, Arnaud Adam, Adeline Decuyper, and Isabelle Thomas (2016) "Belgium through the Lens of Rail Travel Requests: Does Geography Still Matter?" *ISPRS International Journal of Geo-Information*, Vol. 5, p. 216.

Kauffmann, Albrecht (2012) "Delineation of City Regions Based on Commuting Interrelations: The Example of Large Cities in Germany," IWH Discussion Paper 4, Halle Institute for Economic Research, Issue: 4.

Kitchin, Rob (2013) "Big data and human geography: Opportunities, challenges and risks," *Dialogues in Human Geography*, Vol. 3, pp. 262–267.

Lam, Nina Siu-Ngan (2009) "Fractal Analysis," in *International Encyclopedia of Human Geography, Kitchin, R., Thrift, N. (dir.), Elsevier Science, 263–270.*

Le Gléau, Jean-Pierre, Denise Pumain, and Thérèse Saint-Julien (1996) "Villes d'Europe: à chaque pays sa définition," *Economie et statistique*, Vol. 294, pp. 9–23, Number: 1.

Lee, Jungwoo and Hyejung Lee (2014) "Developing and validating a citizen-centric typology for smart city services," *Government Information Quarterly*, Vol. 31, pp. S93–S105.

Martinez de Lejarza, I and J Martinez de Lejarza (2002) "Delimitación de áreas metropolitanas mediante un modelo anisótropo de decrecimiento exponencial. Una aplicación al caso del Área metropolitana de Valencia," *Estudios de Economía Aplicada*, Vol. 20, pp. 471–486.

Liu, Lingbo, Binxin Xia, Hao Wu, Jie Zhao, Zhenghong Peng, and Yang Yu (2019) "Delimitating the Natural City with Points of Interests Based on Service Area and Maximum Entropy Method," *Entropy*, Vol. 21, p. 458, Number: 5.

Liu, Qiliang, Min Deng, and Yan Shi (2013) "Adaptive spatial clustering in the presence of obstacles and facilitators," *Computers & Geosciences*, Vol. 56, pp. 104–118.

Liu, Qiliang, Min Deng, Yan Shi, and Jiaqiu Wang (2012) "A density-based spatial clustering algorithm considering both spatial proximity and attribute similarity," *Computers & Geosciences*, Vol. 46, pp. 296–309.

Longley, Paul A, Muhammad Adnan, and Guy Lansley (2015) "The Geotemporal Demographics of Twitter Usage," *Environment and Planning A: Economy and Space*, Vol. 47, pp. 465–484.

Longley, Paul A and Victor Mesev (2000) "On the measurement and generalisation of urban form," *Environment and Planning A*, Vol. 32, pp. 473–488.

Longley, Paul A. and Victor Mesev (2002) "Measurement of density gradients and space-filling in urban systems*," *Papers in Regional Science*, Vol. 81, pp. 1–28, Number: 1.

Longley, Paul and Victor Mesev (1997) "Beyond the analogue models: space filling and density measurement of an urban settlement," *Papers in Regional Science*, Vol. 76(4), pp. 409–427.

Lowe, D.G. (1989) "Organization of smooth image curves at multiple scales," *International Journal of Computer Vision*, Vol. 3, pp. 119–130.

Lynch, Kevin (1984) *Good city form*, mit press edition.

Lévy, Albert (2005) "Formes urbaines et significations : revisiter la morphologie urbaine," *Espaces et sociétés*, Vol. 122, p. 25, Number: 4.

Mandelbrot, Benoit B. (1982) *The Fractal Geometry of Nature*, San Francisco: W.H.Freeman & Co Ltd.

Marmolejo, C, J Masip, and C Aguirre (2013) "Policentrismo en el sistema urbano español: un análisis para siete áreas metropolitanas," *Ciudad y Territorio Estudios Territoriales*, Vol. XLV, pp. 281–300.

Marmolejo Duarte, Carlos, Carlos Aguirre Nuñez, and Manuel Ruiz Lineros (2010) "¿Hacia un sistema de metrópolis españolas policéntricas?: caracterización de su estructura metropolitana," p. 25.

Marshall, Stephen, Jorge Gil, Karl Kropf, Martin Tomko, and Lucas Figueiredo (2018) "Street Network Studies: from Networks to Models and their Representations," *Networks and Spatial Economics*, Vol. 18, pp. 735–749.

Martino, Nicholas, Cynthia Girling, and Yuhao Lu (2021) "Urban form and livability: socioeconomic and built environment indicators," *Buildings and Cities*, Vol. 2, pp. 220–243, Number: 1.

Masucci, A. Paolo, Elsa Arcaute, Erez Hatna, Kiril Stanilov, and Michael Batty (2015) "On the problem of boundaries and scaling for urban street networks," *Journal of the Royal Society Interface*, Vol. 12, p. 20150763, Number: 111.

Medda, Francesco, Peter Nijkamp, and Piet Rietveld (1998) "Recognition and classification of urban shapes," *Geographical Analysis*, Vol. 30, pp. 304–314, Number: 4.

Miller, Harvey J. and Michael F. Goodchild (2015) "Data-driven geography," *GeoJournal*, Vol. 80, pp. 449–461.

Minkowski, H (1903) "Volumen und Oberfläche," *Mathematische Annalen*, Vol. 57, pp. 447–495.

Montero, Gaetan, Cécile Tannier, and Isabelle Thomas, "Délimitation morphologique des agglomérations urbaines : comparaison de méthodes fractales et non fractales (chap. 5)," in *Analyse et modélisation fractale de la distribution spatiale des implantations humaines, Cécile Tannier (dir.), Encyclopédie Sciences, domaine Géographie et démographie, ISTE éditions.*

Montero, Gaëtan, Cécile Tannier, and Isabelle Thomas (2021) "Delineation of cities based on scaling properties of urban patterns: a comparison of three methods," *International Journal of Geographical Information Science*, Vol. 35, pp. 919–947, Number: 5.

Moudon, Anne (1992) "A Catholic Approach to Organizing What Urban Designers Should Know," *Journal of Planning Literature - J PLAN LIT*, Vol. 6, pp. 331–349.

Moudon, Anne Vernez (1997) "Urban morphology as an emerging interdisciplinary field," *Urabn Morphology*, Vol. 1, pp. 3–10.

Nations, United (2016) "New Urban Agenda."

Nel.Io, O (1996) "Els confins de la ciutat sense confins. Estructura urbana i limits administratius de la ciutat difusa," *en: Castañer, A. ; Falgueras, M. y Vicente, J. : La ciutat difusa i les periferies: experiences de planificació i gestió, Girona, Universitat de Girona*, pp. 55–71.

Newman, Peter W. G. and Jeffrey R. Kenworthy (1989) "Gasoline Consumption and Cities: A Comparison of U.S. Cities with a Global Survey," *Journal of the American Planning Association*, Vol. 55, pp. 24–37.

Nicolau, Rita and Cristina Cavaco (2018) "Automated delimitation of urban areas comprising small-sized towns – Comparison of two methodologies applied to mainland Portugal," *Environment and Planning B: Urban Analytics and City Science*, Vol. 45, pp. 180–201, Number: 1.

de Olde, Clemens and Stijn Oosterlynck (2021) "Taking Implementation Seriously in the Evaluation of Urban Growth Management Strategies: "Safeguarding the Future" of the Antwerp City-Region," *Land*, Vol. 10, p. 159.

Openshaw, S. (1983) "The Modifiable Areal Unit Problem," *Concepts and Techniques in Modern Geography*, Vol. 38, Number: 38.

Parr, John B. (2007) "Spatial Definitions of the City: Four Perspectives," *Urban Studies*, Vol. 44, pp. 381–392, Number: 2.

Pauleit, Stephan and Friedrich Duhme (2000) "Assessing the environmental performance of land cover types for urban planning," *Landscape and Urban Planning*, Vol. 52, pp. 1–20, Number: 1.

Porta, Sergio, Paolo Crucitti, and Vito Latora (2006) "The network analysis of urban streets: A dual approach," *Physica A: Statistical Mechanics and its Applications*, Vol. 369, pp. 853 – 866.

Pruvot, Michel and Christiane Weber-Klein (1984) "Écologie urbaine factorielle comparée : essai méthodologique et application à Strasbourg," *Espace géographique*, Vol. 13, pp. 136–150, Number: 2.

Ramspek, Chava L., Kitty J. Jager, Friedo W. Dekker, Carmine Zoccali, and Merel van Diepen (2021) "External validation of prognostic models: what, why, how, when and where?" *Clinical Kidney Journal*, Vol. 14, pp. 49–58.

Redfearn, Christian L. (2007) "The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area," *Journal of Urban Economics*, Vol. 61, pp. 519–541.

Remy, Jean (2015) *L'espace, un objet central de la sociologie*, Sociétés urbaines et rurales - Poche, Érès edition.

Ren, Zheng (2018) "Tutorials for Extracting Street Nodes from OSM Data and Creating Natural Cities from Street Nodes Using ArcGIS Model, A Workshop on Big Data Analytics: Topological and scaling perspective for better understanding and (re-)making sustainable cities 12 June 2018, Lund, Sweden, in conjunction with AGILE 2018."

Rinkinen, Jenny, Elizabeth Shove, and Mattijs Smits (2021) "Conceptualising urban density, energy demand and social practice," *Buildings and Cities*, Vol. 2, pp. 79–91, Number: 1.

Roca Cladera, J, M Moix Bergadà, and B Arellano Ramos (2012) "El sistema urbano en España," *Scripta Nova. Revista Electrónica de Geografía y Ciencias Sociales*, Vol. 16.

Romero, J (2009) "Geopolítica y Gobierno del territorio en España," *Valencia, Tirant lo Blanch.*

Rozenfeld, Hernán D., Diego Rybski, José S. Andrade, Michael Batty, H. Eugene Stanley, and Hernán A. Makse (2008) "Laws of population growth," *Proceedings of the National Academy of Sciences*, Vol. 105, pp. 18702–18707, Number: 48.

Rozenfeld, Hernán D, Diego Rybski, Xavier Gabaix, and Hernán A Makse (2011) "The Area and Population of Cities: New Insights from a Different Perspective on Cities," *American Economic Review*, Vol. 101, pp. 2205–2225, Number: 5.

Rullan, Onofre (2011) "La regulación del crecimiento urbanístico en el litoral mediterráneo español," *Ciudad y Territorio Estudios Territoriales*, Vol. 43, pp. 279–297.

Rybnikova, Nataliya, Boris Portnov, Igal Charney, and Sviatoslav Rybnikov (2021) "Delineating Functional Urban Areas Using a Multi-Step Analysis of Artificial Light-at-Night Data," *Remote Sensing*, Vol. 13, p. 3714.

Rybski, Diego and Marta C. González (2022) "Cities as complex systems—Collection overview," *PLOS ONE*, Vol. 17, p. e0262964, Publisher: Public Library of Science.

Salom, J (2011) "Procesos territoriales y transformaciones recientes del sistema urbano valenciano," *Scripta Nova. Revista electrónica de geografía y ciencias sociales*, Vol. XV.

Salom, J and J.M Albertos (2010) "Densidad de la red viaria y forma urbana: Delimitación del espacio urbano en ocho aglomeraciones españolas," *en: Feria Toribio J.M. y Albertos Puebla J.M.: La ciudad metropolitana en España: Procesos urbanos en los inicios del siglo XXI, Pamplona, Thomson Reuters*, pp. 49–94.

——— (2014) "Delimitación y caracterización de los nuevos espacios urbanos valencianos," *Boletín de la Asociación de Geógrafos Españoles*, pp. 127–149.

Salom, J and C Zornoza (2015) "La delimitación del espacio urbano de Valencia a través de la red viaria: Análisis de fuentes de información libre y primeros resultados," *en de la Riva, J., Ibarra, P., Montorio, R., Rodrigues, M. (Eds.) 2015 Análisis espacial y representación geográfica: innovación y aplicación : Universidad de Zaragoza-AGE.*, pp. 403–412, OCLC: 1090325339.

Sastry, N, A.R Pebley, and M Zonta (2002) "Neighborhood Definitions and the Spatial Dimension of Daily Life in Los Angeles„" *Technical, UC Los Angeles : California Center for Population Research*, pp. 1–34.

Schirmer, Patrick Michael and Kay W. Axhausen (2015) "A multiscale classification of urban morphology," *Journal of Transport and Land Use*, Vol. 9, pp. 101–130.

Schoning, Per, Marianne Vik Dysterud, and Erik Engelien (1999) "Documents 1999/17 • Statistics Norway, November 1999," p. 19.

Schwarz, Nina (2010) "Urban form revisited—Selecting indicators for characterising European cities," *Landscape and Urban Planning*, Vol. 96, pp. 29–47, Number: 1.

Shen, Guoqiang (2002) "Fractal dimension and fractal growth of urbanized areas," *International Journal of Geographical Information Science*, Vol. 16, pp. 419–437, Number: 5.

Sotomayor-Gómez, Boris and Horacio Samaniego (2020) "City limits in the age of smartphones and urban scaling," *Computers, Environment and Urban Systems*, Vol. 79, p. 101423.

Steel, Wendy, Tooran Alizadeh, and Leila Eslami-Andargoli (2013) "Planning across borders," *Australian Planner*, Vol. 50, pp. 96–102, ISSN: 0729-3682.

Sullivan, William C. and Chun-Yen Chang (2011) "Mental Health and the Built Environment," in Andrew L. Dannenberg, Howard Frumkin, and Richard J. Jackson eds. *Making Healthy Places*, Washington, DC: Island Press/Center for Resource Economics, pp. 106–116.

Sémécurbe, François (2020) "Analyse de la distribution spatiale des implantations humaines: apports et limites d'indicateurs multi-échelles et trans-échelles."

Sémécurbe, François, Cécile Tannier, and Stéphane G. Roux (2019) "Applying two fractal methods to characterise the local and global deviations from scale invariance of built patterns throughout mainland France," *Journal of Geographical Systems*, Vol. 21, pp. 271–293, Number: 2.

Tan, Xingye and Bo Huang (2022) "Identifying Urban Agglomerations in China Based on Density–Density Correlation Functions," *Annals of the American Association of Geographers*, pp. 1–19.

Tannier, Cécile (2017) "Analyse et simulation de la concentration et de la dispersion des implantations humaines de l'échelle micro-locale à l'échelle régionale - Modèles multi-échelles et trans-échelles."

Tannier, Cécile (dir), *Analyse et modélisation fractale de la distribution spatiale des implantations humaines*, Encyclopédie Sciences, domaine Géographie et démographie, ISTE éditions.

Tannier, Cécile and Denise Pumain (2005) "Fractals in urban geography : a theoretical outline and an empirical example," *Cybergeo : European Journal of Geography*.

Tannier, Cécile and Isabelle Thomas (2013) "Defining and characterizing urban boundaries: A fractal analysis of theoretical cities and Belgian cities," *Computers, Environment and Urban Systems*, Vol. 41, pp. 234–248.

Tannier, Cécile, Isabelle Thomas, Gilles Vuidel, and Pierre Frankhauser (2011) "A Fractal Approach to Identifying Urban Boundaries.," *Geographical Analysis*, Vol. 43, pp. 211–227, Number: 2.

Tayyebi, Amin, Bryan Christopher Pijanowski, and Amir Hossein Tayyebi (2011) "An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran," *Landscape and Urban Planning*, Vol. 100, pp. 35–44.

Teller, Jacques (2021) "Regulating urban densification: what factors should be used?" *Buildings and Cities*, Vol. 2(1), pp. 302–317.

Thomas, Isabelle, Arnaud Adam, and Ann Verhetsel (2017) "Migration and commuting interactions fields: a new geography with community detection algorithm?" *Belgeo [En ligne],*, Vol. 4.

Thomas, Isabelle, Camille Cotteels, Jonathan Jones, and Dominique Peeters (2012) "Revisiting the extension of the Brussels urban agglomeration: new methods, new data ... new results?" *Belgeo*, pp. 2–13, Number: 1-2.

Thomas, Isabelle and Pierre Frankhauser (2013) "Fractal Dimensions of the Built-up Footprint: Buildings versus Roads. Fractal Evidence from Antwerp (Belgium)," *Environment and Planning B: Planning and Design*, Vol. 40, pp. 310–329, Number: 2.

Thomas, Isabelle, Pierre Frankhauser, and Dominique Badariotti (2010) "Comparing the fractality of European urban neighbourhoods: do national contexts matter?" *Journal of Geographical Systems*, Vol. 14, pp. 189–208, Number: 2.

Thomas, Isabelle, Pierre Frankhauser, and Christophe Biernacki (2008) "The morphology of built-up landscapes in Wallonia (Belgium): A classification using fractal indices," *Landscape and Urban Planning*, Vol. 84, pp. 99–115, Number: 2.

Thomas, Isabelle, Pierre Frankhauser, and Marie-Laurence De Keersmaecker (2007) "Fractal dimension versus density of built-up surfaces in the periphery of Brussels," *Papers in regional science*, Vol. 86, pp. 287–308, Number: 2.

Thomas, Isabelle, Jonathan Jones, Geoffrey Caruso, and Philippe Gerber (2018) "City delineation in European applications of LUTI models: review and tests," *Transport Reviews*, Vol. 38, pp. 6–32, Number: 1.

Thomas, Isabelle and Y Zenou (1999) "Ségrégation urbaine et discrimination sur le marché du travail: le cas de Bruxelles," in *In M. Catin, J.-Y. Lesueur, & Y. Zenou (Eds.), Emploi, concurrence et concentration spatiales*, paris: economica. edition, pp. 105–127.

Tobler, W. R. (1970) "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, Vol. 46, pp. 234–240, Publisher: [Clark University, Wiley].

Tomás, Lívia, Leila Fonseca, Cláudia Almeida, Fernando Leonardi, and Madalena Pereira (2016) "Urban population estimation based on residential buildings volume using IKONOS-2 images and lidar data," *International Journal of Remote Sensing*, Vol. 37, pp. 1–28, Number: sup1.

Uchiyama, Yuta and Koichiro Mori (2017) "Methods for specifying spatial boundaries of cities in the world: The impacts of delineation methods on city sustainability indices," *Science of The Total Environment*, Vol. 592, pp. 345–356.

Usui, Hiroyuki (2019) "A bottom-up approach for delineating urban areas minimizing the connection cost of built clusters: Comparison with top-down-based densely inhabited districts," *Computers, Environment and Urban Systems*, Vol. 77, p. 101363.

Valenciana, Generalitat (2011) "Conselleria de Medi Ambient, Aigua, Urbanisme i Habitatge, 2011, Objetivo 01. Sistema de ciudades. Estrategia territorial de la Comunitat Valenciana. Valencia. Generalitat Valenciana."

Vallée, Julie, Guillaume Le Roux, and Pierre Chauvin (2016) "Quartiers et effets de quartier. Analyse de la variabilité de la taille des quartiers perçus dans l'agglomération parisienne:," *Annales de géographie*, Vol. 2, pp. 119–142.

Van Criekingen, M, P Cornut, and S Luyten (2007) "Brussels: polycentricity as 'images on the map', not in reality," in *Cattan N. (Ed), Cities and network in Europe. A critical approach of polycentrism*, pp. 105–112.

Van Hecke, E., J-M. Halleux, J-M. Decroly, and B. Merenne-Schoumaker (2009) "Noyaux d'habitat et Régions urbaines dans une Belgique urbanisée," *Monographie 9!: L'Urbanisation. SPF Economie, P.M.E., Classes moyennes et Energie, Bruxelles.*

Vandermotten, Christian (2016) "Adéquations et inadéquations du découpage territorial aux contextes politiques et économiques: Le cas de la Belgique," *EchoGéo*, Number: 35.

Vandermotten, Christian, Ludovic Halbert, Marcel Roelandts, and Pierre Cornut (2008) "European Planning and the Polycentric Consensus: Wishful Thinking?" *Regional Studies*, Vol. 42, pp. 1205–1217, Number: 8.

Vandermotten, Christian and Frank Vermoesen (1995) "Structures sociales comparées de l'espace de trois villes européennes : Paris, Bruxelles, Amsterdam.," *Espace Populations Sociétés*, Vol. 13, pp. 395–404, Publisher: Persée - Portail des revues scientifiques en SHS.

Vanderstraeten, Lieve and Etienne Van Hecke (2019) “Les régions urbaines en Belgique,” *Belgeo*, Number: 1.

Vanneste, Dominique, Isabelle Thomas, and Lieve Vanderstraeten (2008) “The spatial structure(s) of the Belgian housing stock,” *Journal of Housing and the Built Environment*, Vol. 23, pp. 173–198, Number: 3.

Verhetsel, Ann, Isabelle Thomas, and Marjan Beelen (2010) “Commuting in Belgian metropolitan areas: The power of the Alonso-Muth model,” *Journal of Transport and Land Use*, Vol. 2, pp. 109–131, Publisher: Journal of Transport and Land Use.

Wang, Shuliang, Dakui Wang, Caoyuan Li, and Yan Li (2015) “Comment on "Clustering by fast search and find of density peaks",” *arXiv:1501.04267 [cs]*, arXiv: 1501.04267.

Wang, Yizhan, ZhengLiang Hu, and Yanming Zhang (2021) “Delineating the future boundaries of urban development using the FLUS Model: A Case Study of Zhaoyuan City, China,” *IOP Conference Series: Earth and Environmental Science*, Vol. 783, p. 012088.

Weber, C. (2001) “Urban Agglomeration Delimitation Using Remote Sensing,” in *Remote Sensing and Urban Analysis*, , edited by j. donnay, m. barnsley, and p. longley. london: taylor & francis. edition, pp. 145–159.

Williams, Katie, Elizabeth Burton, and Mike Jenks (2000) “Achieving sustainable urban form: an introduction,” *Achieving sustainable urban form*, Vol. 2000, pp. 1–5.

Wu, Bin, Bailang Yu, Qiusheng Wu, Zuoqi Chen, Shenjun Yao, Yan Huang, and Jianping Wu (2018) “An Extended Minimum Spanning Tree method for characterizing local urban patterns,” *International Journal of Geographical Information Science*, Vol. 32, pp. 450–475, Number: 3.

Yu, Bailang, Hongxing Liu, Jianping Wu, Yingjie Hu, and Li Zhang (2010) “Automated derivation of urban building density information using airborne LiDAR data and object-based method,” *Landscape and Urban Planning*, Vol. 98, pp. 210–219, Number: 3-4.

Yu, Bailang, Song Shu, Hongxing Liu, Wei Song, Jianping Wu, Lei Wang, and Zuoqi Chen (2014) “Object-based spatial cluster analysis of urban landscape pattern using nighttime light satellite images: a case study of China,” *International Journal of Geographical Information Science*, Vol. 28, pp. 2328–2355, Number: 11.

Zahn, C.T. (1971) "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters," *IEEE Transactions on Computers*, Vol. C-20, pp. 68–86, Number: 1.

Zhang, Ming and Nishant Kukadia (2005) "Metrics of Urban Form and the Modifiable Areal Unit Problem," *Transportation Research Record: Journal of the Transportation Research Board*, pp. 71–79, Number: 1902.

Zhong, Xu and Matt Duckham (2016) "Characterizing the shapes of noisy, non-uniform, and disconnected point clusters in the plane," *Computers, Environment and Urban Systems*, Vol. 57, pp. 48–58.

Zhou, Qi (2015) "Comparative Study of Approaches to Delineating Built-Up Areas Using Road Network Data: Delineating Built-Up Areas Using Road Network Data," *Transactions in GIS*, Vol. 19, pp. 848–876, Number: 6.

Zhou, Qi and Lei Guo (2018) "Empirical approach to threshold determination for the delineation of built-up areas with road network data," *PLOS ONE*, Vol. 13, p. e0194806, Number: 3.

Zhou, Yi, Mingguang Tu, Shixin Wang, and Wenliang Liu (2018) "A Novel Approach for Identifying Urban Built-Up Area Boundaries Using High-Resolution Remote-Sensing Data Based on the Scale Effect," *ISPRS International Journal of Geo-Information*, Vol. 7, p. 135.

List of Figures

2.1 The four principal components of the PCA based of variables listed in Appendix A.1 17
2.2 Partitioning of the Brussels-Capital Region according to the different methods and data 23

3.1 An example of self-similarity in fractal figure (Sierpinski carpet) 28
3.2 A line, a square and a cube with standard measuring (in red) . 29
3.3 The initiator and the first three stages of construction of a Sierpinski carpet (adaptation from Mandelbrot (1982); Tannier (2017)) 30

4.1 Les vingt-cinq tissus bâtis réels 36
4.2 Les vingt-cinq tissus bâtis théoriques 37
4.3 Caractéristiques des tissus étudiés 39
4.4 Une "Fractale box" à l'itération 5. IFS : $N = 5$ et $\epsilon = \frac{1}{3}$ 40
4.5 Analyses fractales de corrélation à partir des points A de la Figure 4.4 (a) et à partir de tous les points de la Figure 4.4 (b) 40
4.6 Variation de la dimension de boîte D_B estimée en fonction de la zone d'étude considérée 42
4.8 Variation des dimensions de boîte (a) et des dimensions de corrélation (b) selon la taille minimale de la fenêtre de comptage 47
4.9 Variation des dimensions de boîte (a) et de corrélation (b) selon la taille maximale de la fenêtre de comptage 48
4.10 Variation des dimensions de boîte (a) et de corrélation (b) selon le coefficient déterminant la taille des fenêtres intermédiaires de comptage 50
4.11 Variation des dimensions de boîte (a) et de corrélation (b) selon le type de séquence (géométrique ou arithmétique) déterminant la taille des fenêtres intermédiaires de comptage 51

4.12 Analyse fractale de boîte des distributions spatiales des bâtiments du tissu théorique « Dust-1 » : séquence géométrique (a) et arithmétique (b) 51
4.13 Illustration de la fonction border qui permet de prendre en compte les effets de bord de l'analyse de corrélation dans Fractalyse 52
4.14 Variation des dimensions de boîte (a) et de corrélation (b) selon le type d'ajustement (logarithmique ou direct) 54
4.15 Classification des 25 tissus réels et des 25 tissus théoriques : dendrogramme basé sur la distance euclidienne entre les dimensions de boîte (à gauche) et les dimensions de corrélation (à droite) 59
4.16 Relation entre les dimensions de boîte et les dimensions de corrélation des cinquante tissus bâtis 60
4.17 Relation entre le logarithme de la densité bâtie et les dimensions de boîte (gauche) et de corrélation (droite) des cinquante tissus bâtis 61
4.18 Landscape Shape Index des tissus réels et théoriques 61
4.19 Relation entre la différence des dimensions de boîte et de corrélation et l'étendue des cinquante tissus bâtis 62
4.20 Comparaison des tissus d'Ansbach et Santiago 63
4.21 Les huit tissus bâtis pour lesquels la dimension de corrélation est supérieure à la dimension de boîte (différence supérieure ou égale à 0,1) 63

5.1 Methodological process adopted to delineate *Natural Cities*: (a) identification of the street nodes; (b) creation of a *Triangular Irregular Network* (TIN); (c) rank-size distribution of the edges of the TIN according to their length; (d) selection of the edges shorter than the mean; (e) transformation of the selected edges into polygons; (f) selection of the largest clusters (i.e. the *Natural Cities*). 68
5.2 A road network (a) and two variants of the *Natural Cities* method: delineation based on the mean length of the edges of the TIN (b); delineation based on the median length of the edges of the TIN (c). 69
5.3 Effect of the geographical extent of the study area on the delineation of *Natural Cities*. 69
5.4 Covering of a theoretical urban pattern. Source:Tannier (2017) 71
5.5 *MorphoLim*: identification of the dilation threshold marking the limit of the morphological agglomeration (blue points). 72
5.6 Analysis of two theoretical built patterns using *MorphoLim*. The urban core of pattern (a) is denser than the core of pattern (b). 73

5.7 Effect of the geographical extent of the study area on the identification of city boundaries using *MorphoLim*. 74
5.8 Hierarchical Percolation method applied to a theoretical street network. (a) identification of street intersections; (b) fitted logistic function ($x - axis$: size of the dilation buffer; $y - axis$: number of street intersections in the biggest cluster); (c) city delineation. 75
5.9 A polycentric street pattern (two centres) (a), and a double condensation threshold in the logistic growth function of the size of the largest cluster (b). 75
5.10 Morphological agglomeration within a theoretical urban pattern crossed by a large non built area: (a) *Natural Cities*; (b) *MorphoLim*; (c) *Hierarchical Percolation*. 76
5.11 Morphological agglomeration of a theoretical urban pattern where a built ribbon links two urban entities by means of (a) *Natural Cities*, (b) *MorphoLim*, and (c) *Hierarchical Percolation* 77
5.12 Location of the former province of Brabant around the Brussels-Capital Region in Belgium (a). Smoothed densities of building centroids (b) and road intersections (c) in the province of Brabant. Note: the bandwidths correspond to the optimal values proposed by default in the ArcGis software application. . 79
5.13 Morphological agglomeration of Brussels obtained using *Natural Cities* computed on street nodes (a) for the province of Brabant and (b) for the whole of Belgium. 81
5.14 Morphological agglomeration of Brussels obtained using *Natural Cities* on the basis of building centroids: (a) for the province of Brabant and (b) for the whole of Belgium. 82
5.15 Dilation curve and rank-size distribution of built clusters obtained using *MorphoLim* for the province of Brabant (a & b) and for the whole of Belgium (c & d). 83
5.16 Morphological agglomerations of Brussels obtained using *MorphoLim*; (a) computed for the province of Brabant; (b) computed for the whole of Belgium. 84
5.17 Logistic functions and morphological agglomerations of Brussels obtained using *Hierarchical Percolation* on the basis of road intersections; (a) for the province of Brabant; (b) for the whole of Belgium. $x - axis$: width of the dilation buffer; $y - axis$: number of road intersections in the largest cluster; τ: condensation threshold. 86

5.18 Logistic functions and morphological agglomerations of Brussels obtained using *Hierarchical Percolation* on the basis of building centroids; (a) for the province of Brabant; (b) for Belgium. $x - axis$: width of the dilation buffer; $y - axis$: number of building centroids in the largest cluster; τ: condensation threshold. . . . 87

6.1 Les 82 aires urbaines de l'échantillon 98
6.2 Dimension fractale de boite, indice de dendricité et indice de compacité pour deux agglomérations morphologiques théoriques 102
6.3 Courbes de dilatation et points de courbure principale (+) de six aires urbaines françaises 104
6.4 Distribution rang-taille des agrégats bâtis de six aires urbaines morphologiquement bien différentes 105
6.5 Dimensions fractales de boîte des aires urbaines (AU) et des agglomérations morphologiques (AM) pour chaque groupe issu de la classification effectuée sur ces dimensions 110
6.6 Dendrogramme des classifications de 82 aires urbaines françaises au regard des caractéristiques de leur tissu bâti (a) et de leur périmètre (b) 111
6.7 Six tissus bâtis représentatifs de chaque groupe de la classification selon $D_s(AM)$ et $D_s(AU)$: (a) Béthune – groupe 1 ; (b) Belfort – groupe 2 ; (c) Colmar – groupe 3 ; (d) Tours – groupe 4 ; (e) Lorient – groupe 5 ; (f) Troyes – groupe 6 112
6.8 Relations entre le nombre de lignes fermées (fragments) composant le périmètre de chaque agglomération morphologique et les différents indices la caractérisant. 113
6.9 Six périmètres d'agglomérations morphologiques représentatifs de chaque groupe de la classification selon δ, CO et $D_b(MA)$: (a) Reims – groupe 1 ; (b) Nancy – groupe 2 ; (c) Metz – groupe 3 ; (d) Besançon – groupe 4 ; (e) Douai-Lens – groupe 5 ; (f) Brive-la-Gaillarde – groupe 6 115
6.10 Dendrogramme de la classification de 82 aires urbaines françaises au regard des notes factorielles d'une analyse en composantes principales effectuée sur un ensemble de variables caractérisant à la fois les tissus bâtis et les périmètres des villes. 119
6.11 Notes factorielles des agglomérations morphologiques en fonction des groupes de la classification (troisième typologie) 121
6.12 Typologie de 82 aires urbaines françaises selon leur forme. (Analyse en composantes principales suivie d'un classification ascendante hiérarchique sur 9 variables caractérisant les tissus bâtis et les périmètres des villes). 123

7.1 Delimitation of the Metropolitan Area of Valencia 131

7.2 Dilation curve and rank-size distribution of the Valencia urban area at ground level . . . 134
7.3 Morphological clusters of the urban metropolitan area of Valencia (ground level) . . . 135
7.4 Example of main morphological clusters : (a) clusters of the centre of Valencia; (b) north-western axis cluster; (c) clusters of Lliria . . . 135
7.5 Number of buildings by number of floors in Valencia . . . 137
7.6 Rank-size distribution of morphological clusters of buildings of (a) 2 or more floors; (b) 4 or more floors; and (c) 7 or more floors 138
7.7 Morphological agglomeration for buildings of (a) 2 or more floors; (b) 4 or more floors; and (c) 7 or more floors . . . 139

8.1 Different methods to calculate the density of buildings or to perform a spatial clustering of a built-up pattern . . . 145
8.2 Study area (left) and an example of the footprint of buildings (right) . . . 147
8.3 Application of the method on a simple spatial toy structure of buildings . . . 148
8.4 Step I: Flow chart . . . 151
8.5 Distribution of the 95% smallest edges of the Minimum Spanning Tree . . . 153
8.6 Two examples of typical Moran scatterplot (outlier in green) . 154
8.7 Two examples of graphs and related sub-graphs according to the outlier (in green) . . . 154
8.8 Histogram of the coefficients of variation of the edge lengths for all 26,462 subgraphs . . . 155
8.9 Example 1: Built-up footprints (left) and detected clusters (right) of a Belgian village (Ochamps) . . . 156
8.10 Example 2: Built-up footprints (left) and detected clusters (right) of a suburban settlement (Ophain-Bois-Seigneur-Isaac) . 156
8.11 The case of Seneffe . . . 158
8.12 The case of Genappe . . . 158
8.13 Typical cases of topology-based density according to inter-building distances . . . 160

C.1 The two chunkings . . . 182
C.2 Calculation of the spatial lag on a simple spatial toy structure of buildings . . . 182
C.3 Comparing different thresholds in condition (1) (see 8.4) on a given pattern of buildings (a) . . . 183
C.4 Comparing different thresholds in condition (1) (see 8.4) on a given pattern of buildings (a) . . . 184

C.5 Comparing different thresholds in condition (1) (see 8.4) on a given pattern of buildings (a) 185
C.6 The 26 462 clusters following spatial clustering 186
C.7 The topology-based built density index computed for Belgium . 187

List of Tables

2.1 Number of calls between communities 20
2.2 Comparison of the three partitions of Brussels 25

3.1 The dimension for simple shapes 29

4.1 Présentation des tissus théoriques 41
4.2 Récapitulatif des valeurs de paramètres choisies pour les analyses de boîte et de corrélation 54
4.3 Dimension de boîte D_B et de corrélation D_C des cinquante tissus bâtis 55

5.1 Summary of results obtained using the three delineation methods. (*) Mean length of the TIN; (+) distance at which the dilation curve deviates most from a straight line; (.) percolation threshold 88

6.1 Coefficients de corrélation de Pearson entre les différents indices morphologiques caractérisant les 82 aires urbaines étudiées. ns : non significatif à un seuil de significativité égal à 0,05 ; S = superficie des AM ; Dens= densité bâtie des AM ; %B = part des bâtiments de l'AU qui sont dans l'AM ; L = longueur du périmètre. 107
6.2 Les aires urbaines classées selon leur dimension fractale de boîte $D_s(AU)$ et celle de leurs agglomérations morphologiques $D_s(MA)$ 109
6.3 Les aires urbaines classées selon la dendricité, la compacité et la dimension fractale standardisées des périmètres de leur agglomération morphologique. 116
6.4 Saturations de l'analyse en composantes principales (corrélation entre les facteurs principaux et les variables d'entrée) 118

6.5 Aires urbaines françaises classées selon la dendricité, la compacité et la dimension fractale standardisées des périmètres de leur agglomération morphologique. 120

7.1 Main curvature and distance threshold obtained on buildings with different number of floors 138

8.1 Descriptive statistics: length of the edges of the Minimum Spanning Tree 153
8.2 Descriptive statistics: length (in meters) of the removed edges . 155

B.1 Street nodes of the province of Brabant 179
B.2 Street nodes of the entire country (Belgium) 180
B.3 Building centroids of the province of Brabant 180
B.4 Building centroids of the entire country (Belgium) 180