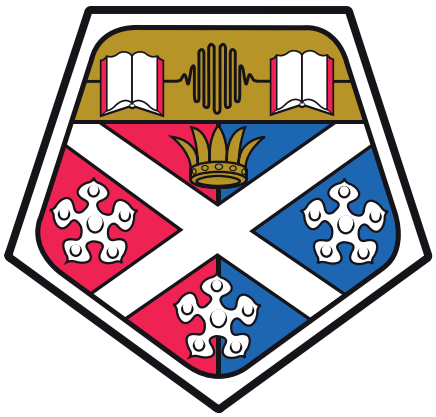
The Urban Atlas: Methodological Foundation of a Morphometric Taxonomy of Urban Form

Martin Fleischmann

A thesis presented for the degree of
Doctor of Philosophy

Supervised by:
Dr Ombretta Romice
Professor Sergio Porta



Urban Design Studies Unit
Department of Architecture
University of Strathclyde, UK
April 2021

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date: 20/04/2021

Abstract

No two cities in the world are alike. Each urban environment is characterised by a unique variety and heterogeneity as a result of its evolution and transformation, reflecting the differences in needs human populations have had over time manifested, in space, by a plethora of urban patterns.

Traditionally, the study of these patterns over time and across space is the domain o urban morphology, a field of research stretching from geography to architecture. Whilst urban morphology has considerably advanced the current understanding of processes of formation, transformation and differentiation of many such patterns, predominantly through qualitative approaches, it has yet to fully take advantage of quantitative approaches and data-driven methods recently made possible by advances in geographic data science and expansion of available mapping products. Although relatively new, these methods hold immense potential in expanding our capacity to identify, characterise and compare urban patterns: these can be rich in terms of information, scalable (applicable to the large scale of extent, regional and national) and replicable, drastically improving the potential of comparative analysis and classification.

Different disciplines with more profound quantitative methods can help in the development of data-driven urban morphology, as now, for the first time, we are in the position where we can rely on a large amount of data on the built environment, unthinkable just a decade ago. This thesis, therefore, aims to link urban morphology and methodologically strong area of quantitative biological systematics, adapting its concepts and methods to the context of built-up fabric. That creates an infrastructure for numerical description of urban form, known as urban morphometrics, and a subsequent classification of urban types.

Conceptually building on the theory of numerical taxonomy, this research progresses the development of urban morphometrics to automate processes of urban form characterisation and classification. Whilst many available methods are characterised by significant limitation in applicability due to difficulties in obtaining necessary data, the proposed method employs only minimal data input - street network and building footprints - and overcomes limitations in the delineation of plots by identifying an alternative spatial unit of analysis, the morphological tessellation, a derivative of Voronoi tessellation partitioning the space based on a composition of building footprints. As tessellation covers the entirety of urban space, its inherent contiguity then constitutes a basis of a relational framework aimed at the comprehensive characterisation of individual elements of urban form and their relationships. Resulting abundant numerical description of all features is further utilised in cluster analysis delineating urban tissue types in an unrestricted urban fabric, shaping an input for hierarchical classification of urban form - a taxonomy.

The proposed method is applied to the historical heterogeneous city of Prague, Czechia and validated using supplementary non-morphological data reflecting the variation of built-up patterns. Furthermore, its cross-cultural and morphological validity and expandability are tested by assessment of Amsterdam, Netherlands and a combination of both cases into a unified taxonomy of their urban patterns. The research is accompanied by a bespoke open-source software *momepy* for quantitative assessment of urban form, providing infrastructure for replicability and further community-led development.

The work builds a basis for morphometric research of urban environment, providing operational tools and frameworks for its application and further development, eventually leading to a coherent taxonomy of urban form.

Keywords: urban morphometrics, taxonomy, classification, measuring, urban form, quantitative analysis, urban morphology, software

Acknowledgements

It would impossible to list everyone who influenced following pages. Let me thank the selection of those whose footprint is the most significant.

First and foremost, I would like to express my gratitude to Ombretta Romice and Sergio Porta for agreeing to embark on this journey with me as my supervisors. Thank you, Ombretta, for guiding me on my first steps through the labyrinth of research. Thank you, Sergio, for your bold ideas shaping my activities.

My PhD would not be possible without Peter Elmlund and Ax:son Johnson Foundation, and the University of Strathclyde providing their contributions towards my funding. Thank you.

Thanks to Alessandra Feliciotti for never-ending peer support and cooperation on the crucial parts of the research. Thanks to David Hořák for entertaining and enlightening discussion of the *nature* of urban form, and to Alessandro Araldi and Alessandro Venerandi for endless discussions on morphometrics. Special thanks belong to Keisuke Sugano and Naoto Nakajima, who taught me to question the presence of universal rules in urban form. Many thanks to Dani Arribas-Bel for giving me enough space in the final stages of writing.

Special thanks go to Joris Van den Bossche and the whole community behind Python ecosystem for geographic data science for their warm welcome and incredible, although mostly indirect support.

Finally, I would like to thank my family and friends who did not mind that I was thousands of kilometres away and supported me in all possible ways. Thank you all.

Previously published work

Fleischmann, M., 2019. Momepy: Urban Morphology Measuring Toolkit. *Journal of Open Source Software*, 4(43), p.1807.

Fleischmann, M. et al., 2020. Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. *Computers, Environment and Urban Systems*, 80, p.101441.

Fleischmann, M., Romice, O. and Porta, S., 2020. Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities, *Environment and Planning B: Urban Analytics and City Science*, 0(0), p. 1-18.

Dal Cin, F., **Fleischmann, M**., Romice, O., Costa, J.P., 2020. Climate Adaptation Plans in the Context of Coastal Settlements: The Case of Portugal. Sustainability 12, 19.

Table of Contents

List of figures

List of tables

Glossary and abbreviations

cell (as tessellation cell)	A developed piece of land associated with one building representing the area of its spatial influence.
character (as taxonomic character)	A characteristics (or feature) of one kind of urban tissue that will distinguish it from another kind.
classification	The ordering of entitites into groups (or classes) on the basis of their similarity.
element	Single geometrical entity representing feature of urban form.
numerical taxonomy	The grouping by numerical methods of taxonomic units into taxa based on their character states. (Sneath and Sokal 1973)
operational taxonomic unit	The lowest ranking taxa employed in a given study.
phenetic relationship	Similarity (resemblance) based on a set of phenotypic characteristics of the objects under study.
phenotype	The appearance of urban tissue, when dealing with characters combinations.
sanctuary area	The portion of the urban tissue enclosed by main streets.
taxon	Taxonomic group or entity (plural taxa).
tessellation-based block	The contiguous portion of land comprised of cells whose centroids are normally bounded by streets or open space.

type	A particular kind, class, or group defined as a conceptual entity. (Bailey 1994)
urban area	The area covered by continuous urban development.
urban form	The physical characteristics that make up built-up areas, including the shape, size, density and configuration of settlements. (Williams 2014)
urban tissue	A distinct area of a settlement in all three dimensions, characterised by a unique combination of streets, blocks/plot series, plots, buildings, structures and materials and usually the result of a distinct process of formation at a particular time or period. (Kropf 2017)
urban tissue type	Morphologically similar group of urban tissues linked by common phenetic characteristics and morphologically distinct from other groups assigned to different taxa.

BIC	**B**ayesian **I**nformation **C**riterion
CAR	**C**overed **A**rea **R**atio
GMM	**G**aussian **M**ixture **M**odel
IQM	**I**nter**Q**uartile **M**ean
IQR	**I**nter**Q**uartile **R**ange
LC	**L**and **C**over
LCZ	**L**ocal **C**limate **Z**ones
LISA	**L**ocal **I**ndex of **S**patial **A**utocorrelation
MAUP	**M**odifiable **A**real **U**nit **P**roblem
MC	**M**orphological **C**ell
MT	**M**orphological **T**essellation
NRMSD	**N**ormalised **R**oot **S**squared **M**ean **D**eviation

OCM	**O**ptimal **C**lassification **M**odel
OTU	**O**perational **T**axonomic **U**nit
POI	**P**oint **o**f **I**nterest
RF	**R**elational analytical **F**ramework
RS	**R**emote **S**ensing
SA	**S**anctuary **A**rea
UM	**U**rban **M**orphology
UMM	**U**rban **M**orpho**m**etrics
UST	**U**rban **S**tructural **T**ype
VC	**V**oronoi **C**ell
VT	**V**oronoi **T**essellation

Any existing, functioning urban area has structure and identity, even if only in weak measure.

Kevin Lynch (1960) p.115

Chapter 1

Introduction

Cities grow, shrink and even disappear. Nevertheless, all inevitably change, reflecting the ever-changing human society. We, people, have built our cities to accommodate our needs of shelter and social interaction. However, those needs were different 50 years ago than they are now and disparate 500 or 5000 years ago. The historical, geographical and societal differences in our needs are imprinted in every small village, town, and metropolis, leaving distinct patterns of development behind. These changes in the way how we design our cities and how their urban form is materialised can be tracked, studied, and can later influence the environment we create for ourselves today.

This thesis aims to contribute to the knowledge of urban morphology, the study of human habitat (Moudon, 1997), by proposing a data-driven method of analysis and classification of urban form able to distinguish the physical imprints of our needs as patterns in the built environment.

1.1 Context of the study

There are two different perspectives when it comes to the study of patterns of development of cities. One tries to capture them to understand their influence on other aspects of life. The other tries to recognise their inner logic and processes of their formation and transformation. Neither of them is new, and both are deeply interconnected.

The study of relation to patterns of urban form is present in a wide range of fields. Economists are interested in the effects of density (Ahlfeldt and Pietrostefani, 2019), a social scientist may look into social mobility (Ewing *et al.*, 2016) sustainability (Bramley and Power, 2009) or include aspects of urban morphology into geodemographic classification models (Alexiou *et al.*, 2016). The role of form is present in energy consumption research (Banister *et al.*, 1997; Ewing and Rong, 2008) or study of biodiversity (Tratalos *et al.*, 2007; Andersson and Colding, 2014). The list could go on. What all have in common is an attempt to understand the consequence of planning decisions and hence influence the future shape of cities. What this perspective needs is a complex characterisation of urban form which does not limit it to one or few particular aspects easy to capture. For that, it needs tools and methods which are universal enough and easy to use and interpret (Boeing, 2020b). Urban morphology, an interdisciplinary study of urban form, focus on such characterisation. However, its tools and methods are not always optimal for the changing needs of today's research.

Urban morphology as a specific field of research was formally established in the early 1960s in the work of MRG Conzen (Conzen, 1960), a geographer, and independently in the work of Saverio Muratori (Muratori, 1959), an architect. The stretch between geography and architecture is typical for urban morphology and forms the core of its interdisciplinarity. Since then, the discipline expanded and proposed different approaches (Oliveira, 2016; Kropf, 2017), some positioned far from the original qualitative works focused on processes and longitudinal aspects (Batty and Longley, 1987; Hillier, 1996; Batty, 1997; Porta *et al.*, 2006). The delineation of patterns of urban form has been studied from various angles, ranging from land use (Caniggia and Maffei, 2001) to the historical origin and geographical location (Conzen, 2004), societal form (Thienel, 2013), building regulations (Forster, 1972) and architectural layout (Beresford, 1971). In recent years, the attempts are more often including computational geography, data science and purely quantitative description of the form (Dibble *et al.*, 2017; Feliciotti *et al.*, 2017; Araldi and Fusco, 2019; Berghauser Pont *et al.*, 2019; Usui and Asami, 2019; Li *et al.*, 2020; Mottelson and Venerandi, 2020; Taubenböck *et al.*, 2020). Such approach, if turned into the systematic and comprehensive method, could react to the needs outside the niche of urban morphology as resulting characterisation could have potential to be adopted by other fields seeking to understand the relation of form and other facets of life.

The rise of data-driven approaches is not coincidental. Current era offers more abundant

geographic data than any other before (Singleton and Arribas-Bel, 2019). Satellite imagery can now bring detailed data on the change of cities at almost real-time in a resolution of 50cm per pixel or less (planet, 2020; Maxar, 2020). Governments and municipalities are increasingly releasing their mapping products under open licenses (UN-Habitat, 2020) and OpenStreetMap, the largest crowdsourced mapping project is enhancing its coverage and quality, making it a reliable source of data for morphological analysis (Barron *et al.*, 2014; Sehra *et al.*, 2020). Data science tools to handle large geospatial datasets (Rocklin, 2015; Yu *et al.*, 2015; Hughes *et al.*, 2015) are readily available, together with general-purpose algorithms helping to make sense of the abundance of data (Pedregosa *et al.*, 2011; Abadi *et al.*, 2016; Paszke *et al.*, 2019). The age of *Big Data* might enable to build better geographical models over space and time (González-Bailón, 2013), but similarly to geography itself (Singleton and Arribas-Bel, 2019), urban morphology needs to bring new methodological tools to increase its relevancy in digital times.

The combination of data abundance, new tools and urban morphology has a potential to deliver detailed analysis on an unprecedented extent as scalable algorithms with a potential to handle *big data* can, in theory, analyse metropolitan and larger areas while keeping information on the granular level. This idea is in the heart of this thesis, and the research presented on the following pages aims to propose steps towards this goal.

1.2 Problem statement

Quantitative (big) data-driven methods are new in urban morphology and far from being matured. While network-based approaches like Space Syntax (Hillier, 1996) or Multiple Centrality Assessment (Porta *et al.*, 2006, 2010) have been around for more than ten years, including tools and wealth of publications, their scope is limited. Recent additions, building on the previous theories, as Multiple Fabric Assessment (Araldi and Fusco, 2019) or street, plot and building types by Berghauser Pont *et al.* (2019) are trying to change the situation and expand the existing scope, but there is a long way towards comprehensiveness able to capture the complexity of urban form.

In particular, a focus on delineation of homogenous patterns of development is scarce. Published literature offers a small number of methods which are data-driven and able

to work on a large scale. If there are such methods, they are limited either in terms of classification detail (e.g. Taubenböck *et al.* (2020)) or granularity (e.g. Jochem *et al.* (2020)). Furthermore, although based on a large sample of features, methods are often based on a small number of variables, limiting their ability to deal with (variable) selection bias and complex nature of urban patterns. Methods which would delineate homogenous areas, systematically classify them and determine the relationship between different types are rare and lack some of the other aspects mentioned above.

Classification of urban form patterns into meaningful, data-driven types is in its infancy. Literature either classify features into predefined types (Lehner and Blaschke, 2019), determine relations between cases (Dibble *et al.*, 2017; Serra *et al.*, 2018) or identify areas without a further interaction (Araldi and Fusco, 2019) between them. The critical aspect which should be studied but it is not to date are relations between automatically recognised urban patterns capturing their similarity, dissimilarity and potentially even phylogenic affinity.

1.3 Aim and scope

This thesis aims to propose a method of derivation of numerical (data-driven) taxonomy of urban form patterns. Numerical taxonomy is a specific type of hierarchical classification which is based on quantitative characterisation of samples, reflecting the relationship between them (see chapter 3 for details). Furthermore, the proposed method should be able to delineate homogenous urban patterns used as samples in the classification in an unsupervised manner, without prior specification of types to minimise the potential bias built in the definition of the types. Optimally, the resulting method will overcome some of the limitations of previous research and provide a comprehensive description of urban form, which will inclusively cover the whole urban fabric instead of predefined case samples. The inclusive taxonomy can be, from a certain perspective, seen as an ultimate classification, one which is able to allocate any urban pattern into a hierarchical structure. Learning from other scientific fields, especially biology, taxonomies are generally accepted as the optimal model of systematisation as they capture not only individual species but also their similarity and relationships. Urban morphology cannot offer such a classification at the moment and existing methods do not aim for delivering one (see the chapter 3 for

details).

The scope of work is limited to a quantitative approach enabling large scale analysis and minimal data input, expanding the applicability of results. This research is purely form-focused to further reduce data requirements, excluding land-use data, points of interest (POI), or any other additional data layers generally used in urban analytics. When describing the built environment, the current abundance of data of various kinds gives us endless opportunities to use the various dataset in the analysis. However, there are two reasons to limit the inputs to the fundamental minimum at this stage of research: 1) as the title suggests, this thesis aims to develop a methodological foundation - a framework on which further research can build. Such a framework should consist only of necessary parts and allow the flexible addition of other components based on the specific needs of future research applications. That said, the inclusion of other data inputs reflecting open spaces, green and blue space or POIs should be considered in later stages and is out of the scope of this work; 2) the framework itself should be applicable across different contexts, esp. regarding varying data availability. If we base the method on a rare dataset representing, for example, the placement and size of trees in streets, the resulting method will be applicable in a handful of cities around the world that can provide such input. However, the goal is to develop a basis that can be applied in the data-rich European context and cities of the Global South with limited cartographic representation.

From the perspective of data sources, it is limited to vector representation of urban form as raster-based earth observation does not yet offer a detailed understanding of morphological elements within remotely sensed data. Nevertheless, it is assumed that earth observation will play a crucial role in morphological research in future as it rapidly evolves and by definition allows the consistent quality of data in any context on the Earth.

The aim and scope drive the background analysis presented in chapters 2, 3 and 4. Explicit research questions and hypothesis (see Chapter 5) are then formulated based on conclusion derived from each of the background chapters.

1.4 Significance of the study

Due to its scope, this research should provide broad applicability, allowing classification of a large number of urban environments while providing their descriptive numerical characterisation. Urban areas around the world are currently covered either by governmental, crowdsources or even private data capturing elements of urban form. As the method results in a classification based on numerical profiling of each recognised urban pattern, it could become an input for the studies analysing the effect of urban form on other aspects of life in cities. This research's role is methodological, aiming to provide a descriptive layer on which other studies can build. For example, to study the effect of urban form on obesity, a researcher typically has to characterise both aspects - urban form and obesity. However, having expertise in both is rare, often leading to the superficial description of one. This research should deliver a thorough evidence-based portrayal of the urban environment, which could be directly embedded in such a study. For that reason, this work is, from a technical perspective, designed as replicable and reproducible research, enabling an easy application by other researchers to other areas.

On a theoretical level, this thesis proposes a comprehensive morphometric description of urban form, including specification of fundamental elements and a framework for their analysis. Furthermore, it revisits the implementation of originally biological concepts of morphometrics and numerical taxonomy in urban morphology, along with the specification of classification units.

On a practical level, it provides a methodological foundation for the construction of an expandable hierarchical classification of urban form. Moreover, it comes with bespoke software tools for quantitative analysis of urban form backing the whole research.

1.5 Overview of the study

The thesis is structured in two major parts, *background* and *core* each with three individual chapters. The structure is graphically represented in a figure 1.1.

Chapter 2 (Existing approaches to classification of the urban form), looks into the theory of classification, introduces different methods of its application followed by the theoretical

proposal of criteria for an optimal classification model of urban form. Furthermore, it provides an overview of approaches to classification of the urban form known in the literature to date and assesses each of them against the aforementioned criteria, allowing the identification of a gap within the existing research.

Chapter 3 (Numerical taxonomy) zooms into detail of one specific classification method - numerical taxonomy. The concept initially developed in biology has been recently introduced into urban morphology (Dibble *et al.*, 2017). However, as chapter 3 shows, aspects of the existing proposal need to be reevaluated before applying it further. Principles of numerical taxonomy are hence introduced and theoretically transposed onto the urban form, specifying the key question which needs to be resolved.

Chapter 4 (Urban morphometrics and its terminological inconsistency) dives into the realm of quantitative characterisation of urban form, providing an overview of nearly all potential measurements literature used to date. That requires to deal with the terminological inconsistency and missing framework for nomenclature and categorisation of measurable characters. Therefore chapter 4 proposes a classification and naming schema and applies it across a wide range of characters, enabling the identification of predominant ways of measuring and potential gaps.

The three background chapters are then synthesised in chapter 5 (Research design statement), which builds hypothesis and research questions on the relevant findings and proposes a framework for reproducibility for the rest of the work. Further, it outlines the selected case studies and links background and core chapters.

Chapter 6 (Morphometric elements of the urban form), the first of the core chapters, provides the basis for morphometric assessment by proposing its fundamental elements, ways of their aggregation and a coherent relational framework binding altogether. The chapter builds on fundamental elements of the urban form known from literature but proposes an implementation of morphological tessellation as a basic spatial unit instead of traditionally used plots. The ability of tessellation to reflect similar phenomena as the plot is then empirically tested together with the various models of location-based aggregation aiming to capture contextual information.

Chapter 7 (Identification of tissue types through urban morphometrics) uses the foundation proposed in chapter 6 and the database of measurable characters from chapter 4

to develop a comprehensive morphometric characterisation of urban form on the level of an individual building. The resulting hyper-dimensional description is used as an input of cluster analysis able to identify distinct types of urban tissues within an unrestricted urban fabric on a case study of Prague, Czechia.

Chapter 8 (Taxonomic relationships of urban tissues) proposes the final methodological step resulting in a hierarchical classification (i.e. taxonomy) of types of urban tissues delineated in the previous chapter. The whole method is then validated using additional variables reflecting historical origin, land use and qualitative classification of urban form. Furthermore, the transferability of the whole method is tested on another case study (Amsterdam, Netherlands) and both cases are then combined to assess the potential of extensibility of proposed numerical taxonomy of urban form.

Final chapter 9 (Synthesis) synthesises the research, discuss its character, potential application, limits and directions of further research.

Furthermore, this thesis contains five appendices, four containing supplementary information to relevant chapters (4, 6, 7, and 8) and one containing Jupyter notebooks allowing reproduction of the major parts of the work. On top of that, two additional annexes contain evidence of dissemination of the work. Annexe 1 consists of the open-source Python package `momepy`, which accompanies the work presented in this thesis, allows its reproducibility and lays a foundation for further morphometric research.

Notice that the structure does not contain an independent chapter dedicated to methods. That is due to the specific design of the core chapters. Each of them has its methodology dependent on the results from the previous chapter.

Chapter 1
Introduction

Background

Chapter 2
Existing approaches to classification of urban form

Chapter 3
Numerical taxonomy

Chapter 4
Urban morphometrics and its terminological inconsistency

Chapter 5
Research design statement

Core

Chapter 6
Morphometric elements of urban form

Chapter 7
Identification of tissue types through urban morphometrics

Chapter 8
Taxonomic relationships of urban tissues

Chapter 9
Synthesis

Figure 1.1: Structure of the thesis and allocation of chapters into parts.

Chapter 2

Existing approaches to classification of urban form

To start the journey towards a taxonomy of urban form, we first need to look at the state of the art of the field. Hence, this chapter aims to understand the classification and its various models from a theoretical perspective and review how are these models applied to urban form. Following the scope of the work defined in the first chapter, the main focus is on quantitative methods of urban morphology based on the vector representation of form. However, some detours to the remote sensing field are necessary. Both to give a full picture of the possible approached and to break the existing firm barriers between the fields.

This chapter is structurally split into two main sections. The first one is outlining the general theory of classification, its principles and different models. That is directly reflected in the specification of a hypothetical, optimal model of classification of urban form, i.e. a set of principles which should a model fulfil to gain a theoretical ability to reflect the complexity of urban form. The second part of the chapter examines models of classification present in literature to this day and compares them to the specified requirements. It first focuses on the spectrum of works ranging from the metropolitan scales to the level of individual buildings (from the perspective of the unit). It then identifies three recurring concepts shared among wide groups of researchers - Urban Structural Type, Land cover classification and Local Climate Zones. The overview leads to the specific gap in the current methodology, as none of the published works fits the optimal model requirements.

In short, the following chapter answers four questions allowing this research to build on the existing knowledge:

1. What is classification?
2. How should it look like?
3. How was it done to date?
4. What is missing?

2.1 Making sense of the world

The world is an inherently complex entity, which needs to be simplified to manageable pieces of information to ensure that the human mind will be able to understand it, i.e. to avoid combinatorial explosion (Fernbach and Sloman, 2017). The natural way of doing so is similar grouping features into chunks and thinking about those. We do not think about individual trees and plants; we consider them all together as a forest. We cluster them based on their functional and geographical relationships into the higher-order entity. That is the essence of classification; we join smaller things into higher-order groups and talk about groups. It is easier for the brain to manage.

2.1.1 Classification as a method

Before we can talk about classification, it is necessary to declutter and define the term itself as it is used in literature in multiple meanings. In an overview of the whole concept and its meanings, Bailey (1994) defines classification as "*the ordering of entities into groups or classes on the basis of their similarity*" (p.1). At the same time, he notes that "*classification is both process and an end result*" (p.2). That can cause terminological confusion as in some situation it might be unclear whether we refer to the resulting systematics or the procedure which generates one. This polysemic nature of the term itself is even more pronounced if we explore the definition offered by the Oxford English Dictionary (OED), which comes with three relevant options:

- "*A systematic distribution, allocation, or arrangement of things in a number of*

distinct classes, according to shared characteristics or perceived or deduced affinities."

- "*The action of classifying or arranging in classes, according to shared characteristics or perceived affinities*"
- "*A category to which something is assigned; a class.*" (Oxford English Dictionary, 2020a)

While the first two coincide with the definition offered by Bailey (1994), the last adds one more option - a class. This research will refer to classification as process and the final result, following Bailey's approach, always trying to specify which meaning is used. The last definition offered by OED will not be used within this research.

Classification systems vary and can be systematised based on the different aspects of the resulting structure and the method used in the process (Bailey, 1994).

The first distinction can be made based on the number of dimensions to *unidimensional* and *multidimensional*. The scope of this work lies in multidimensional classification as a single dimension does not have the explanatory power to describe the complexity of urban form. Dimensions can be both numerical and categorical variables, where unidimensional numerical classification is also known as *binning* or a "*classification scheme*" following Marradi (1990).

Looking at the structure of the classification results, we can talk about *flat* and *hierarchical* models. Flat models generally define all classes as equal and do not specify the relationship between them, either because it could not be specified or because the model does not focus on such aspect. An example is general land use classification, which is flat (e.g. residential use, commercial, industrial) and the relationship between different classes is conceptually complicated to define. Hierarchical models specify, either numerically or conceptually, the relationship between different classes and hence offer a certain level of flexibility of classification as their structure can be interpreted on multiple levels of resolution. That allows the division of elements into two macro classes or multiple micro classes within the same system. A typical example is biological taxonomy of species, where different levels of resolution are represented by domains, kingdoms, families or species. Typical hierarchical classification is, compared to flat models, computationally or conceptually more challenging, but the flexibility which it brings (where applicable) is valuable.

Marradi (1990) uses the term "*typology*" for flat classification models and "*taxonomy* for

hierarchical models. However, Bailey (1994) suggests using the same terms to distinguish between classification models based on different criteria. Methodologically, he recognises two key approaches - one is conceptual, for which he uses the term *typology*, and the other is numerical, named *taxonomy*. Typology is then a conceptual classification, where resulting classes represent concepts, not empirical cases (Bailey, 1994). This means that typologies can be seen as *qualitative classifications* because there is generally no statistics involved. The land-use case mentioned above is a good example of a typology. On the other side is a taxonomy, which is quantitative classification, with classes being empirical entities in the system proposed by Bailey (1994) as well as earlier by Sneath and Sokal (1973).

This confusion in what *typology* and *taxonomy* mean and the criteria used to distinguish between them is present across the literature and is not easy to solve. Furthermore, as with the term classification, *taxonomy* is also used to describe both process and the final result.

Literature knows three terms for quantitative classification based on statistical analysis, to complicate the matter further. One is *numerical taxonomy* (Sneath and Sokal, 1973), which is quantitative, algorithmic classification (more on numerical taxonomy is in section 3.1.2). The other is term *cluster analysis*, describing the process of classification. As Bailey (1994) points out, the methods of numerical taxonomy can be classified as clustering algorithms, making the numerical taxonomy (in the sense of a process) and cluster analysis "*virtually synonymous*" (p.7). The machine learning area, to make things more complex, uses term *unsupervised classification* for cluster analysis, but these two (together with *clustering*) are equal.

Within this research, the term *taxonomy* will be predominantly used to describe the final result of hierarchical quantitative classification, the usage which fits in the definition of both Bailey (1994) and Marradi (1990) but can be used for general hierarchical models if needed. *Cluster analysis* and *clustering* may be used interchangeably and describe the process. The term *typology* will be used within the relevant context to avoid potential disambiguation and its application will generally follow Marradi (1990) rather than Bailey (1994). Unsupervised classification is left only for occasions where it is needed. [1]

[1]Since we are talking about unsupervised classification, it is worth noting that supervised classification is a bit different concept and it is technically rather labelling or classing. It is a tool to sort features into pre-defined categories, unlike all above, which encompass the determination of categories as well. Again,

Bailey (1994) has summarised the reasons why is classification useful in ten fundamental advantages, which are all transferable to urban morphology and classification of the spatial structure of cities. In short, a classification is a tool for *description* (1), giving an overview of all classes within data based on the same criteria. The classified data has *reduced complexity* (2) to a manageable extent. We cannot deal with all individual animals on Earth, but we can work with their taxonomy, significantly reducing the amount of information to work with. Classification can be used to identify *similarities* (3) and *differences* (4) among cases. Identification of similarities allows us to treat all individuals of a single class of similar classes equally (e.g. *beware of snakes* or *go shopping to commercial district*). On the other hand, we can distinguish subtle differences between similar entities (e.g. the difference between venomous and non-venomous snakes can be helpful). A classification, if done properly, is defined by *an exhaustive list of dimensions* (5) on which different classes are based. In such a case, the resulting classification can be comprehensive, while capturing the relationships between classes and dimensions, which is useful for further analysis and profiling of classes. That allows quick, straightforward *comparison* (6) of classes from different parts of the classification structure. Complete list of classes can serve as the *inventory* (7) for management purposes and allows the study of *relationships* (8) among dimensions, relative to the structure of the classification. Moreover, classes can be used as *criteria for measurement* (9), where one class is the criterion, and other are assessed according to similarity with the criterion (e.g. *how close is snake A to a python?* or *how similar is neighbourhood A to Manhattan?*). Finally, the classification may be *versatile* (10), as it can represent both individual units under scrutiny and their location within property space, but also describes the whole sample of units. (pp.12 - 14).

2.1.2 Optimal classification model

Classification model can follow a multitude of pathways and can be defined based on various aspects for a plethora of purposes. However, the aims and scope of each research should drive the classification. Since this research aims to develop a data-driven method, its methodological nature should be a quantitative. The *optimal classification model (OCM)* for this research can be then defined based on the following set of rules derived

this is a terminological issue, but because it is not used in this work, we can leave it out.

from definitions presented in the previous sections. Each of the existing models presented in literature and outlined in the next section is then related to OCM. Furthermore, the specification of the model used within this study should directly reflect the rules in its fundamental design.

The optimal classification model of the urban form should be:

1. Exhaustive
2. Mutually exclusive
3. Empirical
4. Hierarchical
5. Comprehensive
6. Detailed
7. Scalable

An *exhaustive* model covers all entities within the set, meaning that there should not be unlabelled cases in the resulting classification. While the specification of levels of uncertainty in the labelling is welcome, it should not prevail. *Mutual exclusivity* ensures that no entity is at the same time member of more than one class to minimise ambiguities. However, probabilistic clustering specifies the probability of each feature to be a member of each class, but the resulting classification typically uses a single threshold value. *Empirical* nature ensures the data-driven nature of the classification limiting the potential bias in the derivation of *concepts* and dependency on the expert knowledge to assess each entity. Structurally, the model should be *hierarchical* to allow flexibility of its reading, unavailable for flat options. With hierarchical relations between groups, the information encoded in the classification becomes adaptable and provides more straightforward interpretation. *Comprehensiveness* entails the number of dimensions, or descriptors used to cluster entities. The selection of dimensions can be biased and negatively influence the resulting taxonomy. By implementing the large number of dimensions (trying to be as inclusive as possible), such a possibility can be effectively minimised (Sneath and Sokal, 1973). The classification of the urban form should be *detailed* in terms of spatial granularity, meaning that labels should be assigned to individual plots or buildings rather than districts or cities. Finally, the model should be *scalable*. That is both a technical and conceptual requirement, which should ensure that the same model can be used to classify small town and large metropolitan areas.

A classification model which would adhere to these seven principles is currently not available in the published literature. However, the spectrum of existing approaches is vast, and all the points have been addressed in various works. However, not at the same time.

2.2 Spectrum of classification models

There are two main branches of science focusing on urban form classification, which are surprisingly separated from each other - remote sensing (RS) and urban morphology (UM). The diffusion between them is minimal, even though both may conceptually focus on similar questions. Urban morphologists tend to prefer to work with the elements of the urban form directly and build the classification from the constituents parts of urban form upwards (as in Dibble *et al.* (2017)). They need to identify buildings, plots or blocks first, identify the structural features, and then describe and classify them. On the other hand, remote sensing research usually does not dwell into the detail of individual elements, but instead tries to capture the whole pattern directly (Taubenböck *et al.*, 2020). The other approach is to extract the elements first, as in Dogrusoz and Aksoy (2007), but then the methodology is essentially two-step, starting with remote sensing in the extraction part and finishing with morphological analysis of extracted elements in the second. Even though both fields approach the topic differently, there are some similar concepts, although currently unlinked.

Following sections provide an overview of the quantitative trajectory classification of urban form followed in last decades and of recurring concepts stemming from the field of remote sensing. The overview does not include concepts derived from traditional qualitative urban morphology as it is not within the scope of this research.

2.2.1 The trajectory of systematic quantitative classification

The literature on quantitative urban morphology offers a wide selection of proposed methods of classification of urban form, ranging from flat typologies to hierarchical taxonomies and from city to buildings scales (as per the unit of analysis). However, if we want to

focus on the methods following similar aims as this research, we find that there are not many methods, which would come close to OCM. This section presents a trajectory of relevant models, focusing here mostly on urban morphology rather than RS, between the years 2007 and 2020. The earlier work (Barnsley and Barr, 1996; Herold *et al.*, 2002) is seen as too distant from the current research, likely due to the scarce data and tools availability.

The first relevant method proposed by Dogrusoz and Aksoy (2007) attempts to classify urban form based on satellite imagery using automatic extraction of building footprints followed by a cluster analysis of the structure of minimum spanning tree between buildings. The resulting classification can determine whether neighbourhoods tend to be *organized* or *unorganized* as illustrated in figure 2.1. As such, it makes the first step towards the detection of different patterns, which is naturally relatively distant from OCM in some of the criteria.

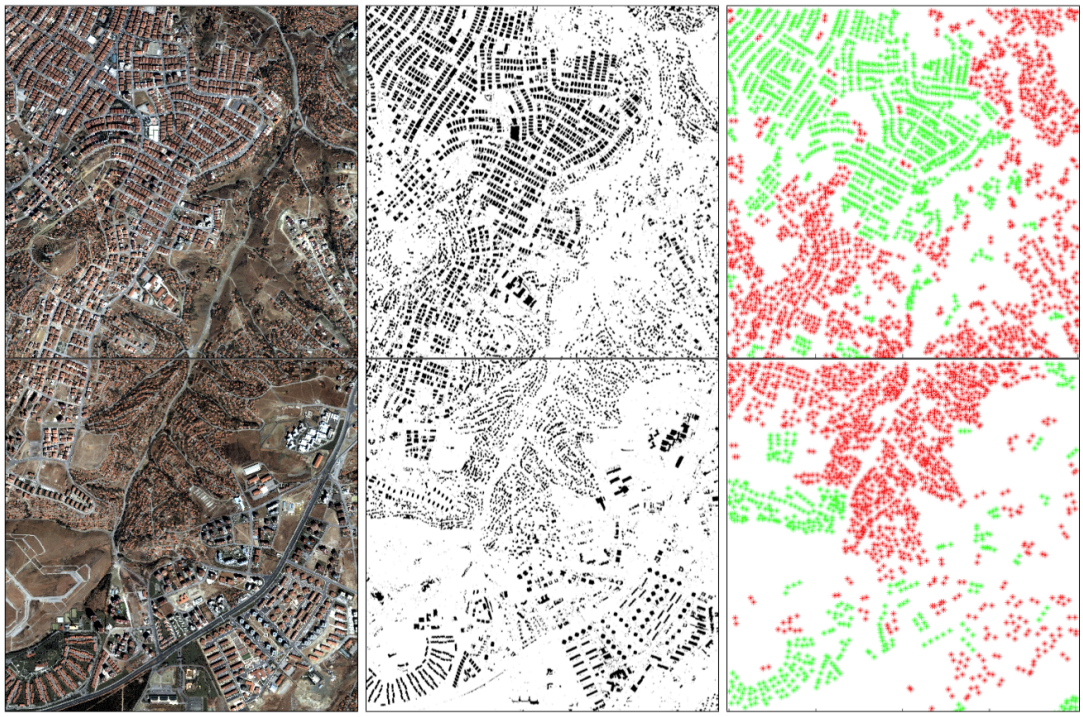


Figure 2.1: Three steps of classification from satellite imagery, through detected building footprints to resulting classes. (Dogrusoz and Aksoy, 2007, figure 4)

Song and Knaap (2007) are proposing a classification of *neighbourhoods* defined as a 1/4 mile (approx. 400m) buffer around selected sites of new single-family development. Using

21 characters, factor analysis and cluster analysis (K-means), they propose flat model of six clusters. By classification of 6788 homes into neighbourhood types, they illustrate the scalability of the model and potential for detailed assessment, as each site has its own neighbourhood defined in a location-based manner (i.e. neighbourhoods are defined independently on each other and can overlap). However, 21 characters and focus on single-family housing only does not ensure the comprehensiveness of the model.

Classification method presented by Steiniger *et al.* (2008) uses a predefined conceptual typology of buildings and uses morphometric assessment to predict its classes. In this sense, the authors provide the first assessment of the validity of morphometrics in the context of classification of urban form, but the typology itself is not defined empirically. Of a similar nature is research presented by Neidhart and Sester (2004), Wurm *et al.* (2016) and Hartmann *et al.* (2016). Even though based on a different set of indicators and data inputs, these three works illustrate the progress in labelling over the years.

Spacemate diagram by Berghauser Pont and Haupt (2010) identifies six building types based on three measurable dimensions - Floor Space Index, Ground Space index and Open Space Ratio (figure 2.2). They propose the same method to be applied on the scale of the building and the scale of fabric. The method has the potential to be scalable and detailed, but not comprehensive as it is based on three characters only. Moreover, the classification, even though based on empirical values, is rather conceptual than data-driven as it is based on nine pre-defined archetypes.

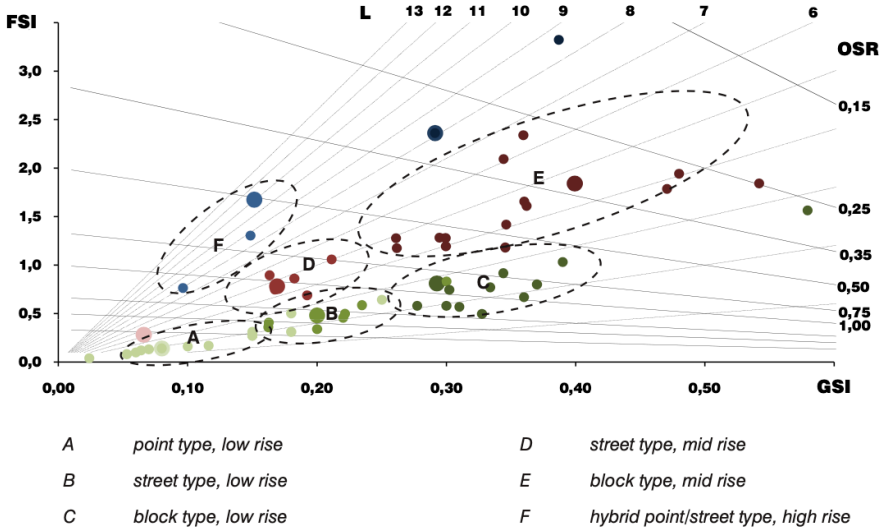


Figure 2.2: Spacemate diagram and building types on the scale of fabric. (Berghauser Pont and Haupt, 2010, figure 23)

Gil *et al.* (2012) characterise streets and block using 25 quantitative characters and cluster them based on K-means into six groups of blocks and four groups of streets. The case study area is covering two neighbourhoods of different origin, and results indicate the potential of cluster analysis based on morphometric values in urban morphology. The method is one of the first which use historical origin as a method of validation of clustering and which results reflect expected distinction (figure 2.3).

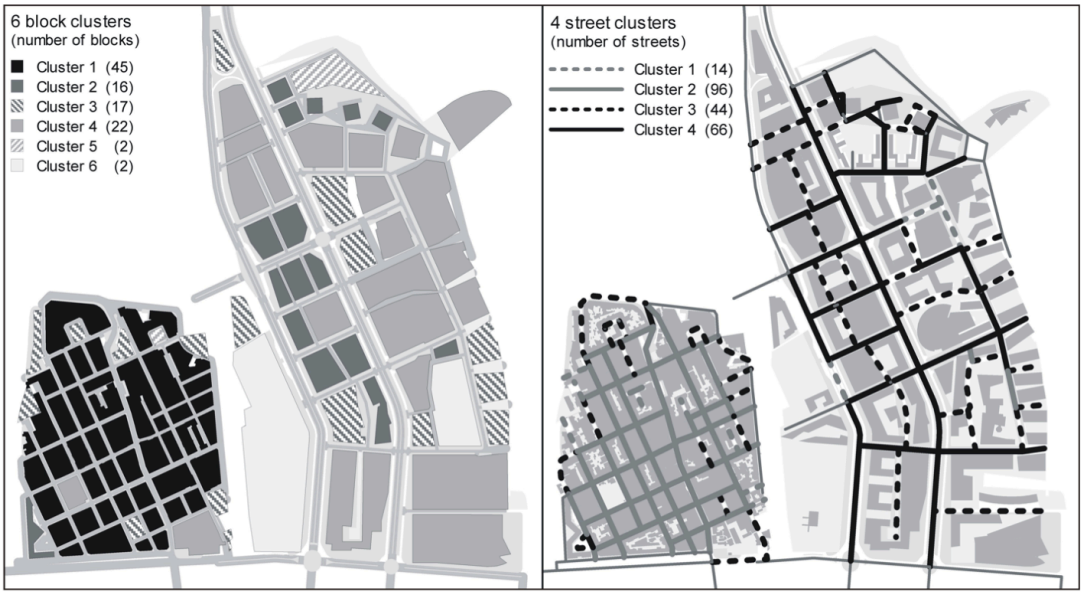


Figure 2.3: Geographical distribution of block clusters and street clusters in two studied neighbourhoods. (Gil et al., 2012, figure 4)

Louf and Barthelemy (2014) propose hierarchical taxonomy of 131 cities based on their street network patterns, with the actual characterisation based on block area and shape. The resulting dendrogram illustrating the classification is shown in figure 2.4. The large set of cases illustrate the scalability of research, but that comes at the cost of granularity and comprehensiveness.

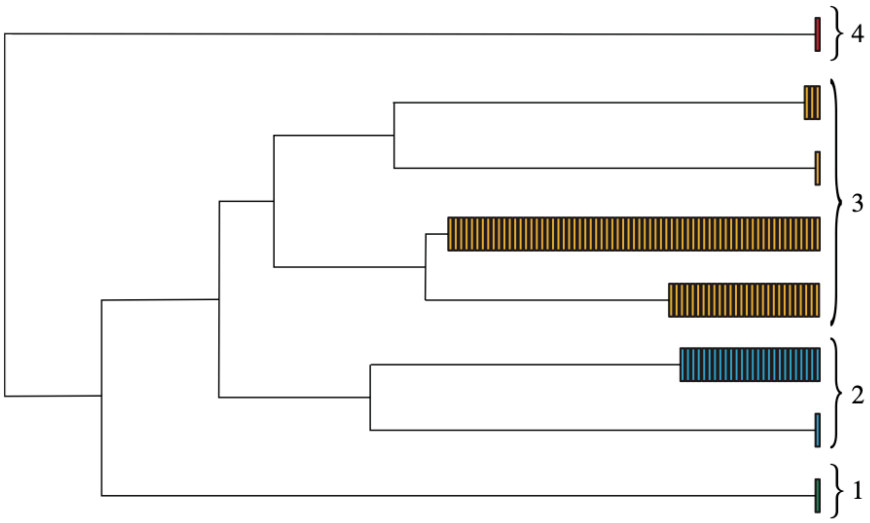


Figure 2.4: Dendrogram representing the structure of classification of cities by Louf and Barthelemy (2014, figure 4). Each bar represents a single case.

Schirmer and Axhausen (2015) proposes classification on multiple levels, where the top one is a municipality, even though the unit is a building defined by centrality and accessibility characters. While their proposal is for a "multiscale typology", methodologically purpose four flat classification models and do not relate one to the other. Their resulting "municipal typology" is illustrated on figure 2.5 below. While all scales combined may be based on comprehensive information, the proposed model is not.

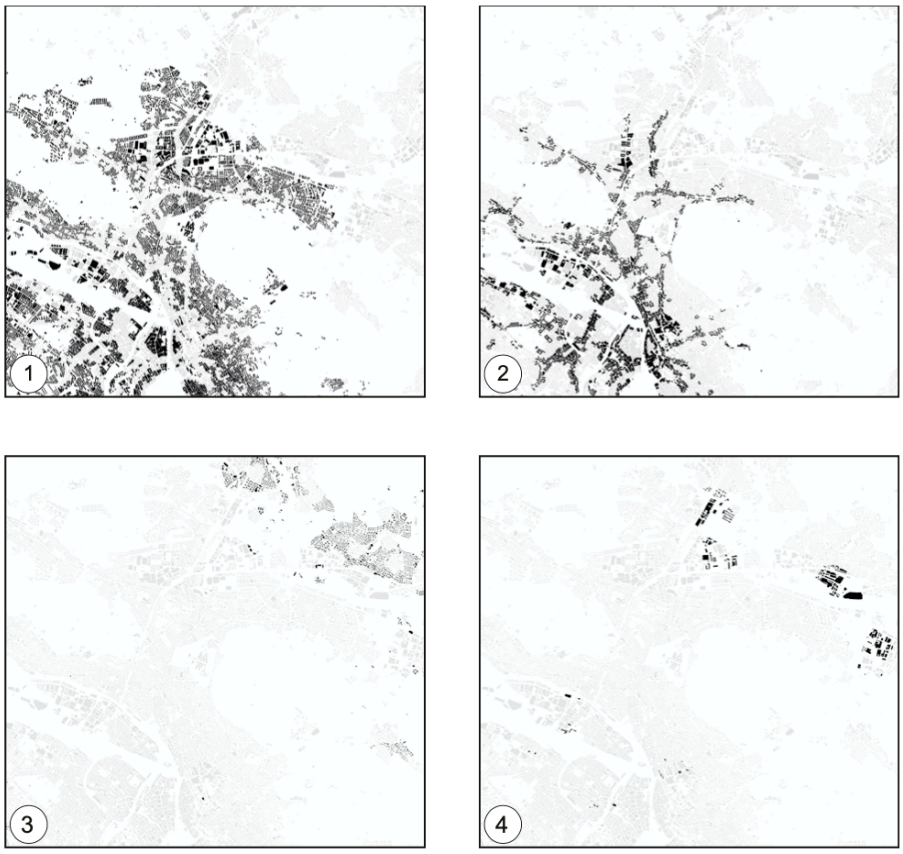


Figure 2.5: Four classes of Schirmer and Axhausen's (2015, figure 11) municipal-level classification mapped in the area of Zurich.

The scale of neighbourhood Schirmer and Axhausen (2015) shares with Serra *et al.* (2018), who are classifying neighbourhoods defined as "*circular areas of 1km radius*" (p.65) characterised by 12 morphological indicators derived from street network, blocks and buildings. Resulting classification is hierarchical taxonomy of selected neighbourhoods, represented by a dendrogram on figure 2.6. In its current form classification is not exhaustive as it covers only pre-defined, yet overlapping neighbourhoods (figure 2.7), and it is not known how would it scale to the continuous classification of whole areas.

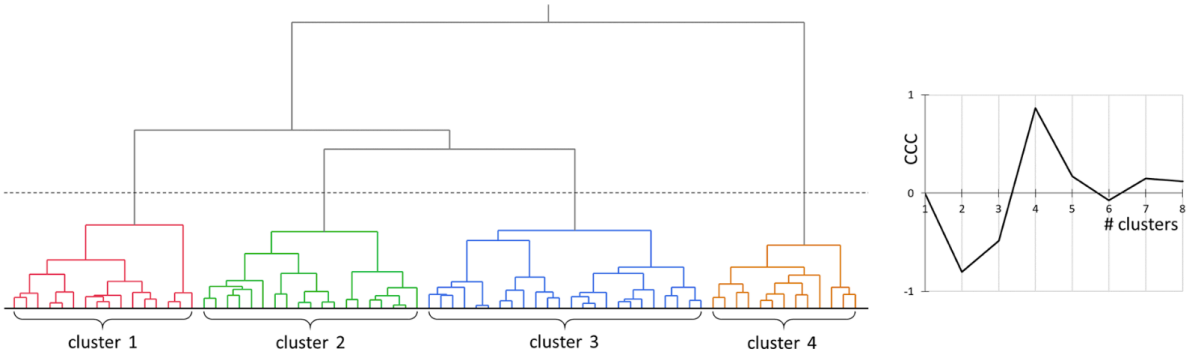


Figure 2.6: Hierarchical classification of neighbourhoods proposed by Serra et al. (2018).

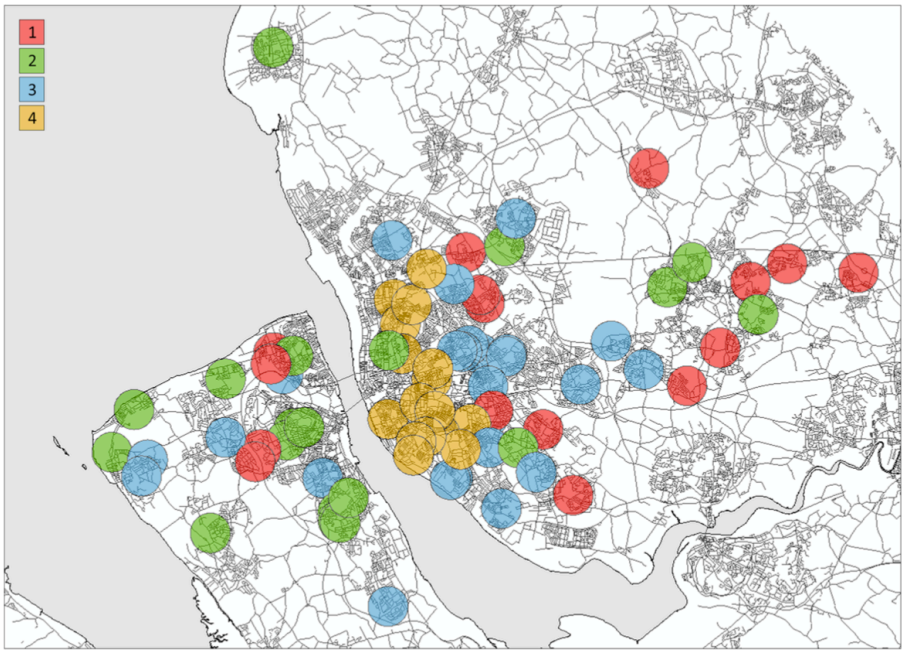


Figure 2.7: Geographical distribution of clusters of neighbourhoods proposed by Serra et al. (2018).

Similar neighbourhood scale is used by Dibble *et al.* (2017), where the unit of classification is Sanctuary Area (Mehaffy *et al.*, 2010). The resulting classification is a hierarchical taxonomy (figure 2.8) based on the comprehensive set of morphometric characters. However, due to the selection of the basic unit, it is not detailed and exhaustive. The method itself is time-consuming (Dibble, 2016), and its proposed form is not scalable. However, the work of Dibble *et al.* (2017) is building foundations of the science of urban morphometrics and will be further examined in section 3.2.2.

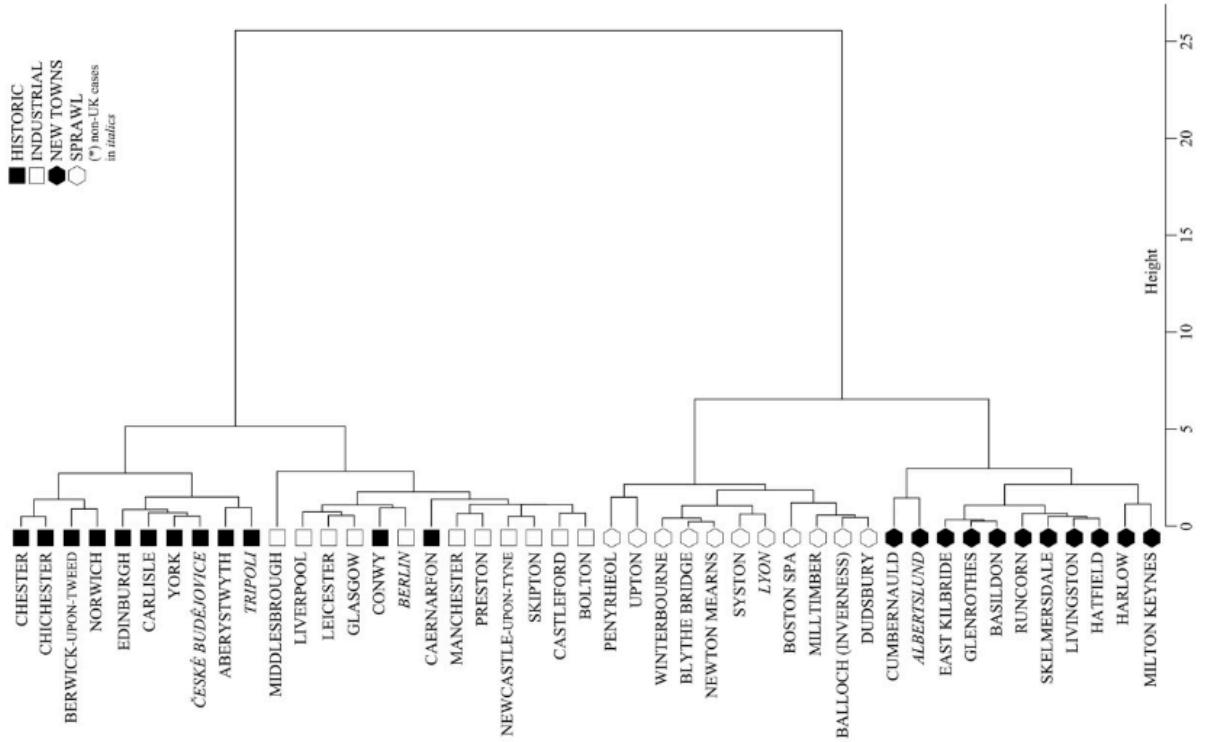


Figure 2.8: Resulting dendrogram illustrating the classification of Sanctuary Areas by Dibble et al. (2017, figure 6, rotated). Notice the consistency of morphometric classification and historical origin of individual cases.

The work on urban typologies presented by scholars at Chalmers University in a series of recent publications (Berghauser Pont and Olsson, 2017; Berghauser Pont *et al.*, 2019; Bobkova *et al.*, 2019) proposes to use three individual typologies of morphological elements: plots, streets and buildings. Each typology is defined through a handful of morphometric characters and cluster analysis, thus making the outputs influenced by this particular selection. Compared to the optimal criteria above, their model is not hierarchical and, notably, not comprehensive (due to the limited number of morphometric characters it uses).

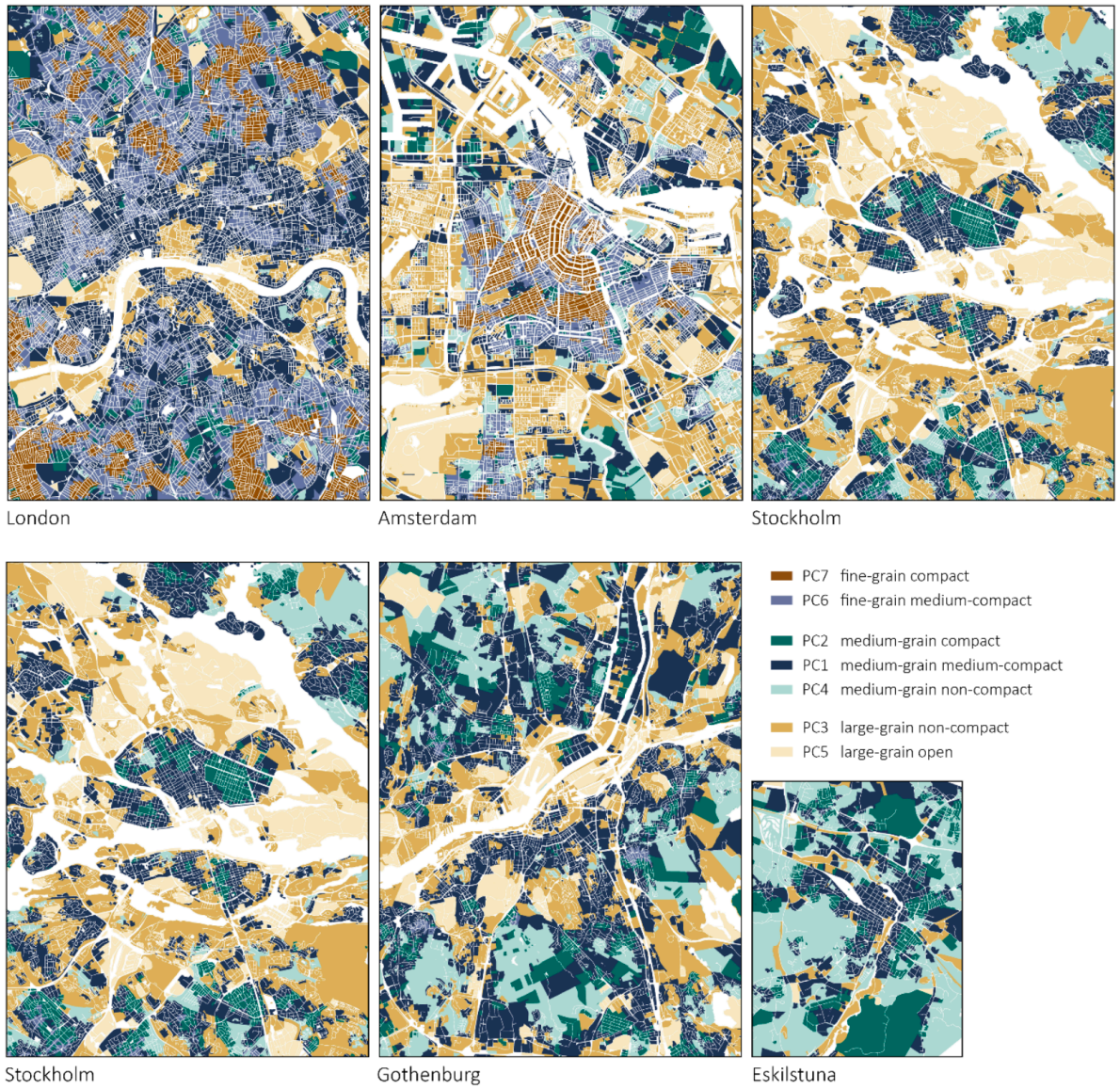


Figure 2.9: Spatial distribution of plot types in selected case studies developed by Berghauser Pont et al. (2019, supplementary material figure 10).

The work of Araldi and Fusco (2019) proposes a classification of street segments from the pedestrian point of view, based on 20+ morphometric characters derived from street networks, building footprints and digital terrain model. The model is powerful in terms of top-level classification of urban form, however, similarly to the Chalmers', it is not hierarchical (the relationship between the types is unknown) and still far from comprehensive (compared, e.g. to others which use a more significant number of characters such

as Dibble *et al.* (2017) with 207). The selection of street as the smallest unit is also a limitation as it assumes homogeneity of the urban form along both sides of the whole segment, which is rarely the case in urban contexts of almost all periods of development.

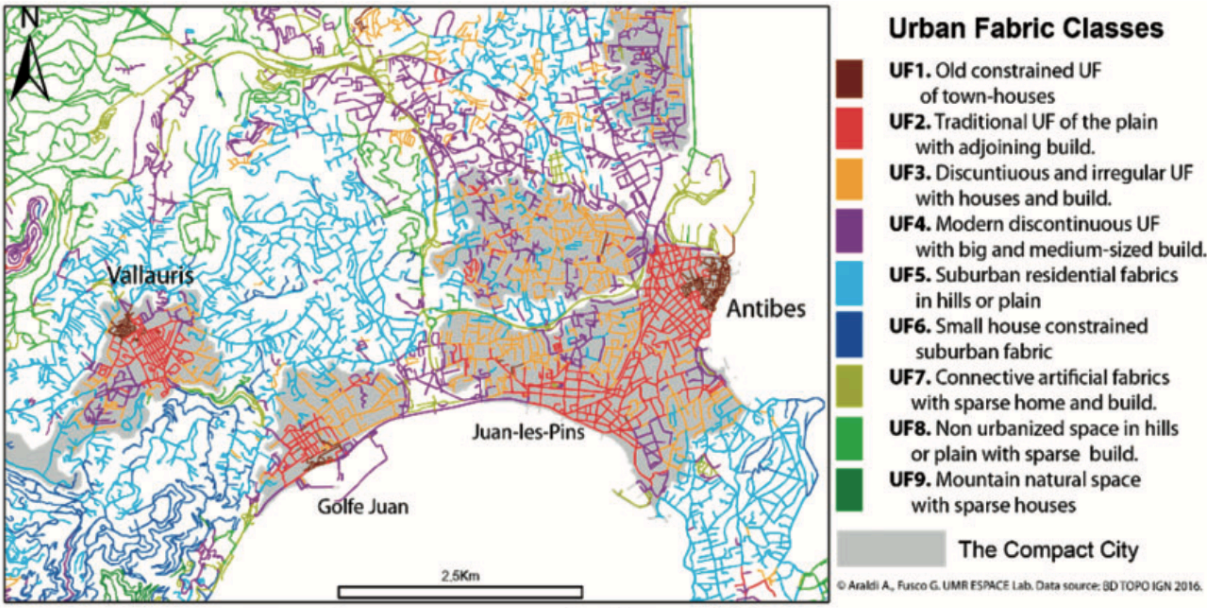


Figure 2.10: Urban fabric classes as a result of the MFA procedure by Araldi and Fusco (2019, figure 2).

Dong *et al.* (2019) proposes a classification of blocks into hierarchical taxonomy using convolutional autoencoder (CAE). The method rasterises the vector representation of block footprints to 64x64 pixels and uses a neural network for image recognition to cluster them. Due to the necessity to keep data for autoencoder of a similar size, oversized and undersized cases were excluded, drawing the method not entirely exhaustive. Resulting hierarchical clustering (figure 2.11) identifies 16 clusters, but in a way which includes all cases, leaving some (approx 40%) unclassified. However, the application of CAE is quite unique in the context of the rest of the field.

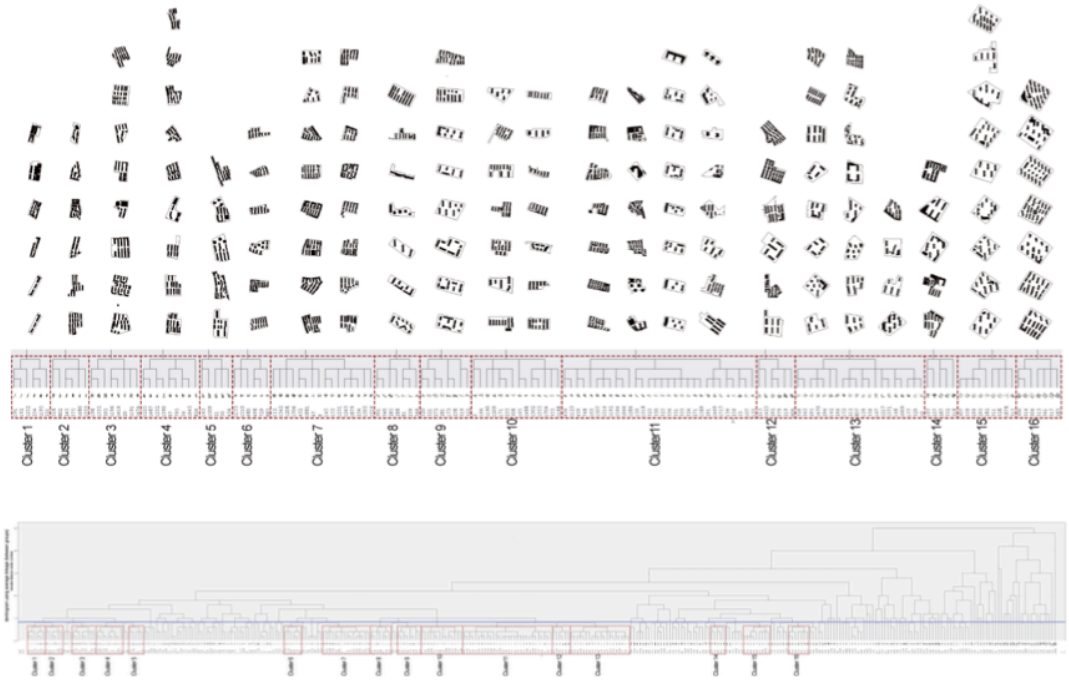


Figure 2.11: Hierarchical clustering of sample of blocks and illustration of different types (rotated). (Dong et al., 2019, figure 6)

Li *et al.* (2020) focus on classification of 83 blocks into hierarchical taxonomy on the basis of 11 indicators (figure 2.12). On top of hierarchical clustering, authors also do K-means analysis, resulting in 5 types, although the relation between K-means clusters and hierarchal one seems to be left unexplored. The sample of blocks covers only a single small case study area, leaving the question of scalability unresolved.

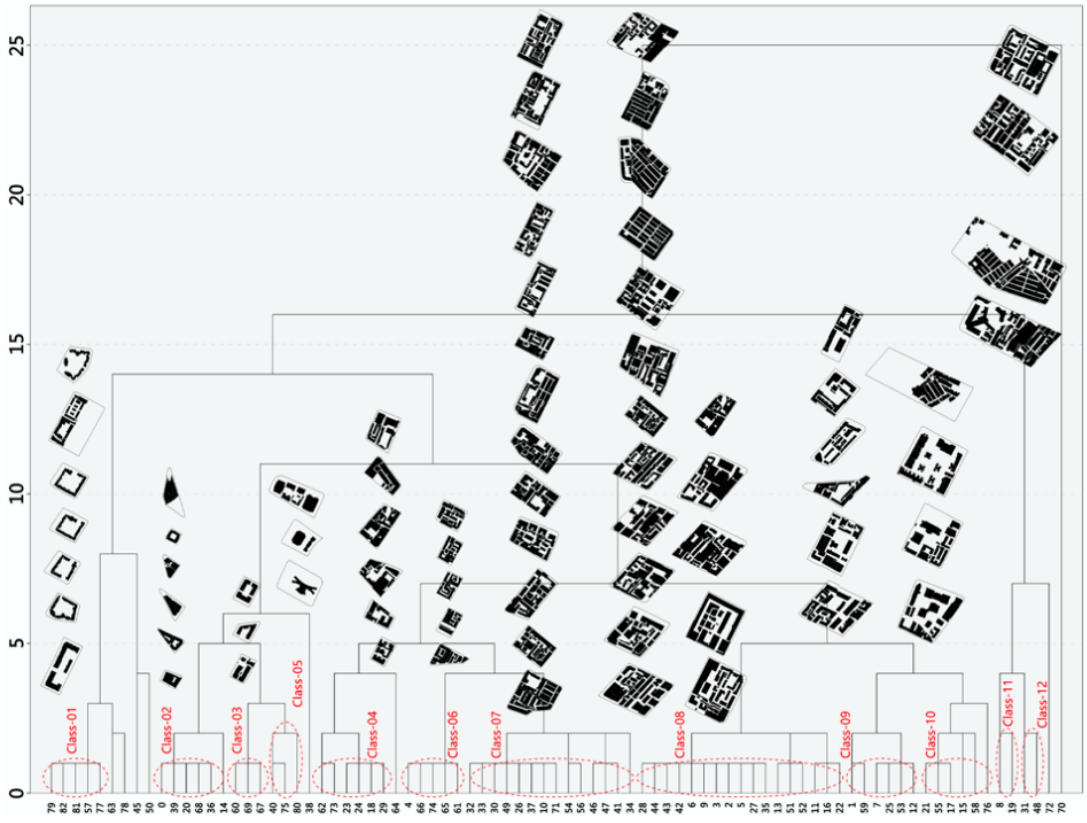


Figure 2.12: Hierarchical clustering of selected blocks and illustration of different types (rotated). (Li et al., 2020, figure 5)

The method proposed by Jochem *et al.* (2020) works on 100m grid and is based on building footprint data and 7 characters measured per grid cell. That resulted in 12 types (figure 2.13) in the first and 5 in the consequent step. The method dependent on the arbitrary grid does not follow the natural composition of urban form and in some case might be unable to recognise certain linear patterns. Although scalable, as presented in the paper, the method is likely not comprehensive enough due to the small number of measurable character influencing the resulting classification.

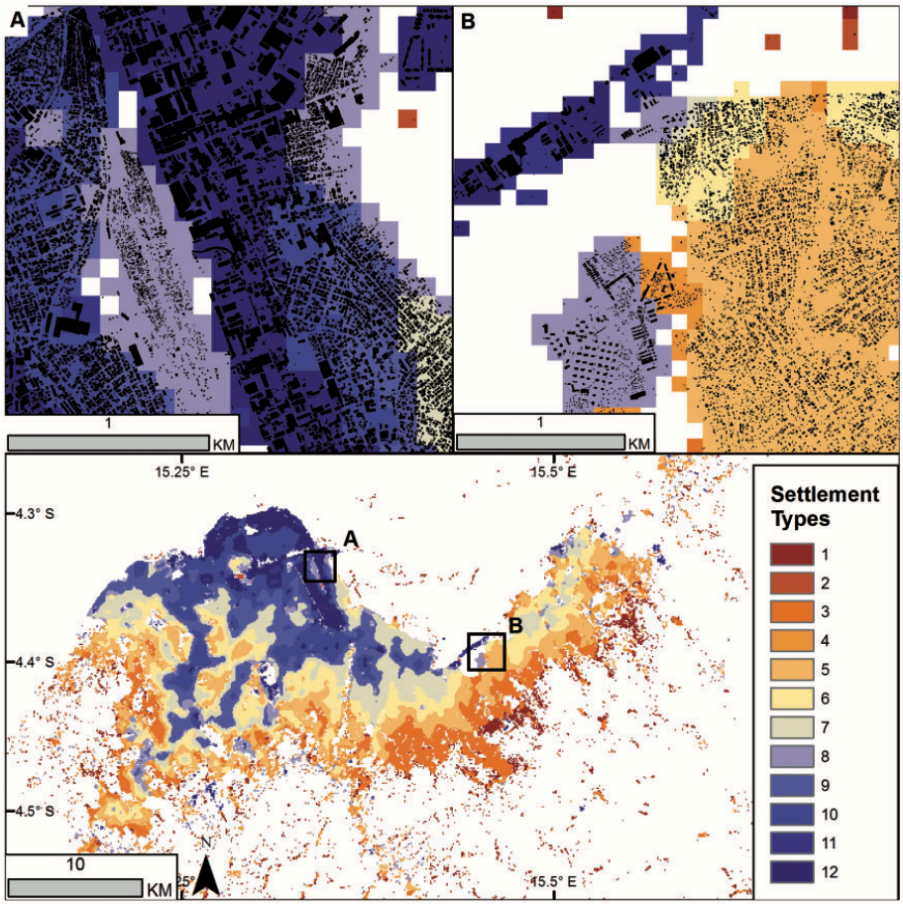


Figure 2.13: Settlement types prediction in Kinshasa, Congo in the resolution of 12 types based on 100m grid. (Jochem et al., 2020, figure 3)

2.2.2 Recurring concepts

The trajectory above presents instead unlinked methods and concepts as urban morphology did not agree on a specific approach yet. Likely due to the relative sub-optimality of all presented above. In remote sensing, partially overlapping with urban morphology, are identifiable three recurring partially linked concepts in the literature - Urban Structural Types (UST), Land Cover (LC) classification and Local Climate Zone (LCZ) classification.

The concept of UST (and related Urban Structural Unit) has been first developed for planning purposes in the 1960s (Lehner and Blaschke, 2019), alongside the UM concepts of the morphological region (Conzen, 1960; Oliveira and Yaygin, 2020) or *tessuta urbana*

(Caniggia and Maffei, 1979), with which it shares the core of the definition based on the internal homogeneity. In the RS, UST started to appear since the early 1990s (Lehner and Blaschke, 2019) and become quickly popular. However, the methods and terms are still remaining inconsistent. Defining UST is not an easy task as the literature is not consistent in the definition and a range of works reviewed by Lehner and Blaschke (2019) agree on principles, but not on specific "definition, description and derivation" (Lehner and Blaschke, 2019, p. 7). Generally speaking, UST can be defined as a unit of a "*specific spatial characteristics, e.g., the morphology and the spatial relationships between urban artefacts such as buildings, streets, trees, lawns*" (Lehner and Blaschke, 2019, p. 2). The practical translation of this, relatively vague description, varies. Some UST classifications are flat while others are hierarchical, work on various scales and are mostly conceptual typologies, with only a few examples of data-driven models (Lehner and Blaschke, 2019). In relation to the OCM requirements, it is complicated to assess the concept as a whole due to its internal inconsistency, but no method found in literature fulfils all the criteria.

Land cover is a related but more straightforward concept coming from Remote Sensing area. Unlike all the other, it does not focus purely on the urban environment but aims to classify all areas together. As such, its detail when it comes to urban form is generally low, with CORINE Land Cover classification (European Environment Agency, 1990) dividing urban form into continuous and discontinuous, plus specialist (industrial/commercial, ports, airports). Copernicus Urban Atlas refines Corine by adding density on top of continuity as a second criterion. One of the most refined is the work of Pauleit and Duhme (2000) which distinguish 10 types of urban form. However, due to the nature of land cover classification, which is usually done as a supervised classification (i.e. labelling), it is usually a conceptual method (which can be both flat and hierarchical).

Classification into Local Climate Zones (Stewart and Oke, 2012) is a mix of both urban structure and land cover into a singular conceptual typology of 10 types of the built form and 7 land cover types (figure 2.14). It is intentionally very generic to produce inclusive classification, covering all possible types of urban development. It is a flat model with classes defined numerically, but still capturing rather conceptual divisions. In RS, LCZ are used on a large scale (Taubenböck *et al.*, 2020) to characterise cites across the world, but the classification itself is still relatively limited by design and very top-level.

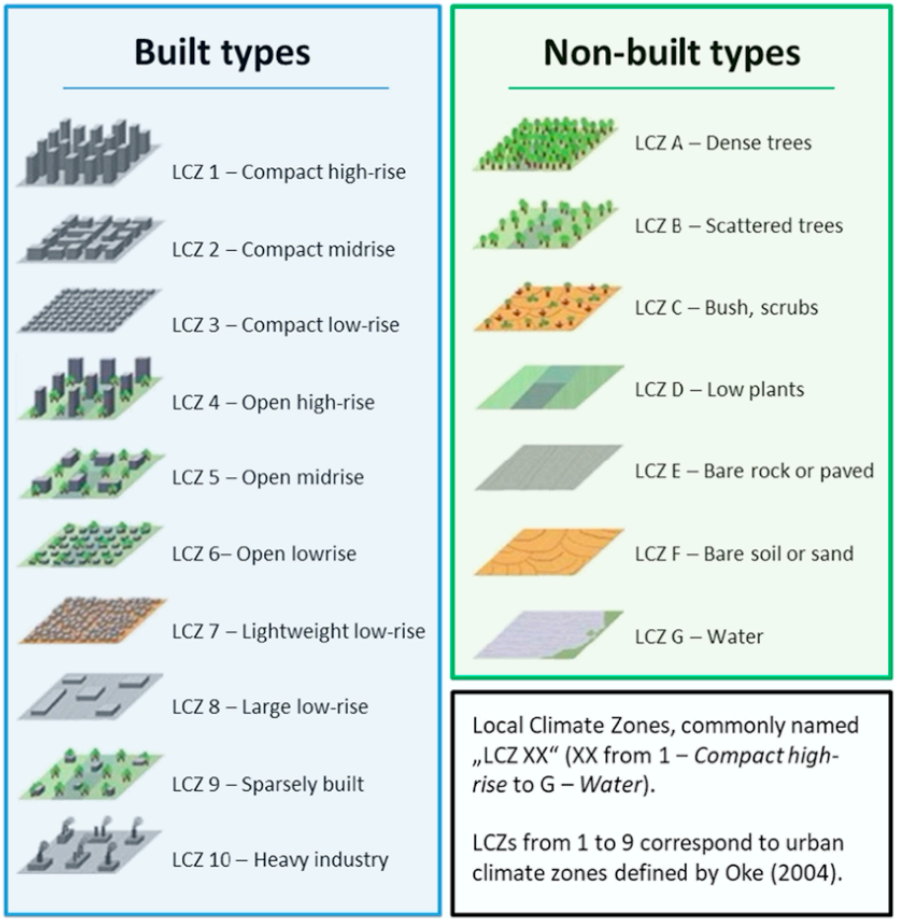


Figure 2.14: Local Climate Zones as reported by Taubenböck et al., 2020, figure 1)

2.2.3 Overview

The relation of each of the concepts mentioned above to the OCM is summarised in the table 2.1 below.

Table 2.1: Relation of methods to OCM. x marks fulfilled criteria, - un-fulfilled and * potentially fulfilled.

reference	exhaustive	exclusive	empirical	hierarch.	compreh.	detailed	scalable
Dogrusoz & Aksoy (2007)	x	x	x	-	-	x	x
Song & Knaap (2007)	-	-	x	-	-	-	x
Steiniger et al. (2008)	x	x	-	-	-	x	x
Berghauser Pont & Haupt (2010)	x	x	x	-	-	x	x
Gil et al. (2012)	x	x	x	-	-	*	-
Louf & Barthelemy (2014)	x	x	x	x	-	-	x
Schirmer & Axhausen 2015	x	x	x	-	*	*	x
Serra et al. (2018)	-	x	x	x	-	-	-
Dibble et al. (2017)	x	x	x	x	x	-	-
Berghauser Pont et al., (2019)	x	x	x	-	-	x	x
Araldi and Fusco (2019)	x	x	x	-	*	*	x
Dong et al. (2019)	-	x	x	x	*	*	x
Li et al. (2020)	x	x	x	x	-	*	-
Jochem et al. (2020)	x	x	x	-	-	-	x
UST	x	x	*	*	*	*	x
LC	x	x	-	*	*	*	x
LCZ	x	x	-	-	-	x	x

The majority of methods provide mutually exclusive classification, are exhaustive and empirical. Assessing scalability is not straightforward as the published research is mostly not reproducible without much effort. The scalability is then inferred from the case study area or the method description and may not be entirely precise. The least fulfilled criteria are hierarchical nature and comprehensiveness of methods, with only a few works being such. However, this seems to be changing, as the proportion is growing in recent years.

2.3 The gap in the systematic classification

Even though literature covers fields of geography, urban morphology, remote sensing or cartography, there is no classification method in the scrutinised body of work which fulfils all seven criteria of OCM. Both exhaustivity and mutual exclusivity are generally well covered as most of the methods follow these rules. The empirical nature is less presented. We have to keep in mind the difference between classifications with clusters based on empirical data and those that use empirical data to label observations properly. The latter is not empirical in the sense of OCM as the model itself is conceptual.

Although we can find methods which produce hierarchical classification, it is not a standard, likely due to a) popularity and performance of flat clustering methods as K-means in case of empirical studies, b) additional complexity in case of conceptual studies; hence they rarely reflect inter-group relationships. The significant issue is with the comprehensiveness of presented models. Most of them are based on a handful of dimensions, making them less robust and more prone to bias induced by the careful selection of indicators. Those few examples which are based on a broad set of descriptors (Dibble *et al.*, 2017)

unfortunately fail to fulfil one or more of the other criteria.

There are studies which are detailed, i.e. their spatial unit is granular (building, plot) but again - they are sub-optimal in other aspects. The scalability is not easy to assess as the majority of methods is not readily reproducible, some do not even describe used tools and computational framework, but that is an issue by itself. In works where the method is described to a sufficient detail, the scalability varies. However, with the rapid expansion of scientific geospatial software and GIS in general, it is assumed that it should not be an issue in future unless the method depends on qualitative data.

All above lead to the conclusion that current methods are not able to provide sufficient details and complexity when it comes to the classification of urban form and hence their applicability in further research or planning practice is limited.

2.4 Summary

Classification is a rich and polysemic term, which needs specific definitions to resolve possible ambiguities. However, even more specific terms as typology and taxonomy are used with different meanings. The ideal path is to be explicit about the context and do not rely entirely on a single definition of the term. Note that since taxonomy can stand for both procedure and the final result, the terminology also needs more refinement. Therefore, this research uses term taxonomy for the final result, whilst term cluster analysis for the procedure.

The classification itself brings a lot of valuable aspects. However, to ensure that all are present, the method for the classification of the urban form should follow the principles of the Optimal Classification Model based on seven simple requirements. It should be *exhaustive, mutually exclusive, empirical, hierarchical, comprehensive, detailed* and *scalable.* None of the models present in literature fulfils all seven criteria to date. The rest of this research will focus on developing a novel method, which embeds all principles into its design to derive the taxonomy (i.e. quantitative classification) of urban form.

Existing models for classification of urban form present in the literature vary and spread across multiple dimensions, from different basic scales (from city to building) to different

methodological prepositions (from remote sensing to typo-morphology). Across the field, all OCM criteria have been met; however, not within a single model.

Chapter 3

Numerical taxonomy

The previous chapter focus on the concept of classification in general and its application in the context of urban morphology. One of the specific classification methods is numerical taxonomy, which is conceptually quantitative, and structurally hierarchical. The following chapter aims to provide a more in-depth overview of numerical taxonomy and its components, from its biological origins to application on urban form.

The first section of this chapter introduces original biological methods of morphometrics and numerical taxonomy, including the overview of core concepts as Operational Taxonomic Unit, taxonomic characters, resemblance and structure. The second section then discusses the transferability of these concepts into urban morphology, starting with a theoretical review of principles. Furthermore, it critically assesses the preceding work of Dibble *et al.* (2017), who tried to provide the first link between the numerical taxonomy and urban morphology.

3.1 Learning from biology

From a conceptual point of view, this research builds on the previous of Dibble *et al.* (2017) (partially published as Dibble *et al.* (2015) and Dibble (2016)), which established the theory of urban morphometrics and made the first link between numerical taxonomy - a classification method used in primarily in biology - and urban morphology. Both morphometrics and numerical taxonomy have a long and rich history outside urban sciences

and the core idea behind the Dibble's work, which is also applied to this research, is that urban form can be in its own way perceived as a group of individual entities which can be consistently measured and numerically classified. In Dibble *et al.* (2017), authors demonstrated that the idea is viable and that the potential of urban morphometrics should be explored further.

Researchers in biology did much work setting the scene for morphometrics, and we should learn from it to avoid reinventing the wheel, where the key principles hold.

3.1.1 Morphometrics as a descriptive tool

Morphometrics is a quantitative analysis of form, i.e. size and shape (Rohlf and Marcus, 1993) of objects under scrutiny, in biology that usually refers to a size and shape of an individual, in which case morphometrics, as a descriptive tool has the ability to determine the level of the resemblance of different individuals, either of the same species or different species (Sneath and Sokal, 1973). Literature knows several branches of morphometrics, all conceptually similar, used for complementary purposes but based on different principles of what to measure and how. The two main approaches are traditional and geometric morphometrics.

Traditional or multivariate morphometrics is based on "*the application of multivariate statistical methods to sets of variables*" (Rohlf and Marcus, 1993, p. 129). The variables used within the analysis usually represent widths, lengths and angles of features or distances between specific landmarks (Rohlf and Marcus, 1993), hence containing little information about the geometry of the structure (Zelditch *et al.*, 2004). The recreation of the original shape is not possible based on the used set of variables (Adams *et al.*, 2004). Among the statistical techniques used within traditional morphometrics Rohlf and Marcus (1993) list principal component analysis, discriminant functions of canonical variate analysis. An example of measurements used within traditional morphometrics is shown in figure 3.2.

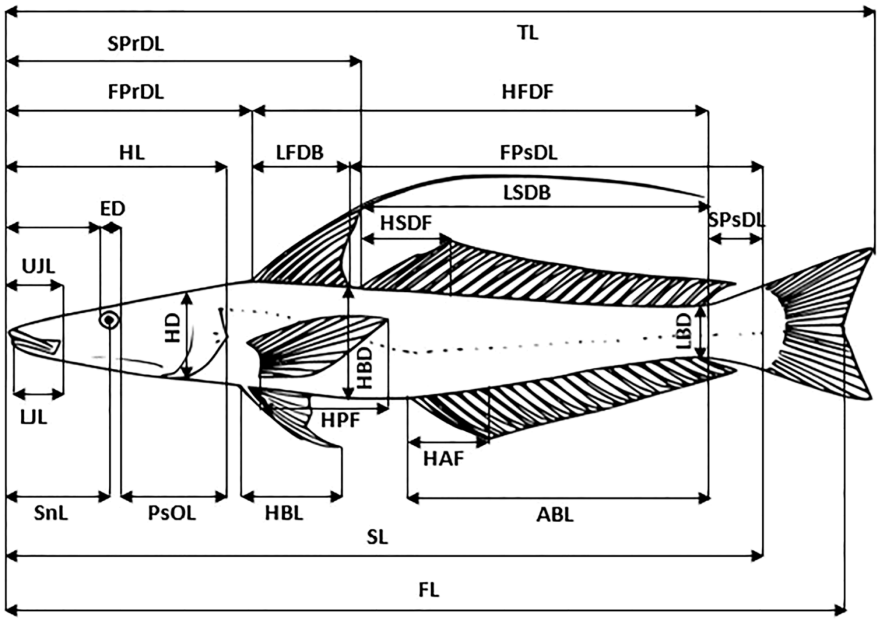


Figure 3.1: Example of the application of multivariate morphometrics. Overview of different morphometric measurements investigated in S.panijus (Siddik et al., 2016)

The typical application of traditional morphometrics is *allometry*, the study of change of the shape with the change of the size. However, in biology, this method has certain limitation mostly related to the normalisation of the shape values affected by size (Breno *et al.*, 2011).

The second major branch of morphometrics, which tries to overcome some limitations of the previous one is geometric morphometrics (Rohlf and Marcus, 1993), focusing on the position of landmarks and semi-landmarks on the grid and its deformation (figure 3.2). The description of both size and shape is then captured through the series of coordinates either on a 2D plane or in 3D space.

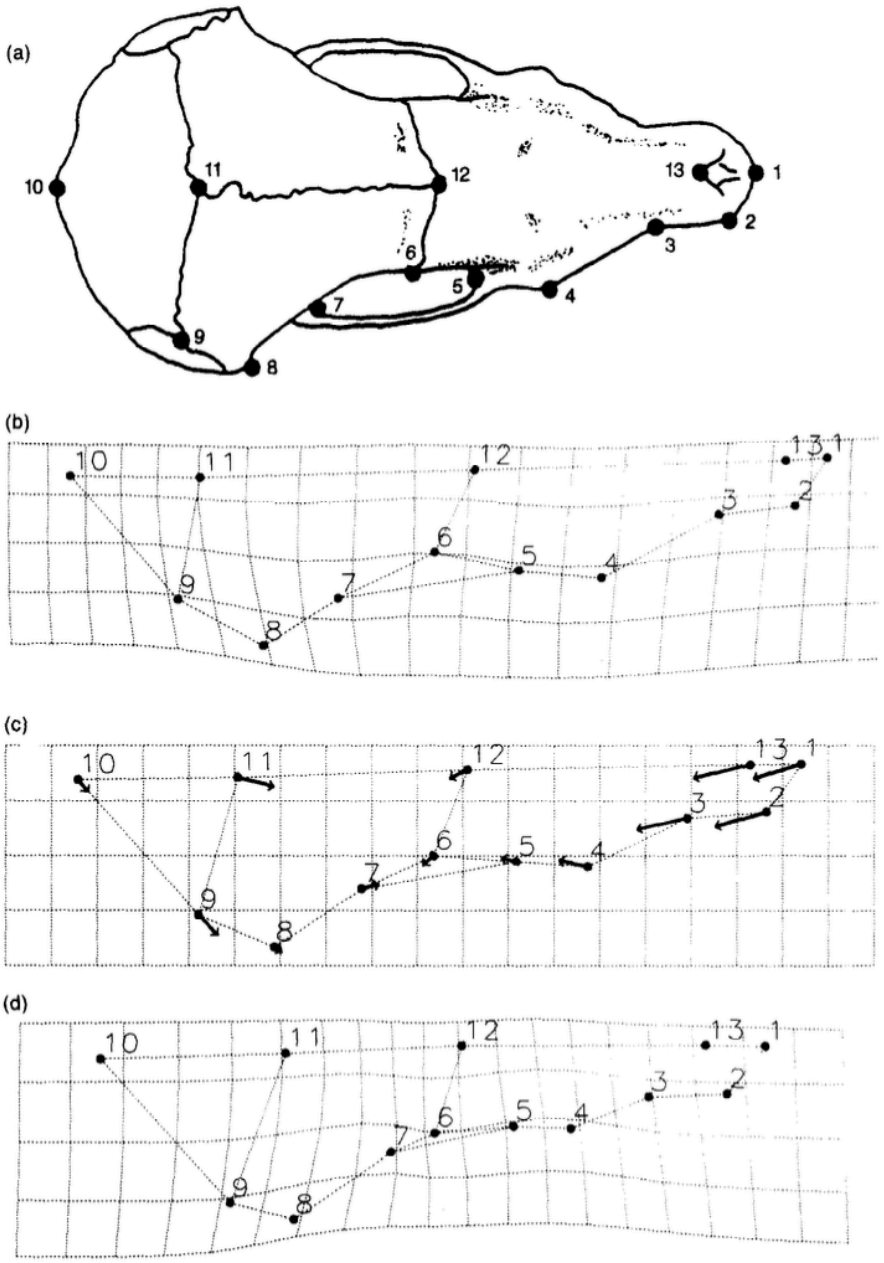


Figure 3.2: Example of the application of geometric morphometrics. Locations of the landmarks on a dorsal view of the skull and the relevant distortion of the grid.

Geometric morphometrics is dependent on the shared landmarks features between an individual in a study. Hence its application in an urban context is difficult due to the complexity of urban form and the variety of patterns, making the definition of shared

landmark features complicated.

Similar to geometric morphometric is an outline analysis, which is trying to fit mathematical functions to points sampled along the outline of the studies shape, an approach which has already been applied in urban morphology either to study the shape of settlements (Batty and Longley, 1987) or different aspects of streetscape (Cooper, 2003, 2005). Furthermore, there are other branches of morphometrics (tissue morphometrics (Bodenstein and Sidman, 1987), geomorphometrics (Coblentz *et al.*, 2014), neuroimaging (Wang and Jernigan, 1994)) which are applying the same principles in their respective fields.

All these approaches and methods can inform the development of *urban* morphometrics.

3.1.2 Numerical taxonomy as a classification tool

One of the specific application of morphometrics is the usage of measurements in the classification of species, for which Sneath and Sokal (1973) use the term *numerical taxonomy* and define it as "*the grouping by numerical methods of taxonomic units into taxa on the basis of their character states*" (Sneath and Sokal, 1973, p. 4). In biology, this requires a distinction between *phenetics* (overall similarity) and *cladistics* (evolutionary branching sequence).

The key principle of numerical taxonomy is that it is a classification which is operational and empirical. As further explained by Sneath and Sokal (1973), "*operationalism implies that statements and hypotheses about nature be subject to meaningful questions; that is, those that can be tested by observation and experiment (p.17)*". That further assumes that numerical taxonomy as a method has the potential to be reproducible and replicable.

Sneath and Sokal (1973) define seven key principles of numerical taxonomy:

1. *The greater the content of information in the taxa of a classification and the more characters on which it is based, the better a given classification will be.*

Selection of morphometric characters which is used for classification is inherently biased. To limit a bias as much as possible, we should employ an as large number of characters as is possible (and meaningful) as long as it does not violate any other rule.

2. *A priori, every character is of equal weight in creating natural taxa.*

Some of the classification models apply weighting, which means that character A is seen as more important than character B. However, this decision often introduces bias into the selection because the rules defining the weighting can be arbitrary. Furthermore, we do not a priori know which character will have the highest discriminatory power.

3. *Overall similarity between any two entities is a function of their individual similarities in each of the many characters in which they are being compared.*

In other words, when measuring the similarity between two entities, all characters need to be taken into account.

4. *Distinct taxa can be recognised because correlations of characters differ in the groups of organisms under study.*

Taxa, i.e. classes of numerical taxonomy, represent different groups of organisms or other entities under scrutiny. The relations between morphometric values within taxa are different from relations to other taxa.

5. *Phylogenetic inferences can be made from the taxonomic structures of a group and character correlations, given certain assumptions about evolutionary pathways and mechanisms.*

Phenetic information used to build a taxonomy is not able to reflect phylogeny entirely, but the difference between the true evolutionary development of species and that observed in phenetic taxonomy is often minimal. In an urban environment, that means that we might be able to, to a degree, trace historical origin in numerical taxonomy of urban form.

6. *Taxonomy is viewed and practised as an empirical science.*

Conceptual assumptions about observations should have no role in the taxonomy.

7. *Classifications are based on phenetic similarity.*

Phenetic similarity represented by the relations (e.g. correlation) between morphometric characters is the only aspect taken into account when building a classification.

While the scope of this work is not to give a detailed overview of biological numerical taxonomy, there are four crucial concepts which need to be understood before the application of numerical taxonomy to urban form - Operational Taxonomic Unit (OTU), taxonomic characters, taxonomic resemblance and taxonomic structure.

3.1.2.1 Operational Taxonomic Unit

OTU is the lowest ranking taxa employed in a given study. In biology that is usually species represented by an individual (e.g. a single bird specimen). In some cases, OTUs can be species or other aggregated groups if the aim is to develop higher-order taxa. Generally, OTU should be the fundamental unit in a large majority of instances. Sneath and Sokal (1973) suggest that in the case of biology, "*a taxonomic unit at any level should be based on individuals*" (p.69) to ensure consistency and rigour. Naturally, the urban form does not offer a simple definition of the individual, and the question of optimal OTU needs to be further studied.

3.1.2.2 Taxonomic characters

Taxonomic characters are, in essence, morphometric characters used to derive numerical taxonomy. For the application in biology, they can be defined as "*a characteristics (or feature) of one kind of organism that will distinguish it from another kind; or any attribute of a member of a taxon by which it differs or may differ from a member of a different taxon*" (Sneath and Sokal, 1973, p. 71) For application on urban form, the term *organism* will refer to a set OTU.

The critical step in the design of numerical taxonomy is the selection of taxonomic characters. The set of characters which is used can significantly affect the results of the analysis; therefore, an extensive set of rules of selection should be defined and followed. The general principle should be driven by inclusivity - all available kinds of characters should be used.

Furthermore, all characters which may bring new information should be employed, and the set should not follow a limited set of conventionally used analytical variables. The advice given by Sneath and Sokal (1973) says *"take as many characters as is feasible and distribute them as widely as is possible over the various body regions, life history stages, tissues, and levels of organisation of organisms" (p. 108)* - again, the terminology refers to biology. However, the principle could be translated into urban morphology.

3.1.2.3 Taxonomic resemblance and structure

Taxonomic resemblance, the similarity between OTUs, is determined using cluster analysis based on the resemblance matrix. Resemblance matrix consists of all morphometric values and all OTUs in the study. As results, clusters, i.e. higher-order taxa, are based on phenetic resemblances in an objective manner.

The optimal interpretable outcome of numerical taxonomy is a hierarchical dendrogram merging lower-order taxa into the higher-order (fig. 3.3). However, the methods of cluster analysis vary, even though the tendency is to employ hierarchical methods. [1]

[1] That does not necessarily mean that flat clustering methods like K-means are not used. Although if the final output should be captured in a dendrogram, a single layer of K-means clustering is not enough and other steps should be introduced.

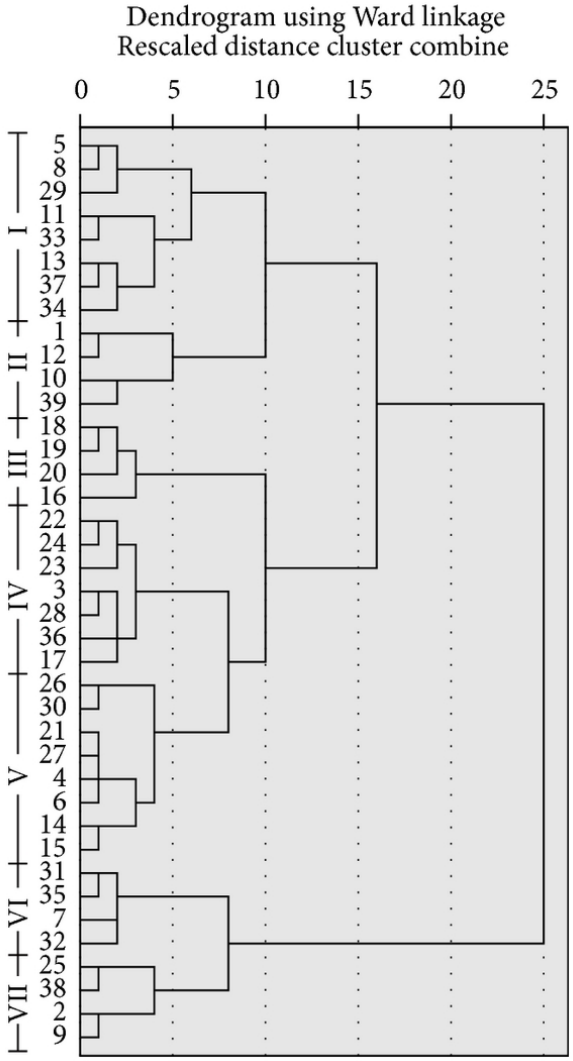


Figure 3.3: Example of dendrogram using Ward's method classification of 39 species of Mallomonas. (Feng and Xie, 2013)

Of high relevance for the urban form is a *mixture problem*. That happens "*when taxonomist assumes that sampled populations consist of a mixture that he first wishes to decompose into separate populations, which are then investigated further or are used as OTU's*" (Sneath and Sokal, 1973, p. 199). If the OTU in place will be an aggregation as is Sanctuary area in case of Dibble *et al.* (2017) or urban tissue, the sample may be composed of individual buildings which need to be first assigned into their respective SA/tissue. In that case, the OTU for taxonomy itself will be the *population of buildings* composing a morphological

aggregation.

3.1.3 Criticism of numerical taxonomy

As any other classification, numerical taxonomy has its drawbacks and limitations, which should be carefully considered before its application. Sneath and Sokal (1973) name three major problems related to the phenetic classification of organisms of which all are relevant for urban morphology.

The first one points out a possible dissimilarity of clusters based on morphometric characters obtained during different time periods. Those would be life stages for an animal but could be seen as stages of the development cycle in urban morphology. Phenetic description of OTUs would differ, potentially causing differences in taxonomy. While it is relatively simple to control such an issue in biology, capturing urban form in the same stage of the development cycle is virtually impossible, and the limitation needs to be taken into account in the interpretation of resulting classification.

The second and third problems are related and focus on the design of a cluster analysis. The selection of similarity coefficients between morphometric characters and selection of clustering method can both result in differences in identified relationships between OTUs and affect the interpretability of the classification. The method of classification should then be selected either as a result of comparative analysis or based on the theoretical understanding of the data structure and the optimal clustering algorithm for the data.

3.2 Transferability of concepts into urban morphology

The conceptual and methodological framework of numerical taxonomy and morphometrics is well defined in biology but less so in urban morphology. Therefore, the transferability of individual concepts needs to be studied before a method leading numerical taxonomy of urban form can be defined.

3.2.1 Morphometrics

Morphometrics has a tradition in urban morphology, although the term itself is not fully established yet. There is a wide range of literature applying quantitative methods of characterisation of form, ranging from building dimensions (Schirmer and Axhausen, 2015) to network-based characters (Porta *et al.*, 2010) and beyond (Araldi and Fusco, 2019). The detailed overview of a published work is available in chapter 4. What can be concluded now, is that the principle of morphometrics, explicitly brought to urban morphology first by Carneiro *et al.* (2010) and then further elaborated by Dibble *et al.* (2017), is applicable to urban form without many constraints. However, to ensure that the set of morphometric characters used for a taxonomy of urban form is inclusive enough, detailed analysis of the potential of urban morphometrics and its current limits is required and provided in chapter 4. It has to be noted that *urban* morphometrics in conceptually traditional, multivariate morphometrics.

3.2.2 Operational Taxonomic Unit

From the theoretical perspective of a numerical taxonomy as a classification concept, there is no reason to believe that it could not be applicable to urban morphology. From a practical perspective, it is necessary to discuss individual components of the procedure, primarily Operational Taxonomic Unit (OTU).

In biology, an OTU for taxonomy on the level of a specimen is an individual. However, in urban morphology, we face the problem defining what is individual in cities. If we try to define an individual generally, we will find the following:

- *a single person or thing, especially when compared to the group or set to which they belong* (Cambridge Dictionary, 2020)
- *a single organism capable of independent existence* (Dictionary.com, 2020)
- *a member of a compound organism or colony* (Dictionary.com, 2020)
- *a particular being or thing as distinguished from a class, species, or collection* (Merriam-Webster.com, 2020)
- *a single organism as distinguished from a group* (Merriam-Webster.com, 2020)

- *a single, separate organism (animal or plant) distinguished from others of a same kind* (Biology-online.org, 2020)

Morphological literature used various elements as the unit of analysis, from a building (Schirmer and Axhausen, 2015) or plot (Bobkova *et al.*, 2017) to sanctuary area Dibble *et al.* (2017) or neighbourhood (Song and Knaap, 2007). However, none gives a definite answer on what is the individual, the smallest indivisible meaningful unit from which we could derive a taxonomy of urban form. The fact that settlements can grow from a single building, and are composed of buildings as (one of) fundamental elements (Moudon, 1997) may lead to an idea of a building being an individual in the city. Alternatively, the same conclusion could be made about a plot. In such a case, a taxonomy of urban form focusing on the classification of built form patterns is not examining the lowest taxa (buildings), but higher ranks, which means that we have to deal with the mixture problem outlined in section 3.1.2.3. We should then look at the problem in a similar way as biologists are looking at taxonomy at the level of the population (put aside the fact that population in the case of cities still needs to be defined) or species level.

That moves the problem to a different terminological issue - how do we define a *species* of urban form. Biology knows several ways of species definition, one based on the deviation of DNA code, other based on the ability of two individuals to interbreed (De Queiroz, 1998). However, the definition which could be helpful in urban morphology is a phenetic one - *a taxonomic species based on morphologically similar populations located in a definite geographic area and morphologically distinct from other populations assigned to different species (Sneath and Sokal, 1973, p. 364).* In practice, phenetic species can be defined as "*the smallest cluster that can be recognised upon given criterion as being distinct from other clusters*" (ibid), which could be seen as the smallest cluster of buildings or plots, characterised by their form and spatial distribution and configuration, which is morphologically distinct from the other.

3.2.3 Critical assessment of preceding work

This work, as mentioned above, is a direct continuation of the previous research at the Urban Design Studies Unit at the University of Strathclyde published as Dibble *et al.* (2017). The authors attempted to bridge urban morphology and numerical taxonomy,

which means that they also provided a conceptual analogy of crucial concepts and notably OTU.

The following section will examine several aspects of preceding work and argue why some of the previous decisions need to be reevaluated.

3.2.3.1 Sanctuary area as OTU

As an OTU, Dibble *et al.* (2017) use Sanctuary area (SA), defined as the portion of the urban fabric enclosed by main streets (Mehaffy *et al.*, 2010). However, such a decision comes with an inherent issue of potential internal heterogeneity. The SAs used in previous research were ideal cases, but cities are not composed of ideal cases only. The classification model should recognise what cities are composed of, which patterns and urban forms and systematise them. Which means that OTU needs to reflect such individual patterns, while a concept of SA applied to the whole city comes with a large portion of SAs which are internally heterogeneous - composed of multiple patterns. In that case, results of classification would not reflect the actual patterns of urban form but the way they coincide with each other.

Consider an illustrative example below, of a small town in Czechia[2]. The whole fabric of the town is composed of few SAs. The working hypothesis Dibble *et al.* (2017) use is that it is composed of several classes of urban patterns following the division of urban fabric into SAs as shown on figure 3.4[3].

[2]The hometown of the author.

[3]The identification of homogenous areas is purely perceptional, based on the personal knowledge of the place and visual interpretation of build form patterns. The map is for an illustrative purpose only.

Figure 3.4: Illustration of sanctuary areas in a town Milevsko, Czechia. Main streets forming boundaries of SAs are marked as black lines. It is clear that SAs are composed of multiple heterogeneous patterns and do not function as a unit of urban form which would be sensible to use as OTU.

In case of using SAs as an OTU, the distinction between different urban patterns (figure 3.5) within this town would be impossible as these SAs are far from homogenous. The problem with SAs is that their definition and identification is essentially a phylogenic approach; it is based on the process of development of the settlement. The rest of the classification is, however, a systematisation based on purely phenetic attributes. Therefore, I argue that Sanctuary Area does not fit into the definition of an OTU, because an SA is not taxa on any level of systematisation, and should not be used for general analysis outside the safe selection of the best examples (like in the case of Dibble *et al.* (2017)).

Figure 3.5: Illustration of homogenous patterns of urban form in a town Milevsko, Czechia. The optimal goal of taxonomy would be to classify patterns as these.

Use of a SA as an OTU assumes that whole cities are ideal according to 'Emergent Neighbourhood Model' (Mehaffy *et al.*, 2010). Even the authors state that they are not (e.g. the three pathologies). While the concept of SA in this model perfectly works, in the case of taxonomy, it does not.

3.2.3.2 Selection of taxonomic characters

While selecting morphometric characters used for classification, one should avoid the empirical correlation of resulting values (Sneath and Sokal, 1973). That means we should not include two or more characters capturing conceptually the same aspect of an OTU as such concept would be overrepresented in the set of characters and skew the result of cluster analysis (see section 7.2.1 for a detailed discussion). In the case of Dibble *et al.* (2017), the assessment of collinearity is missing, and there are likely characters which are

collinear. The first indication is a theoretical assessment of all characters, but the clear manifestation is their cost-benefit analysis (CBA). CBA results show that with only 9 variables out of 207, the classification model reached more than 90% accuracy. Such a radical reduction of dimensionality would not be possible if the initial set of characters avoided correlated ones.

Therefore, the whole set of characters used within numerical taxonomy needs to be revised alongside OTU and other conceptual and methodological aspects.

3.3 Summary

Chapter 2 provides an overview of models of classification of urban form, methodologically limiting it to quantitative approaches. Chapter 3 focuses on one specific way of classification - numerical taxonomy, the approach which is established in another field of research, notably biology where it originates, and which has undoubtedly potential to provide useful insights into urban morphology.

Numerical taxonomy is a purely quantitative classification based on the morphometric assessment of taxonomic units. Morphometrics itself is present in urban morphology in recent years, although often not explicitly, but under the umbrella of quantitative urban morphology. The current expansion of data availability and enhancements in computational tools and environments allow us to generate a detailed morphometric description of urban form, as is shown in detail in the next chapter. Principles of numerical taxonomy are slowly finding its way to urban morphology as well, but still lacking methodological comprehensiveness.

The potential of morphometrics and numerical taxonomy in the urban context is high, as Dibble *et al.* (2017) previously shown. However, the transfer of the method from biology and other fields is not yet optimal as the previously proposed method was exploratory, which limits its further applicability.

Numerical taxonomy of urban form therefore requires a detailed study of each aspect of the method. Urban morphology will have to deal with several interconnected issues. The core issue lies with the identification of OTU, which will likely be linked with a

mixture problem. The OTU of urban form will therefore be a *population* of fundamental elements, which needs to be identified first. Another important aspect which needs to be thoroughly revisited is the selection of morphometric characters used within the study to avoid collinearity issues and ensure a detailed description of patterns at the same time. The final task will lie in the actual method of numerical taxonomy. There is a variety of options on how to derive the phenetic relationship between OTUs, but only some will be fit for the context of urban morphology.

Chapter 4

Urban morphometrics and its terminological inconsistency

The content of this chapter was partially published in Fleischmann and Romice et al. *(2020).*

The previous chapter outlined the need for morphometrics as a key component of numerical taxonomy and pointed out that literature in urban morphology implements a wide spectrum of morphometric characters.

In the age of urbanisation, urban planning and design still struggle to offer reliable models to address the challenges of the 21st century (Cuthbert, 2007; Romice *et al.*, 2020), while the discipline's shift towards an evidence-based approach and a "new science of cities" is still in its infancy (Batty, 2012, p. S15). Despite remarkable growth and progress, urban morphometrics is no exception. In particular, two issues still hinder a quantitative approach to the analysis of urban form: first, the availability, quality and consistency of data across geographical regions; second, the discipline's inherent difficulties in offering a rigorous and consistent definition of urban form, its fundamental components and the relationships between them. This chapter contributes to the resolution of this second problem.

The high variety of morphometric characters, defined as *a characteristic (or feature) of one kind of urban form that distinguishes it from another kind* (adapted from Dibble

et al., 2017; Sneath and Sokal, 1973), used in literature is fragmented across numerous unrelated sources, and despite several attempts to systematise it, (Larkham and Jones, 1991; Caniggia and Maffei, 2001; Conzen, 2004; Dibble *et al.*, 2017) a comprehensive overview is still lacking. This gap of knowledge creates uncertainty as to which research areas are covered and which need further research. Moreover, the terminology is not consistent nor univocal, resulting in weaker methodological compatibility and higher hurdles in comparing research outputs. According to Whitehand (2012),

> "comparative research is faced with a plethora of case studies that use different, or sometimes unspecified, definitions. […] In addition to problems of non-comparability of definitions, methods and concepts, differences between the sources of information employed need to be overcome" (p.60).

In this chapter, this research: a) proposes a coherent and comprehensive classification system of measurable urban form characters, and b) uses this system to resolve current inconsistencies and redundancies and identify areas of weakness in the existing literature.

4.1 Selection and systematisation of methods and characters

The first section presents: 1) the criteria utilised to select relevant literature used to map the field of UM; 2) the process of systematisation of such literature, which is then used to 3) identify, cross-compare, (re)define and 4) the re-classify morphometric characters.

As for terminology, terms such as "attribute", "variable", "measurement", "metric", "index", "character", "indicator" or "proxy" are often used interchangeably in urban morphology to signify the measurable feature of an object (Schirmer and Axhausen, 2015; Bobkova *et al.*, 2017; Dibble *et al.*, 2017; Vanderhaegen and Canters, 2017; Araldi and Fusco, 2019). This research follows Dibble *et al.* (2017) where the term "character" defines "*a characteristic (or feature) of one kind of organism that will distinguish it from another kind*" Sneath and Sokal (1973). Here, however, "organism" refers to a distinct kind or type of urban form. "Urban form" as a term has been used to loosely signify different aspects of space's configuration in cities along with its use and agents. It is, therefore, a polysemic term, while this work refers exclusively to the physical components

of urban space, i.e. the built-up fabric (blocks, streets, buildings…) and its fundamental spatial subdivision (plots) after Moudon (1997).

4.1.1 Literature selection and systematisation

To review the literature (figure 4.1), this research selected sources that: a) explicitly undertake a *quantitative* examination of urban form characters[1]; b) include urban form characters that are not present in already selected sources, to avoid unnecessary duplication and overlapping.

First, it looked at papers published in two leading journals of urban analytics and morphology: "Environment and Planning B" and "Urban Morphology". From here, it extracted keywords, which were then used to identify several academic citation databases (Google Scholar, Scopus, Mendeley Search, ResearchGate, Taylor and Francis Online) and to undertake a broader snowballing exploration. The process of keyword search and snowballing was iterated whenever new inputs were found and adopted to ensures that the selection is rigorous and inclusive.

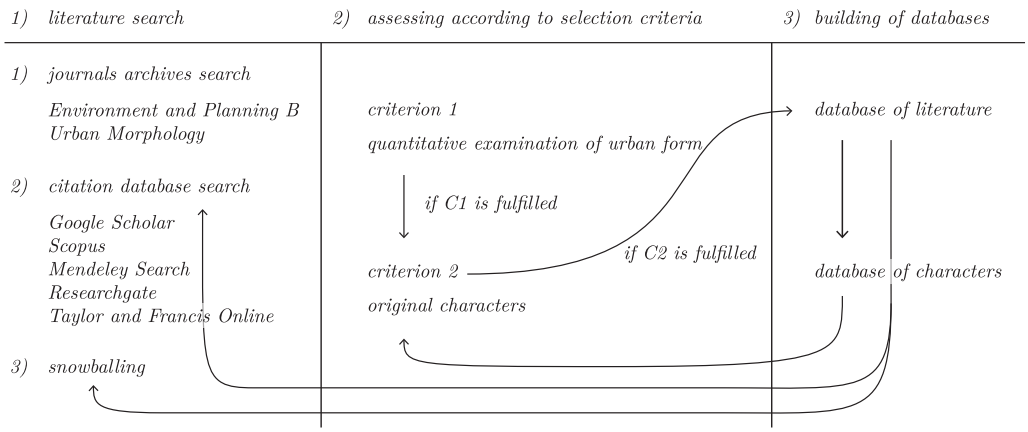


Figure 4.1: Scheme of the process of selection of literature and its usage.

All selected papers were then classified according to grain, i.e. the scale (size) of *the basic spatial unit* on which descriptors are calculated; extent, i.e. the scale (coverage)

[1]In some cases, the research focused both on physical and non-physical characteristics and was therefore included, but only the physical part of the method was used in the analysis.

of *the case study*; purpose; potential comprehensiveness, i.e. the number of urban form characters measured; timeframe, whether *synchronic* (comparing different cases at the same time) or *diachronic* (comparing the same case at different times).

While there is a partial overlap with the literature presented in chapter 2, the scope of this chapter is different and focus on a wider spectrum of research methods.

As a *grain* is considered the basic spatial unit as the smallest element being measured, while for *extent*, is coverage as the total area of the case study analysed. Both are taken into account and then organised from 1 (small) to 10 (large)[2].

4.1.2 Classification of characters

From the sources classified as above were then extracted individual morphometric characters. Those influenced by non-morphological data, such as distance to the nearest bus stop (Song and Knaap, 2007) or land use (Dibble *et al.*, 2017), were excluded.

To overcome terminological inconsistencies among the morphometric characters adopted in different studies[3], this research comprehensively redefined them (see Section 4.3.1). On these new definitions is then designed a classification framework of characters, based on their nature and the spatial unit they belong to. Finally, the framework is tested in the classification of all urban form characters initially extracted from literature, followed by a discussion of the emerging gaps and redundancies and suggestions for further developments.

4.2 A state of art

While the existing literature on urban morphology shows a historical inclination towards qualitative methods (Dibble *et al.*, 2015), through the iterative literature review process

[2]This classification is based on conceptual ranking rather than metric size: the building scale is smaller than the plot scale, in that the former is conceptually contained in the latter, even though in terms of sheer size some buildings may be larger than some plots.

[3]This phenomenon, occurring when the same urban form characters are presented under different names or different urban form characters under the same name, is called nicknaming in this work.

illustrated above are identified 72 predominantly *quantitative* works (peer-reviewed articles, conference papers, book chapters, PhD theses). In figure 4.2, selected literature items are *positioned* according to their grain and extent scales and *classified* by their purpose (colour), and a number of urban form characters considered (size).

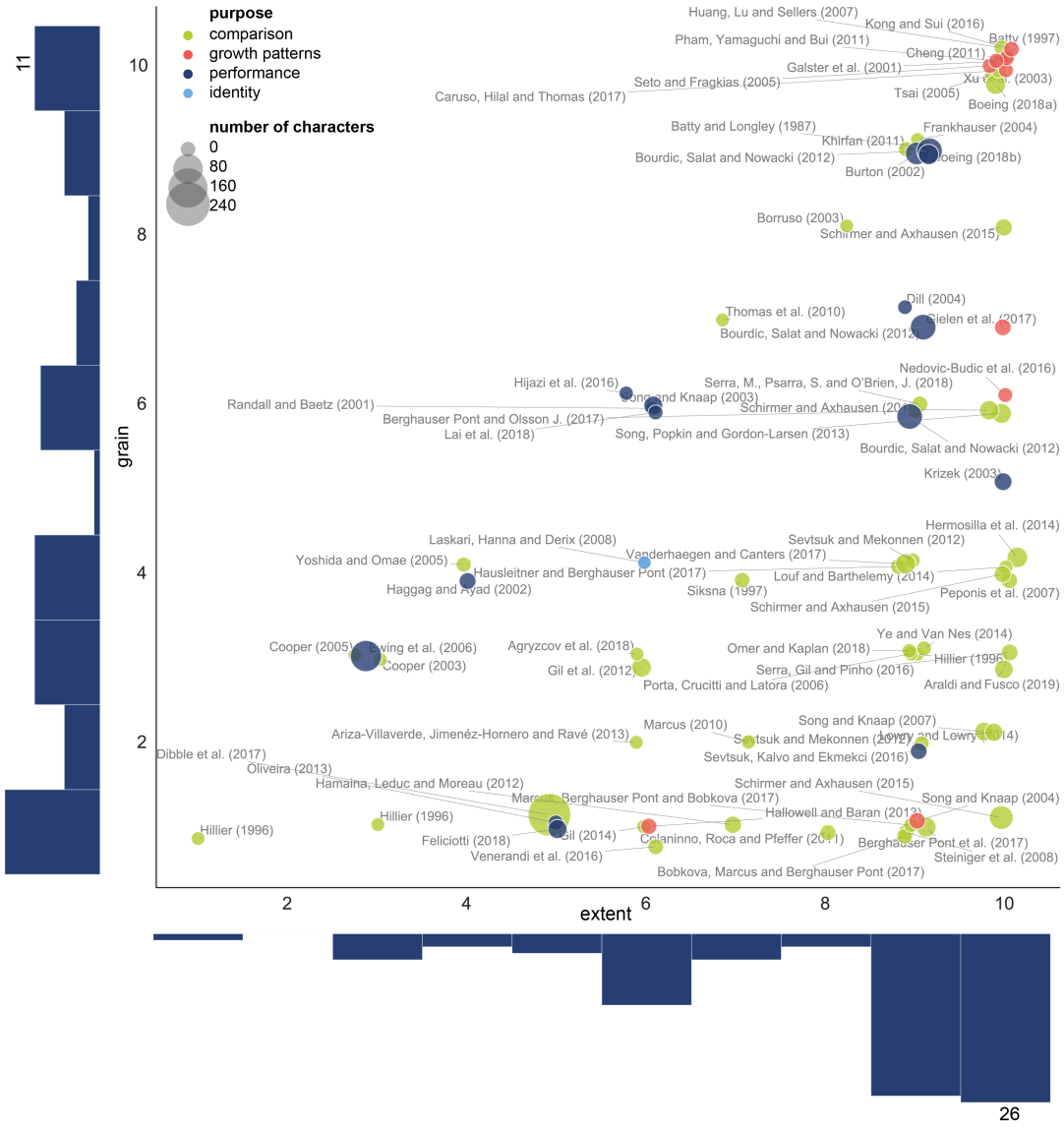


Figure 4.2: Classification of Literature. Predominantly quantitative studies in urban morphology classified according to grain scale (Y axis), extent scale (X axis), purpose (colour) and number of urban form characters (size). The histograms show a relative balance in terms of scale of grain and a tendency towards large scales of extent. Note: placement of points is jittered to minimise overlaps.

4.2.1 Patterns of research

Quantitative analysis in urban morphology appears to have three distinct research purposes in particular: to enable comparison among cases, to measure the performance of urban form and to monitor or predict urban growth. *Comparison* is the largest group containing 45 out of the 72 selected works (62%) and is significantly synchronic (95%). It includes studies which cover a range of urban form characters from one only (Batty and Longley, 1987; Frankhauser, 2004; Thomas *et al.*, 2010; Ariza-Villaverde *et al.*, 2013; Agryzcov *et al.*, 2017) to many (Dibble *et al.*, 2015, 2017); however, those covering more than 10 urban form characters are only the 33%, and those with more than 25 the 15%, demonstrating a lack of comprehensiveness in literature. In terms of scales, comparative studies tend to be lower in grain scale (more detailed) and higher in extent scale (more extensive case studies).

Papers measuring *performance* refer in particular to one specific aspect of urban form, such as sustainability (Haggag and Ayad, 2002; Bourdic *et al.*, 2012), resilience (Feliciotti *et al.*, 2016), urbanity (Oliveira, 2013), or network-based accessibility (Krizek, 2003; Sevtsuk *et al.*, 2016). Similar to the comparison group, the majority of works in this second group is synchronic. However, unlike comparative studies, they tend to use similar scales for both grain and extent.

Not surprisingly, studies on urban *growth* are mostly diachronic. Many publications in this group focus on the analysis of urban sprawl (Galster *et al.*, 2001; Song and Knaap, 2004) to capture sprawl indices (Gielen *et al.*, 2017); here data are often aggregated and classified in a built-unbuilt binary framework (Galster *et al.*, 2001; Seto and Fragkias, 2005), enriched by Cellular Automaton (Batty, 1997; Kong and Sui, 2016) or machine learning (Cheng, 2011) techniques. As growth is measured mostly at a metropolitan scale, with a few exceptions (Hallowell and Baran, 2013) all works focus on a large scale of extent, while mostly using the same scale of grain.

Crucial for the success of a comparative method is *complexity*. It is represented by both the cross-scale extent of the research, as reflected for example in the work of Song and Knaap (2007), later refined by Song *et al.* (2013) or Schirmer and Axhausen (2015)[4], and the number of urban form characters measured (potential comprehensiveness). Still, over

[4]In the figure 4.2, cross-scale research is listed at all relevant scales (as Schirmer and Axhausen (2015)).

the whole set of 72 literature items selected, those measuring a number of urban form characters large enough to minimise biases and errors (i.e. > 25 urban form characters) is relatively rare (15%). Only recently, a few such comprehensive studies started to emerge (Ewing *et al.*, 2006; Bourdic *et al.*, 2012; Oliveira, 2013; Schirmer and Axhausen, 2015; Dibble *et al.*, 2017), contributing to the growing area of urban morphometrics (Carneiro *et al.*, 2010; Dibble *et al.*, 2017; Feliciotti *et al.*, 2017). However, the sheer number of urban form characters scrutinised (comprehensiveness) does not necessarily ensure complexity, as many of them may be collinear and hence capture the same information.

4.3 Classification of morphometric characters

The review of the 72 quantitative studies illustrated above produced a list of 465 individual morphometric characters, which are further studied and classified.

4.3.1 Nomenclature

Of these 465 characters, many were duplicated or hidden under the same name ("nicknaming"), suggesting the persistence of significant nicknaming even in the quantitative area of urban morphology analysis. For example, the term "connectivity" is in some cases used to signify a broader group of urban form characters (usually related to network analysis) (Dibble *et al.*, 2017), while in other cases is attributed to one single one of them, and yet with different meanings (Hillier, 1996; Lowry and Lowry, 2014); in some instances, the term is used in both ways in the same study (Bourdic *et al.*, 2012).

Hence, this research applied a process of "character redefinition", and introduced the "*Index of Element*" aimed at achieving a higher degree of consistency between the name of urban form characters and their *substance*. This index essentially defines each morphometric character according to *the measure that it calculates* (the *Index*) and *the element of urban form that it measures* (the *Element*). Let us consider the "connectivity" of the pedestrian grid in Bourdic *et al.* (2012), for example. We can easily distinguish the measure being calculated (Index), which is a *weighted number of intersections*, and the "thing" the urban form character of which is calculated *(Element)*, which is the *pedestrian*

network. This brings redefinition of the measure as *"Weighted Number of Intersections of Pedestrian Network "*, leaving much narrower room for interpretation. The use of a rigorous terminological criterium such as the Index of Elements is, even in quantitative urban morphology analysis, still occasional, though not absent (Schirmer and Axhausen, 2015). The Index of Element helps achieve an understandable definition of urban form characters *by their same name*: the *Index* part of the name captures the nature of the measure, independently from what is measured, while the *Element* part of the name captures the nature of what is measured, independently from how it is measured. Urban form characters defined by the combination of the two become consistently understandable and comparable across different methods. Application of this method on 465 identified urban form characters led to the elimination of 104 cases of duplication (22.4%), leaving 361 uniquely defined ones.

Table 4.1: Examples of Index of Element conversions. In some cases, urban form character's redefinitions bring in crucial information about the urban form character, in others only minor change. However, adding *Element* into the urban form character's name helps to develop quantitative urban morphology by making it more intelligible, hence comparable.

original name	index	element	reference
Urban Form	Continuity	Built-up area	Gielen *et al.* (2017)
Connectivity of the pedestrian grid	Weighted Number of Intersections	Pedestrian network	Bourdic *et al.* (2012)
Redundancy index	Redundancy	Street network	Feliciotti (2018)
Block section	Longest diagonal of/between	Block	Feliciotti (2018)
Building size - footprint	Area	Building	Hallowell and Baran (2013)
Built-up area	Built-up area	Block	Gil (2014)
Distance	Distance	Building	Hijazi *et al.* (2016)
Angle	Angle	Building	Hijazi *et al.* (2016)

4.3.2 Classification

Having tackled the terminology issue, this research proposes a typology of morphometric characters directly based on their name (which now captures their definition). This is a *"concept-based classification"*, i.e. one *" which conceptually separates a given set of items multidimensionally. … the key characteristic of a typology is that its dimensions represent concepts rather than empirical cases"* (Smith, 2002, p. 381). In this sense, by examining the urban form characters' names we can classify them along three dimensions: 1) the nature of the *Index,* 2) the scale of the *grain* of the character, and 3) the scale of the *Element*'s extent.

While most authors classify their observed urban form characters in groups, which are usually case-specific, these classifications vary. Generally, we can identify two approaches: one refers to the character's scale, as the sequence *Object*, *Composition, Neighbourhood*, *District*, *Municipality* and *Region* in Schirmer and Axhausen (2015); the second refers to the *Element's nature*, for example in Song *et al.* (2013) *Permeability*, *Vitality*, *Variety*, or equally in Bourdic *et al.* (2012) *Intensity, Distribution, Proximity, Connectivity, Complexity, Diversity, Form.* This research proposes that the first step in the classification of urban form characters follows the nature of the measure itself, which is captured in the *Index* part of its *Name.* On this ground, it builds on Bourdic *et al.* (2012) classification, adapting it to reflect the needs of a general analysis of urban form[5].

Hence, a classification firstly distinguished in the *Index* six categories that are *ontological* (they express the nature of the Index): *1. Dimension, 2. Shape, 3. Spatial distribution, 4. Intensity, 5. Connectivity, and 6. Diversity.* These six categories are in ranked order from the simplest (*1. Dimension*) to the most complex (*6. Diversity*). For example, *"Weighted Number of Intersections of Pedestrian Network"*, where the term *"Weighted Number of Intersections"* is the *Index* and *"Pedestrian Network"* is the *Element* will be classified as a character of Index category *"4. Intensity"*. The six categories are not purely independent, as we can identify functional relationships between them. Often characters in latter groups are mathematically dependent on others in the former: for example, those indexed by *Elongation*, which fall in the *"2. Shape"* category, are functionally dependent

[5]Research of Bourdic et al. (2012) focuses on measuring urban sustainability, and one of the categories is defined as 'form' which is refined into 'dimension' and 'shape', while 'proximity' is excluded as it is referring to non-morphological elements.

on those indexed by *Width* and *Length*, which fall into *"1. Dimension"*, since *Elongation = Width/Length* ratio. Also, the proposed system classifies the character into three categories that capture its *grain*– the scale of the spatial unit in which the unique value is stored. Finally, it distinguishes in the *Element* three categories that are *descriptive* of the scale at which the element itself occurs (the equivalent of the scale of spatial extent in figure 4.2), is observable and measurable in urban morphology.

Since many measurable urban form characters in literature work at multiple scales, the classification needs to maintain a certain level of breadth in defining the amplitude of scale. Therefore, it is proposing three *conceptual* levels of scale only: Small (S) representing the spatial extent of the building, plot, street or block (and similar), Medium (M) representing the scale of the sanctuary area (Mehaffy *et al.*, 2010), neighbourhood, walkable distance (5 or 10 minutes) or district (and similar) and Large (L), representing the city, urban area, metropolitan area or similar. Thus, to continue with the example, the character "*Weighted Number of Intersections of Pedestrian Network*", would be classified based on 1) its grain, and 2) the *scale* of its *Element* "Pedestrian Network". In this case, networks as physical entities occur and have meaning, and therefore can be observed and measured at the larger (M, L) scales, while they do mean very little at the small scale. Because the network, in this case, refers solely to pedestrian use, the urban form character falls into the category M of scale, or alternatively M/L if we allow more flexible cross-scale definition which might be desirable in general, as it softens the hard boundaries which might not be applicable to some, accounting for the authors' specific conceptualisation of spatial scale (such as Space Syntax). As this urban form character measures a single number per network, the scale of its grain and that of the extent of its *Element* coincide. However, that is not the case in all situations: for example, *Closeness Centrality of Street Network* is measured on the larger network (M, L scales), while the value is specific for each node (S scale)[6].

The resulting typology offers an unambiguous identification of each urban form character based on its very nature, as reflected in its name (table 4.2).

[6]Note that the extent in section 4.2 refers to the whole method used in each paper, in this section it refers to the spatial extent of single character only. The two concepts of extent are not the same.

Table 4.2: Table of Urban Form Characters (extract). A sample of measurable urban form characters, showing: definition of each category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale. The complete version of the Table, including all 361 urban form characters identified at this stage of the research, is provided as an Appendix 4.1.

category	definition	Index		Element	grain	extent	reference
dimension	*the basic geometrical dimensions of individual objects*	Length	*of*	Street	S	S	Dibble et al. (2017)
		Height		Building	S	S	Schirmer and Axhausen (2015)
		Bounding box area		Building	S	S	Schirmer and Axhausen (2015)
		Core area index		Building	S	S	Colaninno, Cladera and Pfeffer (2011)
		Number of floors		Building	S	S	Ye and Van Nes (2014)
		Mesh size		Grid network	M	M	Siksna (1997)
		Area		Built-up area	L	L	Seto and Fragkias (2005)
		Length		Urban edge	L	L	Boeing (2018a)
shape	*the mathematical features of geometrical dimensions of individual objects*	Height to width ratio	*of*	Street	S	S	Schirmer and Axhausen (2015)
		Compactness index		Plot	S	S	Schirmer and Axhausen (2015)
		Form factor		Building	S	S	Bourdic, Salat, Nowacki (2012)
		Fractal dimension		Axial map	M	M	Ariza-Villaverde et al. (2013)
		Rectangularity index		Sanctuary area	M	M	Schirmer and Axhausen (2015)
		Complexity index		Built-up area	L	L	Seto and Fragkias (2005)
spatial distribution	*the spatial distribution of objects in space and their reciprocal positioning*	Built Front Ratio	*of*	Block	S	S	Schirmer and Axhausen (2015)
		Solar orientation		Building	S	S	Gil et al. (2012)
		Distance		Buildings	S	S	Hijazi et al. (2016)
		Continuity		Built-up area	L	L	Galster et al. (2001)
		Concentration index		Built-up area	L	L	Gielen et al. (2017)
intensity	*the intensity of space occupation, referring to the density of elements within a set context*	Covered Area Ratio	*of*	Plot	S	S	Schirmer and Axhausen (2015)
		Floor Area Ratio		Block	S	S	Schirmer and Axhausen (2015)
		Number of plots		Accessible radius	S	M/L	Marcus, Berghauser Pont, Bobkova (2017)
		Weighted number of intersections		Street network	M	M	Araldi and Fusco (2019)

Table 4.2: Table of Urban Form Characters (extract). A sample of measurable urban form characters, showing: definition of each category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale. The complete version of the Table, including all 361 urban form characters identified at this stage of the research, is provided as an Appendix 4.1.

category	definition	Index		Element	grain	extent	reference
		Proportion of dead-ends		Street network	L	L	Boeing (2018a)
		Proportion of 4-way intersections		Street network	L	L	Boeing (2018a)
connectivity	*the spatial interconnection of the segments of the networks (usually street networks)*	Closeness centrality	*of*	Street network	S	M/L	Porta et al. (2006)
		PageRank		Street network	M	M	Boeing (2018a)
		Self-loop proportion		Street network	L	L	Boeing (2018a)
		Clustering Coefficient		Street network	L	L	Boeing (2018a)
		Node/edge connectivity		Street network	L	L	Boeing (2018b)
		Node connectivity		Street network	L	L	Boeing (2018b)
diverstiy	*the variety and richness of the elements and their characteristics in the study area*	Power law distribution	*of*	Blocks	M	M	Louf and Barthelemy (2014)
		Plot area heterogeneity		Sanctuary area	M	M	Feliciotti (2018)
		Plot area diversity		Accessible radius	S	M/L	Bobkova, Marcus and Berghauser Pont (2017)
		Intersection type proportion		Street network	M	M	Song and Popkin (2013)

4.4 Interpretation

The summative statistics of the complete Table of Urban Form Characters offers in-depth information into the current state of how terminology is defined and used in the field.

4.4.1 Distribution of characters

The distribution of characters across the scales of extent shows a slight decline as we proceed from Small to Large scales, but the distribution is relatively balanced (figure 4.3b). In terms of the scale of grain, it is naturally skewed towards Small scale (figure 4.3a). The situation changes if we explore the distribution of urban form characters among the 6 different categories established at the start of our classification. In this case, *spatial distribution* and *diversity* are underrepresented (with respectively 27 and 13 urban form characters), while all other categories each contain relatively high numbers each (from 55 in *connectivity* to 115 in *intensity*) (figure 4.3c). One of the reasons for this distribution is that *dimension, shape, intensity* and *connectivity* are much easier to capture than *spatial distribution* or *diversity* and their urban form characters are simpler to define.

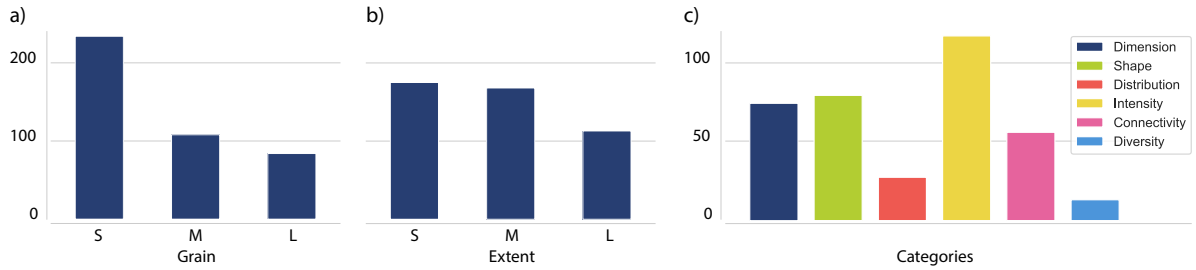


Figure 4.3: Number of urban form characters per scale of grain (a), the scale of the extent (b), number of characters per category (c). Note that some characters are present at multiple scales.

To understand the distribution of urban form characters in better intra-category detail, figure 4.4 shows decomposed statistics, which helped understand the relationship between categories and both definitions of scales. *Dimension* and *shape* categories tend to be significantly more present at the Small scale, from both perspectives. At this scale, physical features tend to be more precisely defined; hence it is natural that their dimensions and shapes are measured at the same scale. On the other side is *connectivity*, being present exclusively at larger scales (M, L) of extent, but skewed towards smaller scales of grain. This is an inherent consequence of the nature of this urban form character which is typical of networks, more comfortable to identify at larger scales of the environment in which they are observed, while the values are often unique for each component of network (as mentioned above).

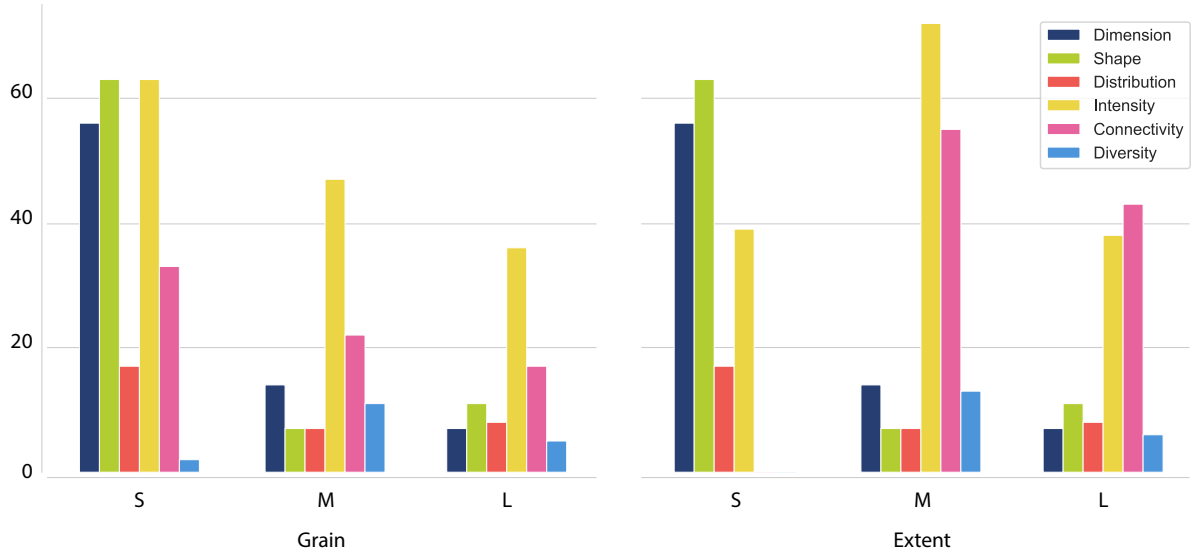


Figure 4.4: Number of urban form characters per each scale decomposed to each category. Note that some urban form characters are present at multiple scales.

The overview of urban form characters shows some clear recurring patterns from the perspective of (both) scales as well: it is worth noting that complex urban form characters are more likely to be measured at larger scales of extent (M, L). It seems to be partially caused by the nature of the classification system, where the limited amount of data inputs at a small scale makes results for more compound and the aggregated urban form characters less reliable. However, at the same time, this pattern is posing the question of whether the information is being missed out in this overview. Not even one of the six categories shows a balanced coverage of all three scales (for both grain and extent). It can be questioned which parts of the classification are less comprehensive for a logical reason (smaller scales are not suitable for complex relational urban form characters) and which are so just because some may have been missed out.

Back to the issue of *spatial distribution* and *diversity*, the former seems to differ across scales (the scales of grain and extent are identical for all urban form character in *spatial distribution* and, except for 2, in *diversity* as well). 17 out of 27 urban form characters in *spatial distribution* category are present at S scale. While the number is still lower than for the other groups (except *diversity*), the gap seems to be more significant at larger scales. The situation with *diversity* appears similar, featuring a majority of urban form

characters at M scale (15 in terms of grain and 17 in terms of extent), but the overall number is too low to conclude scalar dependency, even though such a tendency might be present.

An issue revealed by the proposed classification of urban form characters is the *overlap* and at times *redundancy* of some of them (the empirical correlation between urban form characters which makes some redundant). This is most evident among those capturing *shape* at the level of the block and below. Here a high number of such characters is utilised in the literature to capture the objects' geometry and form. Basaraner and Cetinkaya (2017) assessed the capacity of some of the urban form characters to capture the complexity in the shape of building footprints and concluded that only six out of 20 generally used are appropriate (p. 1972). Similar assessment should be done for other types, to rule out redundancy and increase the effectiveness and reliability of the fewer selected. On the other hand, the fact that certain types of urban form characters are abundant and might overlap or even lead to redundancy suggests that there is a general agreement on their value as descriptors of urban form.

4.4.2 Terminological inconsistency

Finally, terminological inconsistency could be explained by two causes. On the one hand, the current lack of a comprehensive framework for the systematisation and comparability of morphometric characters, on the other, the relative novelty of quantitative methods in urban morphology. There is, therefore, an urgent need for coherent terminology, as the amount of quantitative studies is expected to rise with the development of Geographic Information Systems (GIS), big data science, data mining as well as open data and volunteer-based mapping services. The problems of comparability of studies defined by Whitehand (2012) could be limited if a more rigorous typological system such as the *Index of Elements* proposed in this chapter was applied which would leave room for the interpretation of urban form characters, while making them comparable. In this regard, this work is dependent on the scope of existing research, and its validity is affected by the limits of the initial literature review. However, it could be argued that the method used to select papers ensures a reasonable level of representativeness as demonstrated by the fact that we were able to extract and successfully systematise 465 characters covering a significant number of measurements. The consequent systematisation exemplified in a tab.

4.2 and reported in full in the Appendix A4 seems to be inclusive and coherent enough to make sense of all of them, and yet this should be seen as just an initial framework. The proposed systematisation is meant to be refined and expanded as research progresses, in an open repository of tested urban form characters which would be ideally a collective product of the urban morphology scientific community as a whole. Moreover, the work could be expanded by the inclusion of other ways of conceptualisation of urban form, to cover land use or behavioural patterns (among others).

In reviewing the literature, it was necessary to rely on previously defined descriptions of characters. In several cases, these proved to be vague, sometimes lacking any definitions and/or mathematical formulas. Therefore, the classification of such characters might not align perfectly with the original source work. Even if it was able to classify all relevant characters successfully, it still might be possible to find in the future some that just do not fit into any of the six proposed categories (yet, it would still be possible to define it through the *Index or Elements* naming approach).

4.5 Summary

Quantitative approaches to urban morphology are critical to inform the long-overdue undertakings of a new "sciences of cities". The current state of the discipline is, however, to some degree, inconsistent. To make further progress, it is essential to understand what the limits and potentials of existing measuring methods are, and where the gaps of knowledge are.

The terminology used is often unclear, methods and urban form characters vary in ways that are at times difficult to understand. This limits the development of comparative studies, which however are essential to evidence-based research.

This chapter presented the first attempt at systematically and comprehensively organising existing measurable characters of urban form while overcoming terminological discrepancies. It collected a significant and representative sample of published literature and identified the main purposes of the research that underpinned it. From this sample, it extracted individual urban form characters capturing the physical structure of urban form and identified significant terminological inconsistencies ("nicknaming"), which were seen

as undermining the comparability of research outcomes across cases and methods. The chapter then introduced a new terminological framework based on an *Index of Element* approach, which then tested to redefine all the 465 urban form characters extracted from literature. As a part of a newly proposed conceptual typology, it organised them into six distinct and inclusive categories. The new framework allowed to identify a degree of redundancy in both the definition of urban form characters and their measurements, which led to producing a more rigorous set of final 361.

Analysis of how these urban form characters have been deployed identified a few anomalies in the distribution of their qualifying categories: the most significant tendency is the underrepresentation of *spatial distribution* and *diversity*. Moreover, *shape* and *dimension* are predominantly used at smaller scales, *connectivity* at larger scales (this tendency does not seem to be a consequence of the nature of the urban form character, but rather the lesser production of research on this topic).

Future research on the quantitative analysis of urban form, or urban morphometrics, should aim at collectively building a reasonably reliable and stable typology of measurable urban form characters, in order to achieve consistency across methods and case studies. Furthermore, the area should progress in recognising and measuring the full scalar and structural complexity of urban form, and we should be more comprehensive with regards to scales.

From the review within this chapter, it is clear that the state of urban morphometric is matured enough to provide a stable basis for numerical taxonomy. However, there are still issues to resolve.

Chapter 5

Research design statement

The following chapter bridges three previous background chapters and three subsequent core chapters. It provides a partial research design, formulates hypothesis and research questions and introduces case studies and general approach to the research driven by the reproducibility of the whole study.

In the first section, this chapter summarises the core findings from background chapters (2, 3, 4) and uses them to formulate the hypothesis, main research question and supplementary research questions provided in the second section. The third section introduces the principle of reproducibility into technical aspects of the work and provides an overview of an open-source software package design for morphometric research. The fourth section presents three case studies used within the remaining chapters of the work, and the last section outlines the contents of the three core chapters 6, 7 and 8 coming immediately after.

5.1 Synthesis of background chapters

The goal of chapters 2, 3 and 4 was to provide a necessary theoretical, conceptual and methodological background for the work proposed within this study. It is clear that the field of urban morphometrics is advancing in recent years at a fast pace. However, as shown in each of the chapters, there are still clear and wide gaps to fill.

Existing literature, as shown in chapter 2, offers various methods for classification of urban form. While approaches vary from simple duality *organised-unorganised* form based on remotely sensed data (Dogrusoz and Aksoy, 2007) to comprehensive small scale studies based on detailed manual digitisation of urban form elements (Dibble *et al.*, 2017), none of them is ideal. The optimal classification model, which consists of 7 principles (see section 2.1.2), is not fully reflected in any of the existing proposals (*problem 1*).

One of the potential pathways which could lead to better classification models is morphometrics and related numerical taxonomy stemming from the phenetic studies in biology. Conceptually, morphometrics itself is already well established in urban morphology (as shown in detail in chapter 4) and previous work of Dibble *et al.* (2017) showed the potential of application of numerical taxonomy as well. However, as shown in chapter 3, the existing proposals bridging the two are not fully operational on a large scale and unrestricted spatial units. The critical element, an operational taxonomic unit, needs to be revisited to reflect the nature of urban form better and allow an exhaustive analysis of metropolitan areas, rather than a specific carefully selected set of case studies (*problem 2*).

The related biological concept which may help with the issue of undefined OTU is a *mixture problem*. That arises when a taxonomist needs to identify populations within samples to perform classification on a population level then. The parallel in urban morphology is apparent. From a pool of fundamental units, i.e. buildings, plots and streets (Moudon, 1997), morphologist needs to identify distinct patterns of form and classify those into a taxonomy (*problem 3*).

Chapter 4 then provided a deep dive into urban morphometrics and existing methods of measuring of urban form. For the rest of the thesis, it is possible to draw three conclusions out of the chapter. First, although there is a rich pool of relevant studies, we lack those that are at the same time granular and extensive, while providing a comprehensive description of the form (*problem 4*). Second, the field has a nomenclature issue to tackle. The so-called *nicknaming issue* prevents comparability and brings a layer of inconsistency hard to declutter without a detailed decomposition of each individual method to the level of characters' formulas (*solution discussed in the chapter 4*). Last, urban morphometrics can measure an abundant number of characters covering all aspects of urban form. However, the focus is not very balanced, and some categories of characters require the

development of new approaches. At the same time, it forms a stable ground for numerical taxonomy of urban form.

5.2 Hypothesis & research questions

This work aims to fill some of the gaps identified in previous chapters and develop a method of classification of urban form which reflects *optimal classification model* and follows principles of morphometrics and numerical taxonomy. The scope of the work is primarily technical, focused on reproducibility and replicability of quantitative science of urban morphology, eventually leading to the establishment of the atlas of urban form. However, in its nature, the work is still exploratory. If the principles hold, the work should pave the way for more robust implementations. Furthermore, its main theoretical contribution lies in the revaluation of the bridge between numerical taxonomy and urban morphology previously proposed by Dibble *et al.* (2017).

5.2.1 Hypothesis

Drawing on the background knowledge presented in previous chapters, the main hypothesis behind the proposals laid in the rest of the thesis is then as follows:

> Methods of morphometrics and numerical taxonomy established in the classification of biological species can be applied in the context of urban morphology to lay the foundations of numerical taxonomy of urban form.

The rest of the thesis builds on the hypothesis and proposes an implementation of numerical taxonomy, which is later validated in chapter 8. The validation of the method then indirectly indicates whether the hypothesis is valid, which in turn tells whether the application of numerical taxonomy in urban morphology is a viable direction of research.

5.2.2 Research questions

The process of verification of the hypothesis needs to provide an answer to the main research question and a series of subsequent supplementary research questions.

5.2.2.1 Main research question

The main research question (RQ) focuses on the operationalisation of numerical taxonomy in the context of urban morphology:

> RQ: How to adapt methods of numerical taxonomy to study of urban form?

Answering the question gives us enough ground to understand whether such adaptation provides meaningful information for further analysis of the built environment. The task itself could then be subdivided into four supplementary research questions, which together outline the research proposal, eventually answering the RQ.

5.2.2.2 Supplementary research questions

All four supplementary research questions (SRQ) focus on the operationalisation of numerical taxonomy as a method based on urban morphometrics.

The first question focus on the identification of features which at the same time form urban fabric and could be studied using morphometric methods, and the connections between them:

> SRQ1: What are the fundamental morphometric elements and how to model their relationship?

Chapter 6 is dedicated to answering the question.

The second question builds on the previous and fills the gap identified in chapter 3. The question of OTU is critical as it directly affects the very nature of the taxonomy. The

question itself seeks an answer to both theoretical definition of the fundamental unit of urban form and the technical aspects of its delineation:

> SRQ2: What is the optimal Operational Taxonomic Unit of urban form and how to identify it in the continuous urban fabric?

Since an OTU is the result of a population delineation (see *mixture problem* in section 3.1.2.3), it is necessarily an aggregation of fundamental morphometric elements. The overview of aggregation models is available in chapter 6, together with a theoretical discussion and definition of OTU for this study. The delineation part is the content of chapter 7.

Morphometric assessment of any kind is based on morphometric characters, i.e. measurable aspects of elements used within the study. Their selection and implementation are the key drivers influencing the results of the analysis; hence it is necessary to give them enough attention and dedicate one SRQ to the topic:

> SRQ3: What are the taxonomic characters describing urban form?

The first part of chapter 7 aims to provide answers, building on the database collected in chapter 4.

The final SRQ focuses on the final aspect of the creation of numerical taxonomy, the quantification of similarity of OTUs:

> SRQ4: How to determine the taxonomic relationship between OTUs to derive taxa of urban form?

The method of creation of numerical taxonomy in the context of urban form and related answers to the last question is part of chapter 8.

5.3 momepy: Urban Morphology Measuring Toolkit

The content of this section was partially published in Fleischmann (2019).

Urban morphology is based on the analysis of space, traditionally mostly visual and qualitative (Dibble *et al.*, 2015); its objects are the fundamental elements of urban form (building, plot, street) (Moudon, 1997) as well as a range of analytical constructs such as axial maps (Ariza-Villaverde *et al.*, 2013) or proximity bands (Araldi and Fusco, 2019). The increased availability of morphological data and computational power have led in time to more emphasis on quantitative forms of analysis, and the emergence of urban morphometrics (Dibble *et al.*, 2017). Since morphometric analysis is addressed both in-depth and at large scale, it is grounded on the intensive use of GIS software (proprietary ArcGIS, MapInfo, open-source QGIS) either through built-in processing tools or specific plugins like Urban Network Analysis (Sevtsuk and Mekonnen, 2012) or Place Syntax Tool (Ståhle *et al.*, 2005). However, essential functions to conduct measurements of specific urban morphometric characters or tools to generate required geometry as axial maps or proximity bands are not always available: current plugins offer only a limited number of functionalities as they are mainly application or case-specific.

This thesis is hereby proposing momepy, a Python toolkit which aims to overcome such limitations by enabling a systematic in-depth analysis of urban form, to comprehensively include a wide range of measurable characters, with a prospect of expanding future development due to its open-source nature and independence on proprietary software or operating systems. The development of momepy is timely, as the role of measurable characters is vital to recognise form-based patterns and establish descriptive and analytical frameworks of human settlements, in the "age of urbanisation".

Momepy holds all morphometric algorithms used within this thesis (and some more) to simplify the adoption of urban morphometrics and allow easy reproducibility of the whole work presented in this study. Furthermore, momepy is designed to be more flexible than other toolkits as its functions are generally not restricted to specific morphological elements but to geometry types only and as such, can be used in various analytical models.

The six core modules of momepy represent six categories of urban morphometric characters: `dimension`, `shape`, spatial `distribution`, `intensity`, connectivity (`graph` module), and `diversity` identified in chapter 4. These six modules together provide a wide range of algorithms measuring different aspects of urban form and are able to describe its complexity with a significant degree of precision. Each of the characters tested or used in chapters 6 and 7 are included in the respective module. Additional modules help with

the generation of necessary morphometric elements, preprocessing of input data and other utilities.

Internally, momepy is built on the GeoPandas Python package (Jordahl *et al.*, 2020), using its GeoSeries and GeoDataFrame objects to store and handle large amounts of geospatial information. Under the hood uses PySAL (Rey and Anselin, 2007; Rey, 2019), mostly taking care of spatial weights matrices capturing the adjacency of elements of urban form. The graph module uses the capabilities of networkX (Hagberg *et al.*, 2008) for the analysis of urban street networks. Basic Python knowledge is required to use momepy, but exemplar Jupyter notebooks should provide enough information to allow using momepy with a standard GIS knowledge only.

Version 0.1. of the package was released in November 2019. Current version 0.3 (as of November 2020) was released in July 2020, and the software is being picked up by the research community (Mottelson and Venerandi, 2020). Thanks to the Journal of Open Source Software, the whole package is now peer-reviewed.

The reproducibility and replicability of research is a critical asset which should be ever-present, especially in the data-driven studies like this one. Momepy is seen as a fundamental component of the significance of this work as it enables further research and minimises barriers to urban morphometrics.

The software and its documentation are publicly available on momepy.org.

5.4 Case studies

The following chapters will present the work done on the three case studies - Prague, CZ, Amsterdam, NL and Zurich, CH.

The research focus of this work requires morphological richness in its case studies, which should capture various situations and assess the versatility of the proposed method. The second requirement is the availability of data representing selected morphometric elements of inconsistent quality and level of detail. Furthermore, it is advantageous if the researcher knows the place or has the ability to do a study visit to verify findings and interpretation on the ground.

The core case study used throughout the following chapters is Prague. Prague was selected as a case for its morphological richness of multiple historical layers. It was only mildly affected by bombing during the WW2, leaving the large part of the medieval core and other historical neighbourhoods intact. Furthermore, during the era of communism between 1948 and 1989, the city built its modernist belt, offering, together with the other areas, wide scope of urban patterns within a relatively compact extent. The study area is limited by the administrative boundary of the city (figure 5.1), which still extends the built-up area of the city by a large margin. That ensures the minimisation of the edge effect (Gil, 2016) which could otherwise adversely affect morphometric values. Prague's Institute provided the data used within this study for Planning and Development (www.geoportalpraha.cz) and further preprocessed using bespoke algorithms.

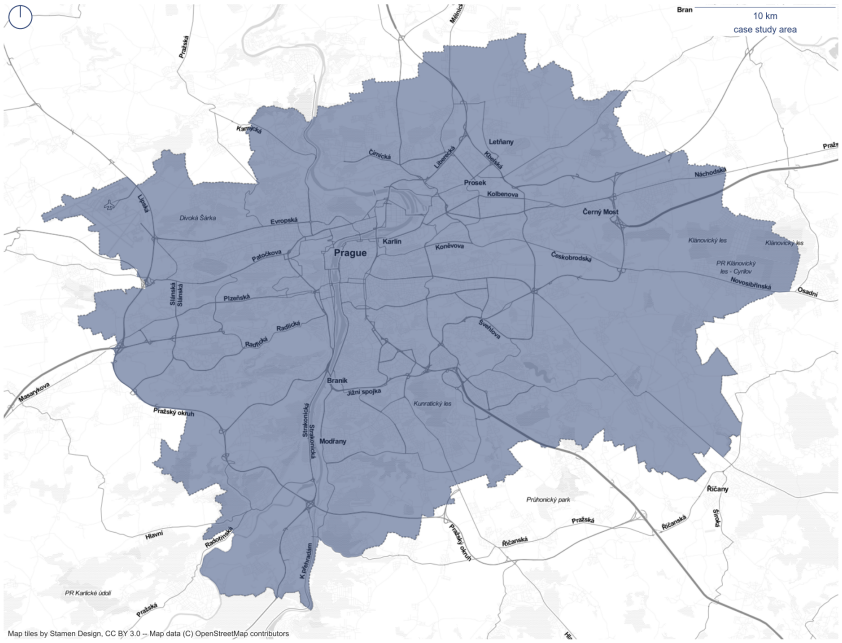


Figure 5.1: Prague case study area, which matches the administrative boundaries. Data source © IPR Praha, CC BY-SA 4.0

The second case study to which Prague will be compared in chapter 8 is Amsterdam. Both cities share some of the characteristics as both have (relatively) preserved medieval cores and a range of patterns of development spanning through historical periods. However, Amsterdam's planning context is very different. Prague is historically mercantile city on the river, which was never fully exploited by industry. On the other hand, Ams-

terdam is a port city with a high presence of port-related industry. Its historical centre is interlaced with artificial canals bringing different spatial order. Furthermore, the development of the second half of the 20th century and the beginning of 21st followed different planning paradigms that the one in Prague. The study area could not be limited to the administrative boundary as that does not reflect the morphology of a city. Instead, contiguous built-up land is used to avoid cutting through urban fabric (figure 5.2). The data are obtained from Dukai (2020) and Basisregistratie Grootschalige Topografie, BGT (http://data.nlextract.nl/). Road network was further preprocessed to eliminate dual carriageways, roundabouts and similar transport-focused features following the procedure proposed in Krenz (2018).

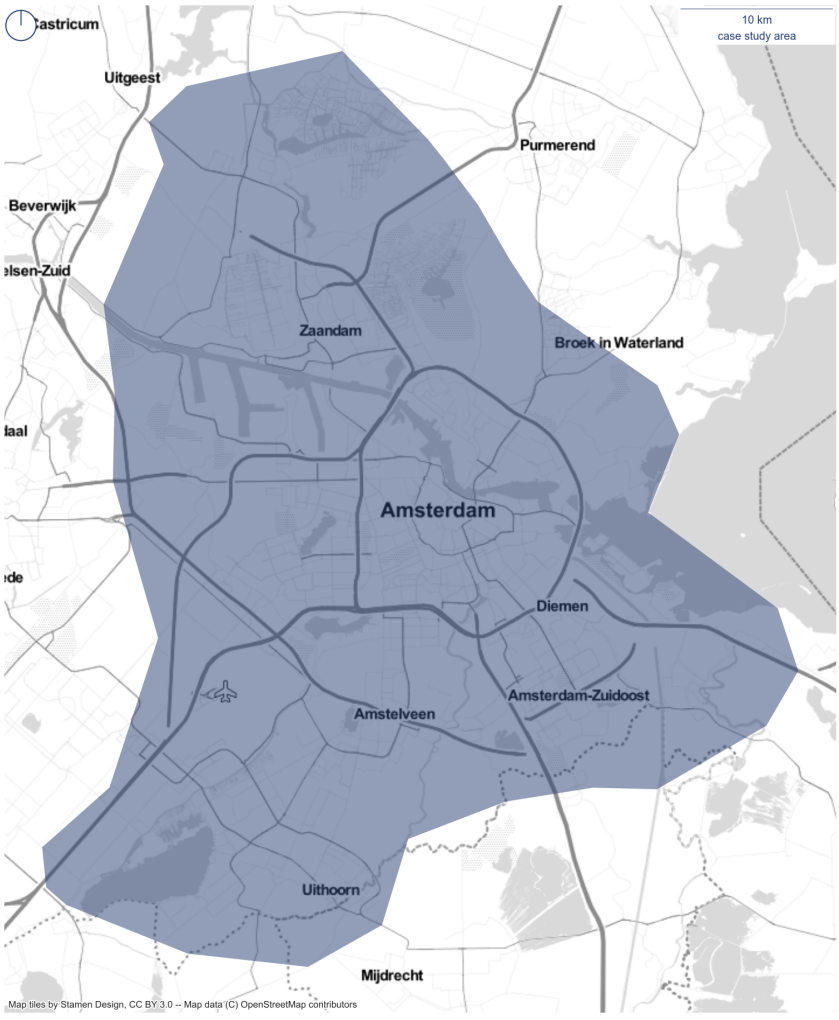


Figure 5.2: Amsterdam case study area, which follows continuous built-up area. Data source © BGT

For a one specific step requiring cadastral layer, Zurich, CH is used as a case study. The relevant section of chapter 6 has been published as an independent research paper (Fleischmann and Feliciotti *et al.*, 2020) prior specification of Prague and Amsterdam as case studies for numerical taxonomy as a boundary is used administrative area of Zurich (figure 5.3). Zurich was chosen for its historically characterised and heterogeneous urban fabric as well as for the availability of the '*Amtliche Vermessung*' dataset, a freely-accessible resource containing high-quality information on cadastral plots and building footprints.

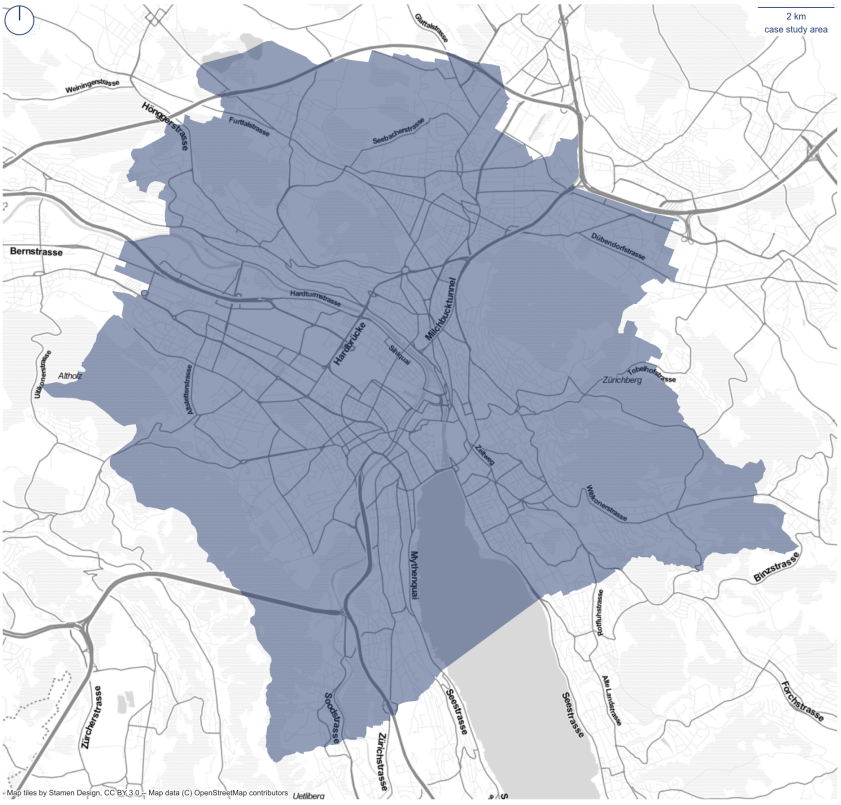


Figure 5.3: Zurich case study area, which matches the administrative boundaries. Data source © Kanton Zürich, CC BY-SA 4.0

5.5 Outline of the second part

The second (core) part of the thesis is formed of three chapters (6, 7, 8), followed by synthesis.

Chapter 6 studies morphometric elements of urban form. It provides an overview of fundamental ones and method of their aggregation, both area-based and location-based. It discusses in detail the issue of the smallest spatial unit, leading to the proposal and later test of morphological tessellation, Voronoi-based partitioning of space, as a basic unit of analysis. The topological aggregation capability of tessellation is further tested in comparison to conventionally used methods. Furthermore, chapter 6 proposes urban tissue as an operational taxonomic unit for numerical taxonomy and defines a relational framework of urban form, which will be later used for the morphometric characterisation of both main case studies.

Chapter 7 builds on the proposals and empirical tests presented in the previous chapter and presents a selection of morphometric characters of two kinds - primary and contextual. A complex quantitative description of urban form patterns linked to the level of individual buildings is then used as an input for cluster analysis identifying urban tissue types in the urban fabric in the form of homogenous contiguous clusters.

Chapter 8, the last core chapter then proposes a method of composition of numerical taxonomy based on urban tissue types identified in chapter 7. Furthermore, it extensively validates the whole method using additional data layers (historical origin, land use, municipal typology) and additional case study (Amsterdam). The final part of the chapter assesses extensibility of a taxonomy allowing further expansion of the database, eventually leading to the atlas of urban form.

The final chapter 9 synthesises the whole research, discusses the implications, limits and potential application of the proposed method, outlines potential further research and concludes the thesis.

Chapter 6

Morphometric elements of urban form

The content of this chapter was partially published in Fleischmann and Feliciotti et al. *(2020).*

The first half of the thesis presented the work done in the field of urban morphometrics to date, used as a basis for the formulation of hypothesis and research questions made in the previous chapter. The following chapter 6 is the first of the three chapters which focus on operationalising top-level ideas and answering questions posed in chapter 5.

The aim of this chapter is to lay the building blocks for morphometric analysis out. It focuses on morphometric elements of urban form, their relation to traditional morphological elements and range of possible issues from conceptual to practical ones. To overcome some of them, this chapter introduces the concept of morphological tessellation as the smallest spatial unit of analysis. The critical question which needs to be answered is the way of combining all elements together, as literature offers a range of potential frameworks, but few of them are explicitly defined, and minimum has the ability to link all elements and their aggregations together.

Structurally, the following chapter is divided into five major sections. Section 6.1 starts with the introduction of analytical representation of urban form, covering fundamental morphometric elements of urban form and their aggregations into larger features. The

specific focus is on the issues related to the plot as the smallest spatial unit and potential use of morphological tessellation. In its second part, the section follows with the overview of analytical frameworks (6.1.2), i.e. the ways how to combine elements and their aggregations into a singular schema. Section 6.2 builds on the previous in a proposal of a relational framework of urban form for comprehensive morphometric applications, relying on topology and morphological tessellation. To ensure that tessellation holds as a spatial unit, the rest of the chapter is dedicated to empirical experiments assessing its ability to capture the similar information as the plot and the potential of topological aggregation in comparison with other location-based methods. Hence section 6.3 outlines the method of testing and section 6.4 presents the results of the empirical study. The chapter concludes with a short discussion of presented results and paves the way to the application of relational framework and tessellation in chapter 7.

6.1 Analytical representation of urban form

This chapter aims to talk about the analytical, measurable representation of urban form. It does not attempt to build a new theory of urban morphology or re-conceptualise its understanding. It focuses on the operationalisation of the analysis within the context of urban morphometrics. For that, we need to understand two aspects - what are the basic components of urban form we can capture in relevant data, and how they relate to each other. In other words, what are the morphometric elements of urban form and how to work with them in an analytical framework.

6.1.1 Elements of urban form in the context of urban morphometrics

Morphological theory talks about a wide variety of elements, but only some are seen as fundamental. Furthermore, we have to take into account the issue of data availability - since high-quality data representing urban form are not always available, we need to use as a small number of resources as possible to keep any morphometric method widely applicable. For this reason, this research attempts to work with the minimal input needed - fundamental elements only.

6.1.1.1 Fundamental morphometric elements

Urban morphology knows three fundamental elements of urban form. Both (Moudon, 1997) and later (Kropf, 2017) agree on buildings, streets and plots as features representing three fundamental aspects of human behaviour: *sleep*, happening in a shelter, i.e. building, *movement*, first reflected in tracks and today in streets and roads and *local activity*, initially happening in a core territory, which can be today seen as a plot (Kropf, 2017).

Fundamental *morphometric* elements are based on fundamental morphological elements, and their goal is to reflect the same phenomena. However, the difference between morphological and morphometric elements is that the latter are operational, measurable and represented as (usually vector) GIS data. Hence we can talk about three layers of data: buildings (either as footprints or models), street networks and plots as smallest spatial units.

It is essential to ensure that the data are good enough to represent morphometric elements. That could be an issue for all types of elements, so there are cases when the data needs to be prepared for morphometric analysis. The pre-processing can be in some cases automatised, in other, unfortunately, manual or at least semi-manual to have the data in a correct form in the end.

Whilst each dataset coming from a different source is specific; hence the cleaning procedure needs to be tailored to each source, there are some common issues which are not unique to specific datasets. The following section presents all fundamental morphometric elements and outlines these common issues and ways of resolving them or at least minimising the error under a significant level. We cannot expect data to be perfect all the time (they are never perfect). This limitation has to be taken into account during the design of any morphometric analysis, which needs a certain level of robustness to accommodate potentially erroneous data.

6.1.1.1.1 Buildings Buildings can be represented as footprints, i.e. two-dimensional projections of building shape to the ground or 3D models. Somewhere in between lies what we can call the 2.5D model, which is a building footprint with an attribute of a building height. All the options are being used in morphometric research. However, since one of the key goals of the method being developed within this work is wide applicability,

it cannot depend on 3D representation due to its limited availability. The most accessible representation of buildings are vector-based footprints (figure 6.1), either generated by official national or municipal mapping agencies (like Ordnance Survey in Great Britain) or by volunteers within OpenStreetMap crowdsourcing mapping movement (Haklay and Weber, 2008). Hence this research depends on 2.5D vector representation, providing a balance between availability and precision.

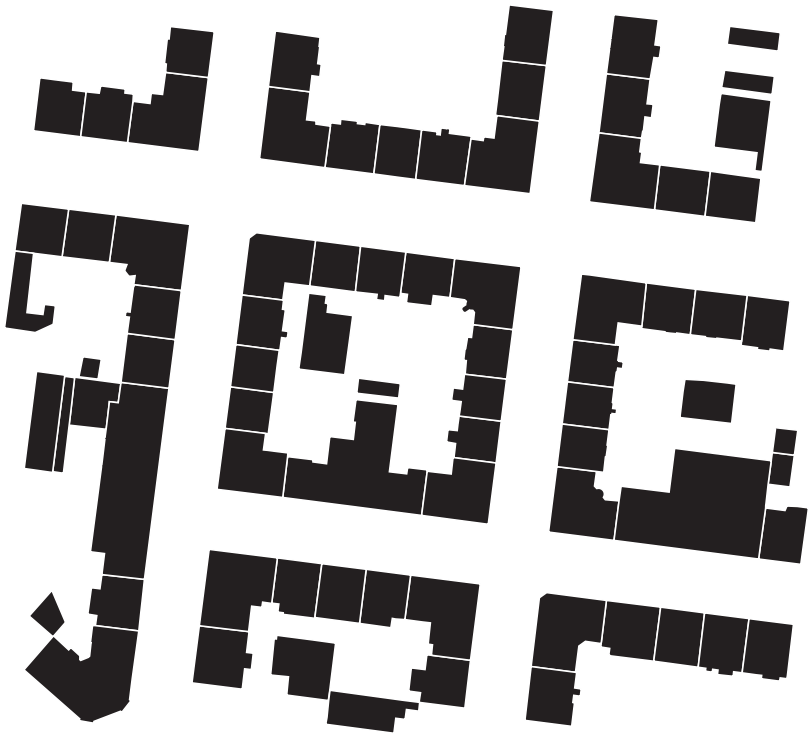


Figure 6.1: Building footprints in the optimal resolution and data quality.

Having data layer representing building footprints correctly and consistently is the first condition for successful morphometric analysis. There are several aspects which need to be fulfilled - topological correctness, consistency in detail, representation of individual buildings and building height attribute presence. Overall, it is expected to have a building data representing Level of Detail 1 (LoD1) (Biljecki *et al.*, 2016) (figure 6.2).

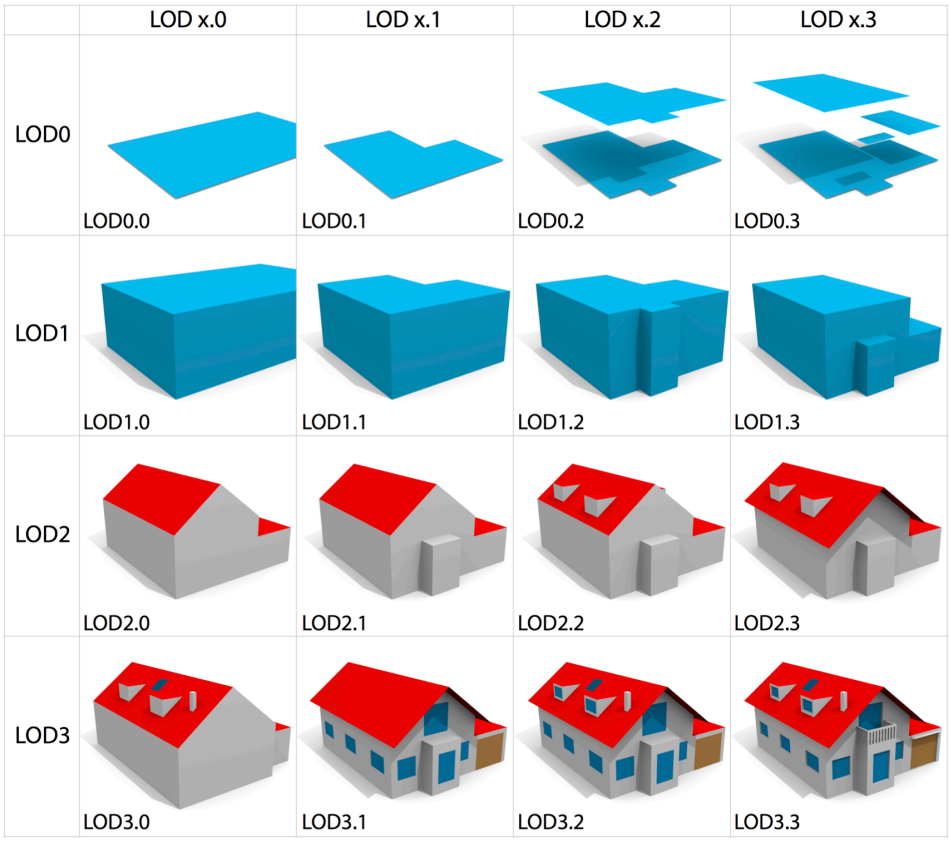


Figure 6.2: Diagram of LoD classification on the example of single-family housing. (Biljecki et al. 2016, figure 3)

Topological correctness ensures that geometry represents the relationship between buildings on the ground. There are characters measuring continuity of a perceived wall in a joined buildings or shared walls ratio which require building polygons to be correctly snapped together when two buildings touch. In that case, it is expected that neighbouring polygons will share vertices and boundary segments. There should not be a gap between polygons when there is none in reality. Also, polygons must not overlap at any case as that could cause significant disruption of analysis.

The building detail should be consistent across the dataset and represent an optimal approximation of building shape based on LOD specification as proposed by Biljecki *et al.* (2016). The approximation should represent LOD1.1 (no details, but the shape is kept) or LOD 1.2 (minor details), building shapes should not be overly detailed nor overly simplified. In the case of inconsistency, simplification of more detailed shapes should be

done before morphometric assessment.

Each polygon should represent a single building. There are datasets (often of a remote sensing origin) capturing all structures which are joined by any means as a single polygon. Such data do not represent the morphological truth on the ground. Their pre-processing is complicated as it requires splitting of existing geometries based on an additional dataset. The second extreme is the opposite situation when multiple polygons represent a single building. These usually represent different height levels, through routes or similar features. If these polygons, representing parts of buildings, have a common ID which allows joining them together to get a single polygon representing a single building, the pre-processing of such a data is only a simple dissolution. However, there are many cases when this ID is missing, and correct pre-processing require either clever heuristics to understand which polygon belongs to which or laborious manual work.

A number of morphometric characters use building height attribute, which, in that case, has to be present in the original input dataset. The resolution should be able to capture the distinction between floor levels.

6.1.1.1.2 Street network Street network represents a street, but that itself can be, form the data perspective, defined as multiple features. We can understand it as a movement, as an area or as a network. Furthermore, an important aspect of a street (network) is also a junction - node relationship. Therefore, within this study, a street is represented as a street network consisting of edges representing centrelines and nodes representing junctions. Both can be abstracted to a simple LineString (polyline) geometry and its configurational graph representation.

Figure 6.3: Illustration of street network represented as street centrelines, from which can be derived both street edge and node.

The similar pre-processing situation as with building layer is with a street network. Incorrectly drawn street network may cause significant errors in morphometric results and consequently in taxonomy. There are three critical cases which need to be checked before the analysis - topological correctness, morphological correctness and consistency in classification.

Topological correctness ensures that each street segments is represented by a single `LineString` geometry, that neighbouring segments share end vertex and that geometry is not split if the segments intersect only on the projected plane and not in reality (typically multi-level communications, when one is on the bridge across the other so that projected intersection is not a real intersection).

Moreover, street networks have to be morphologically correct, which means that geometries represent morphological connections, not other, usually transport-focused elements. That often mean simplifications of networks to eliminate transport geometries like roundabouts or similar types of junctions, or dual lines representing dual carriageways. In

certain cases, networks have to be snapped together, because due to traffic-calming measures, some junctions might not be connected when they should be (morphologically).

Finally, the network needs to be consistently drawn in terms of inclusion of different levels of network hierarchy. Hence the definition of what is considered a street and what is a minor pedestrian connection is crucial and needs to be consistent throughout the study.

Subject to data availability, networks are widely available. However, geometries mostly represent transport network and often do not follow ideal topological rules. The preprocessing to ensure that all three points above are fulfilled is hence necessary and can be partially automatised either using momepy or using methodology outlined by Krenz (2018), using conventional GIS tools. However, there might be cases when more complicated procedures should be employed, either to provide a more accurate algorithm or to include manual steps.

6.1.1.1.3 Spatial unit The identification of a reliable, significant and universal spatial unit of analysis is of crucial importance. However, the situation with a traditionally used plot is complicated.

6.1.1.1.3.1 Plot In traditional Urban Morphology, a *plot*, is considered to be the smallest meaningful unit of spatial subdivision and a fundamental component to understanding the spatial structure of the ordinary fabric of urban settlements (Moudon, 1997; Panerai *et al.*, 2004; Porta and Romice, 2014) and their processes of formation and transformation in time (Whitehand, 1981).

However, despite its significance, the plot remains a problematic construct. At *ontological* level, there is no agreement on *what exactly a plot is*: indeed, it has been variously defined as "*a land-use unit defined by boundaries on the ground*" (Conzen, 1969, p. 128), *a module* of the urban tissue constituted by a built-up area and its open pertinent area (Caniggia and Maffei, 1979), a piece of property, subject to subdivision and amalgamation as a result of successive patterns of occupation (Moudon, 1986), or again, according to Bobkova *et al.* (2017), as "*a basic unit of control*", "*a fundamental link between spatial and non-spatial medium*", "*a connection between built space and space of movement*" and "*the framework for building evolution over time*" (p. 47.5). Furthermore, crucially, more often than not,

these definitions may represent very different entities on the ground "*potentially leading to misinterpretation and so a somewhat obscured picture of the dynamics of urban form*" (Kropf, 1997, p. 1).

The second problem has to do with the relevance and applicability of the plot to different urban contexts. In literature, plots have been predominantly used to study and characterise traditional urban tissues that having evolved incrementally at the plot level (Conzen, 1969; Bobkova *et al.*, 2017), are quintessentially plot-based (Panerai *et al.*, 2004). It is, however, not the case for urban forms that came about after the Second World War, which appear to respond to substantially different rules of the organisation (Dibble, 2016; Feliciotti, 2018). For these tissues, "*plots no longer have a structuring role*" (Levy, 1999, p. 83), and hence can hardly be a suitable unit of analysis. While the process of identifying plots in traditional tissues is somewhat less controversial, the same is not true in contemporary ones (figure 6.4).

Figure 6.4: Comparison of traditional (left) and modernist (right) urban tissues in Glasgow. Plots are clearly better identifiable—even just visually—in the former, where distinction of public and private space is clear-cut, than in a modernist housing estate, where the transition between public and private is blurred. Source: Ordnance Survey MasterMap, January 2019 (EDINA Digimap Service)

In addition to this issue, the identification of plots in the urban fabric also poses a series

of *analytical* problems: given a map or a satellite image, how to determine the plot boundaries consistently? Moreover, in the case of existing datasets, what do they *actually* represent? What definition of plots do they adopt? Are different datasets comparable?

Not all mapping agencies explicitly report plots and, even when they do, not all of them define or represent plots in the same way. In some spatial databases, as in the Swiss Katasterwesens, plots are represented as a unitary land parcel, whilst in other cases, ownership-based plots can be made of multiple unlinked features, as in the French Cadastre, limiting comparability between different datasets. In other cases, the identification of plots from available sources is inferred by the analyst via resource-intensive manual interpretation. However, that makes the resulting procedure on the one hand unsuitable for large scale analysis, and on the other potentially biased, as heavily dependent on both individual interpretation and the often uneven quality of the underlying data. Indeed, while through open-data policies (Huijboom and Van den Broek, 2011) and Voluntary Geographical Information System (VGIS) (Barrington-Leigh and Millard-Ball, 2017) the availability of free-to-use geo-data is growing dramatically, their quality, coverage and resolution are often insufficient to determine individual plots and generally limited to building footprints, street centrelines, natural features and administrative boundaries. All of this reduces the reliability of the analysis and the universality of its results considerably.

Given the aforementioned issues, and despite plots being still widely used in urban morphology to capture the "*pattern of human intention and activity*" (Kropf, 1997, p. 5), they are ill-suited as a basic unit for morphometric applications.

6.1.1.1.3.2 Morphological tessellation One of the few alternatives of plots proposed in the literature (Hamaina *et al.*, 2012, 2013; Schirmer and Axhausen, 2015, 2019) is morphological tessellation (MT). A method of deriving a spatial unit of analysis, the morphological cell (MC), which is able to convey reliable, universal and meaningful plot-scale information and, at the same time, to minimise manual labour, subjective interpretation and data dependence. Hence it is proposed to use morphological tessellation as a spatial unit instead of the plot in urban morphometrics.

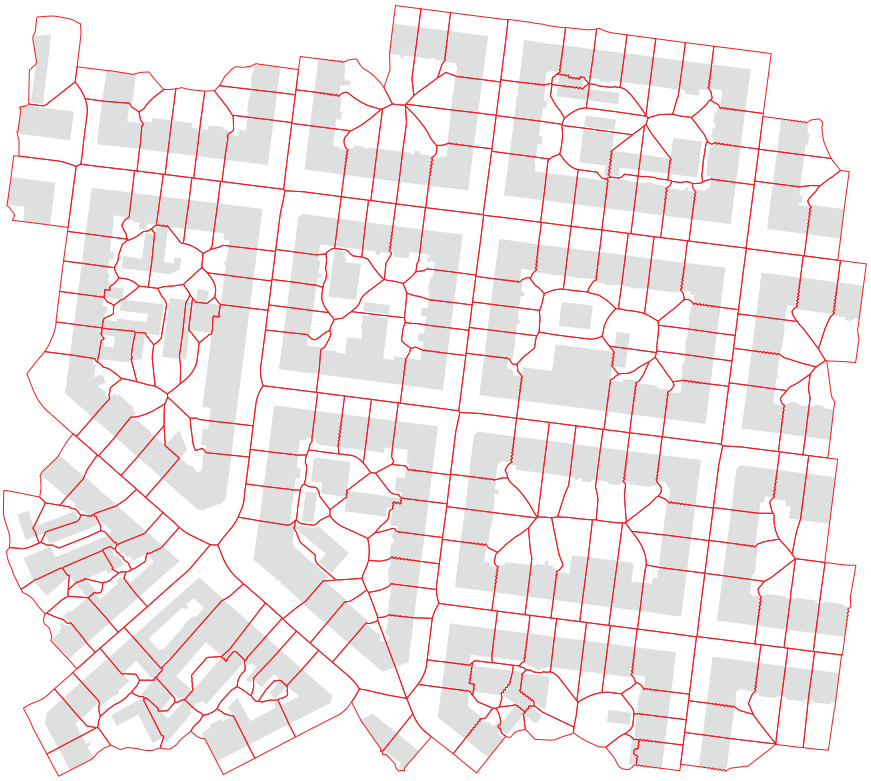


Figure 6.5: Illustration of morphological tessellation as the smallest spatial unit with the ability to partially replace the plot while being fully derived from building footprint data.

At the core of the proposed implementation of MT lies the *Voronoi tessellation* (VT), a method of geometric partitioning that from a planar set of *'seeds'* generates a series of polygons, known as *Voronoi Cells* (VC). Each Voronoi cell encloses the portion of the plane that is closer to its seed than to any other (figure 6.6), representing its '*influence zone*'[1].

[1]The term Voronoi Tessellation can be used to describe both the process of partitioning space (method) and the geometric mesh it generates (output). In this text, the two meanings are used interchangeably.

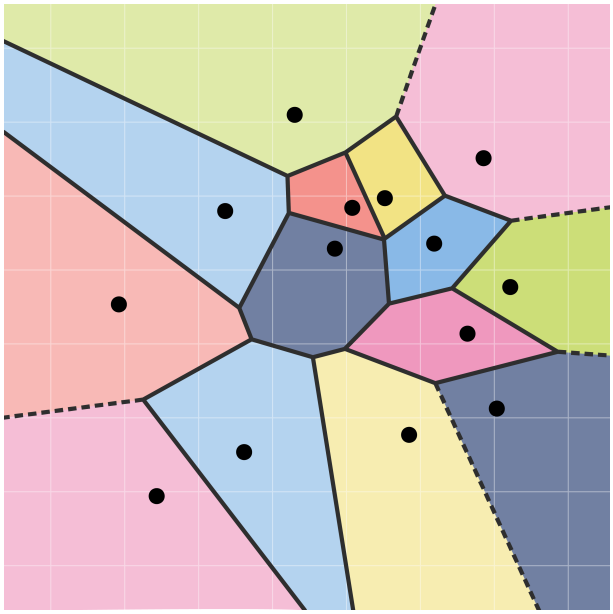


Figure 6.6: Voronoi tessellation based on randomised seeds. Each colour represents the area of one tessellation cell (influence zone). Dashed lines end in infinity

Voronoi tessellation has been already used in relation to urban form, in the context of spatial clustering algorithms (Dogrusoz and Aksoy, 2007) and built-form geometry generalisation techniques (Basaraner and Selcuk, 2004; Li *et al.*, 2004; Ai and Zhang, 2007; Liu *et al.*, 2014), or as input for image-based pattern recognition (Yu *et al.*, 2017). In recent years, morphological tessellation was used to study the micro-scale properties of the urban fabric (Hamaina *et al.*, 2012, 2013) in order to produce a reliable method for urban form patterns' recognition, which pioneered the generation of VC from building footprints. Later, Schirmer and Axhausen (2015, 2019) devised a method to define "*influence zones*" around buildings using a "*topological skeleton*" of unbuilt space that is mathematically similar to morphological tessellation. In parallel, Usui and Asami (2013, 2017, 2019) included the street network as an additional input alongside the building footprint to the Voronoi tessellation algorithm, to mimic the plot structure of traditional Japanese urban fabrics. Whilst the generated mesh shows remarkable similarity to the plot pattern, its main limitation is the inability to capture the spatial pattern of modernist (post-WWII) urban tissues and the highly variable distance between building and street that is typical of such fabrics. On a similar vein, Araldi and Fusco (2017, 2019) developed an approach based on Voronoi tessellation and street segments to define a spatial unit based on the

pedestrian point of view.

In all these cases, the use of Voronoi tessellation helped to rigorously and reliably cluster components according to their configuration although, as pointed out by Usui and Asami (2019), the relationship between morphological cell and 'conventional' plots has never been directly tested to date. In this sense, the morphological tessellation approach is to be intended as a continuation of this line of works, and insofar it too utilises the Voronoi tessellation procedure. However, unlike previous studies, this research aims to provide a fully operational and replicable method by examining the details of the tessellation process and its parameters and testing the similarity of morphometric characters as measured on both morphological cells and plots through direct comparison.

6.1.1.2 Analytical aggregations

Since morphological analysis aims to capture patterns of urban form, it must describe single elements as well as their spatial configurations and relationships. Therefore, larger analytical units have to be identified. Generally, we can distinguish two approaches of aggregation of fundamental elements into larger units: area-based and location-based (Berghauser Pont and Marcus, 2014).

6.1.1.2.1 Area-based Area-based approaches divide space into preselected units, i.e. administrative boundaries (Gielen *et al.*, 2017), abstract projected boundaries (grid) (Galster *et al.*, 2001), or larger morphological structures such as a block (Gil *et al.*, 2012) or a Sanctuary Area (Dibble *et al.*, 2017). However, such methods may face two connected issues, together named "Modifiable Areal Unit Problem" (MAUP) (Openshaw, 1984): scale issue (how big the area of aggregation should be) and aggregation issue (where should we draw its boundaries). Specific non-morphological area-based approaches are prone to both of them, particularly the latter: a change of the boundary, for example, the voting district, might affect the analysis' results.

However, there is still a significant scope to study morphological aggregations. Starting from the smallest plot scale, we recognise street edge, block, street, up to either urban tissue if we follow one definition (Kropf, 2017) or in sanctuary area if we follow another (Mehaffy *et al.*, 2010; Feliciotti *et al.*, 2016).

The smallest morphological aggregation of plots is a street edge, which is a "*series of one or more plots served by the same street*" (Feliciotti *et al.*, 2016, p. 5). To resolve the issue of corner plots, Feliciotti *et al.* (2016) add "*bound to the centrality of the street is sits on*", which can be translated as the importance of each street. A corner building simply belongs to the more important street of the two. More massive related aggregation is a block, or what Kropf (2017) calls a plot series. Block can be defined "*an aggregate of plots surrounded on all sides by street spaces*" (Kropf, 2017, p. 47). Both street edges and blocks are combinations of plots, limited by a street network and as such, both are relatively easy to define even algorithmically given appropriate data.

On the larger scale, area-based aggregations result either in Sanctuary area (Mehaffy *et al.*, 2010; Feliciotti *et al.*, 2016; Dibble *et al.*, 2017) or urban tissue, morphological region or another homogeneity-based structure (e.g. urban structural unit, character area). Sub-optimality of Sanctuary area has been explained in chapter 3 and hence is not scrutinised again here.

Urban tissue (figure 6.7) and related concepts, on the other hand, are worth considering as they repeatedly emerge from various schools of urban morphology. As a *morphological region*, or *plan unit*, it has a prominent role in Conzenian historic-geographical approach (Oliveira and Yaygin, 2020). While different terms often capture different concepts, the underlying logic is always the same. It is well summarised in Kropf's definition of urban tissue: "*a distinct area of a settlement in all three dimensions, characterised by a unique combination of streets, blocks/plot series, plots, buildings, structures and materials and usually the result of a distinct process of formation at a particular time or period.*" (Kropf, 2017, p. 89)

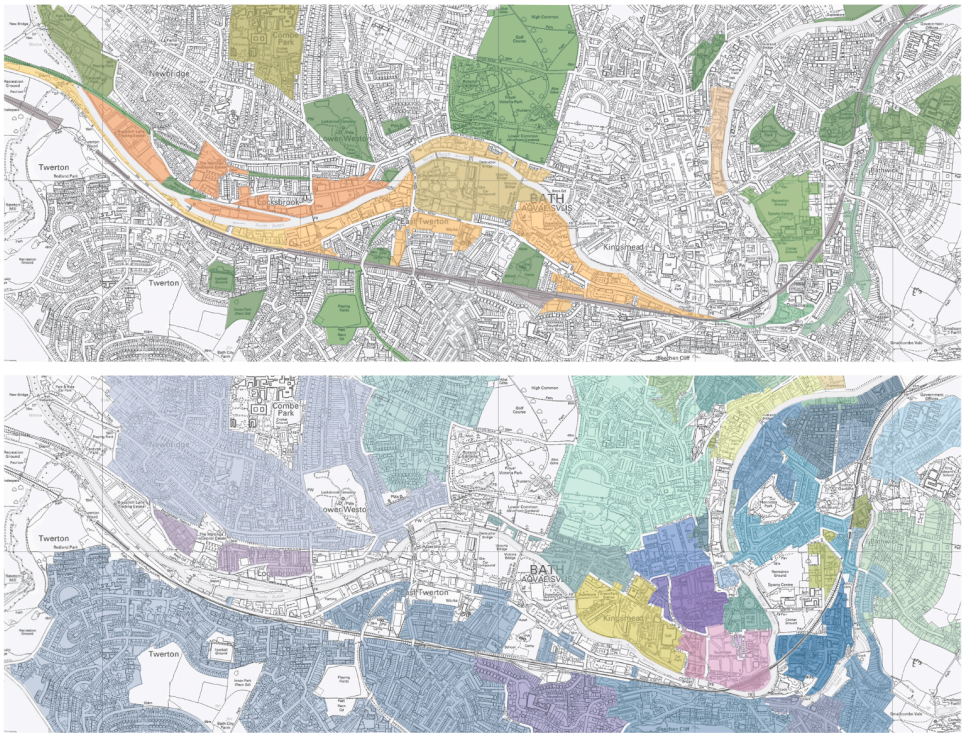


Figure 6.7: Urban tissues in the fringe (top) and central areas (bottom) of Bath identified by Kropf 2017 (figure 8.4)

While some of the area-based aggregations, mostly those following morphological nature of form (edge, block, tissue), are entirely meaningful and minimise MAUP, others are not ideal and should be generally avoided (administrative units, census blocks, voting districts).

6.1.1.2.2 Location-based Location-based approaches generate analytical units independently for each source-element as a unique aggregation around it, typically at walking or driving distance, where distance is measured either along with the street network (network distance) or an approximation of it (for example as the crow flies). Therefore, the aggregated values are uniquely and consistently generated for each source-element (e.g. building), and the effect of arbitrary data aggregation is minimised, resolving MAUP's aggregation issue. For this reason, literature prefers location-based analytical units, as their nature partially resolves MAUP. The scale problem part of MAUP is present also in location-based methods, and it is up to the methodology adopted on a case-by-case base

to limit its effect to a minimum.

Currently, morphological literature relies on a few methods to define an aggregation of units in a location-based manner. The most straightforward is based on simple Euclidean (as-crow-flies) distance from the elements of analysis (typically a radius of 400 metres around a building) (Schirmer and Axhausen, 2015). However, such an approach does not reflect the actual morphological situation on the ground. In some instances, as in traditional compact urban tissues, it can capture hundreds of buildings within 400 metres. However, only a few in sparse modernist urban tissues, leading to fundamental differences in the amount of information captured, causing issues of comparability of such information.

Excluding the effect of specific tissue types from the definition of aggregations overcomes method based on reach. Following the street network or axial map of urban form (Berghauser Pont and Marcus, 2014; Marcus *et al.*, 2017), it captures the area which is possible to reach within a set distance (mostly metric). As a location-based method, reach is useful because it reacts to unequal morphologies, but only through constraints that limit access to space, rather than through detection of a difference in urban form itself. The logic is based on the cognitive experience of cities but limited to accessible open spaces, excluding the intra-block relationships. It can generate situations of two buildings facing each other across the block (hence directly influencing each other) not being aggregated together, ignoring their relationship. Furthermore, unlike any other method, network-based reach adds a requirement of street network data input limiting its applicability, e.g. in the context of remote sensing-based building footprints in the informal context where no street network is available.

Both Euclidean distance and metric reach methods cannot capture the change in the granularity of urban tissues, hence are effectively measuring different information in granular and sparse tissues. In the case of reach, the distance could be defined topologically as a number of steps on the network (represented by a graph structure) (Berghauser Pont and Marcus, 2014), allowing to recognise the change in the pattern of aggregations, but it still does not eliminate the issue of intra-block relationships. On top of that, network-based methods face issues in data availability - street networks usually need significant adaptations before they can be used, as they are typically drawn for traffic purposes, not morphological ones; axial maps are scarcely available, and their generation needs a very

specialised type of morphological knowledge limiting the applicability of such method.

The third method present in literature is K-nearest neighbour (KNN) analysis, which is based on Euclidean distance, but it is using it differently. It defines an aggregation as a set number of nearest neighbours, defined via as-crow-flies distance. Whilst only scarcely used in urban morphology (Liqiang *et al.*, 2013), it has potential as such an approach might reflect changes in the granularity of urban tissues. However, due to the Euclidean definition of nearest neighbours, it cannot react to the detail of some spatial configurations (e.g., be able to detect linear patterns with natural boundaries between as features across the boundary might be closer than those within the pattern). Theoretically, KNN could be used together with reach analysis, joining both the ability to capture morphology represented by networks and scalability of KNN, but this concept has not been applied in morphological literature so far. However, it would still not resolve the issue of intra-block relationship.

It should be pointed out that the morphological cells do offer added values that are relevant on their own in the context of aggregation, regardless of their similarity to the plots. These have to do with the potential innovations – yet largely unexplored – which are triggered by the very nature of this geometry. For example, unlike other methods of urban form partitioning, the morphological tessellation covers the totality of space uniformly within the set study area, allowing to capture the topology of contiguous space at the plot-level. Indeed, since all MCs are determined by adjacency, by using morphological tessellation it is possible to think in terms of topological distance (set number of topological steps between cells) rather than geographic distance (set metric distance around elements, either "as the crow flies" or along with the street network). Moreover, thinking in terms of topological distance as opposed to metric, the morphological tessellation can be used to define new aggregated analytical units that are able to capture the immediate area of influence of a building on its surrounding fabric and, at the same time, of the surrounding fabric on the building. Indeed, since the size of each morphological cells depends on the granularity of the urban structure, the spatial representation of a set topological distance would be far smaller for a morphological cell located in a fine-grained built-up area than for the same located in a coarse one (figure 6.8). Crucially, this is a kind of information that would not be possible to access with plots alone, which allows for reframing the very idea of 'proximity' by rethinking the relationship between scale and spatial meaning, thereby enhancing the ability to capture the context in morphometric analysis.

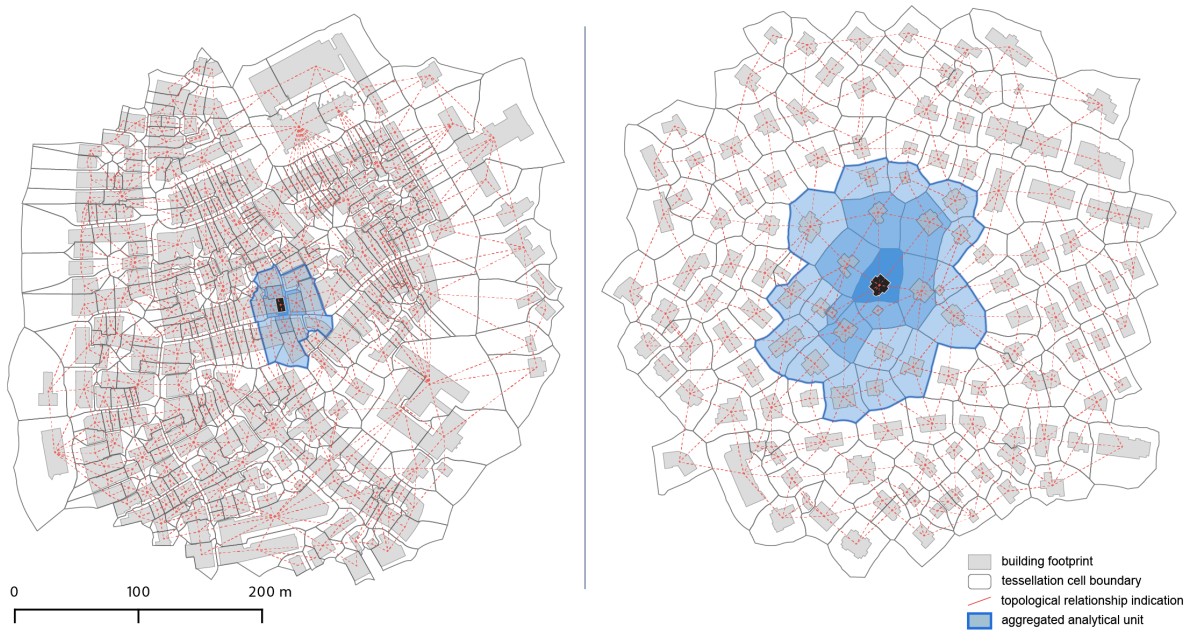


Figure 6.8: Relationship between morphological cells of topological distance 2: the red geometry represents the adjacency network of neighbouring elements (buildings, morphological cellss) at topological distance 1 (adjacent neighbour of first order), while blue geometry represents the boundary of the aggregated analytical unit of topological distance 2 for each of the highlighted buildings. In the image, a fabric characterised by fewer and sparser buildings (b) generate larger cells and aggregated units compared to a denser and more compact fabric (a).

Topology captures the information on adjacency of neighbouring elements (cells) - two cells are neighbouring if they share at least one point (so-called Queen contiguity) or one segment (so-called Rook contiguity). It defines the proximity of elements in terms of the number of steps needed to get from each element A to each element B. Topological relationships can be of two types - unconstrained, if not limited by any other element than tessellation itself, and constrained if the step between two neighbours is impeded by constraint (a block is the maximum number of topological steps from element without the need to cross the street network, while the street network is the constraint in this case). Thus, we can define an aggregation around each element based on several topological steps (topological reach) on the morphological tessellation, where aggregation defined by n steps includes all morphological cells which we can reach within x <= n steps.

6.1.1.3 Operational Taxonomic Unit as a morphological aggregation

Having defined aggregation models, the question of Operational Taxonomic Unit arises again. It was noted that definition of OTU in for numerical taxonomy of urban form follows in principle *mixture problem* (see section 3.1.2.3), meaning that an OTU is necessarily an aggregation of fundamental elements. Moreover, it has to be area-based aggregation to avoid overlaps of elements brought by location-based techniques.

Assuming that the initial pool of fundamental elements is a *mixture* of separate "*populations*" which need to be identified, the population is a group of individuals of a single species. Even though the term species is abstract in urban morphology, its phenetic definition is very much applicable. Following Sneath and Sokal (1973), species is "*the smallest (most homogenous) cluster that can be recognised upon some given criterion as being distinct from other clusters* (p.365). That definition is conceptually the same as the definition of urban tissue in urban morphology as both are primarily based on the distinctiveness of each group identified either as species or as a tissue. However, a species is a taxon; therefore, its counterpart would be urban tissue type.

Therefore, within the framework of a mixture problem, we can consider urban tissue type an Operational Taxonomic Unit of numerical taxonomy of urban form.

6.1.2 Analytical frameworks of urban form

Analytical frameworks of urban form are conceptual schemas linking fundamental elements together for the purpose of morphological analysis. The way we link fundamental elements and their aggregation's matters and frameworks are often not specified in literature, just assumed.

The literature mostly offers frameworks which are hierarchical in nature, meaning that on a single level, each element can be part of a single aggregation. One of them is a framework used by Feliciotti *et al.* (2016) and Dibble *et al.* (2017) (figure 6.9) which stacks elements together to form Sanctuary areas. A similar concept is used by Kropf (2017) in his multi-level diagram of built form (figure 6.10), which itself is based on older

Canniggian theory (Caniggia and Maffei, 1979).

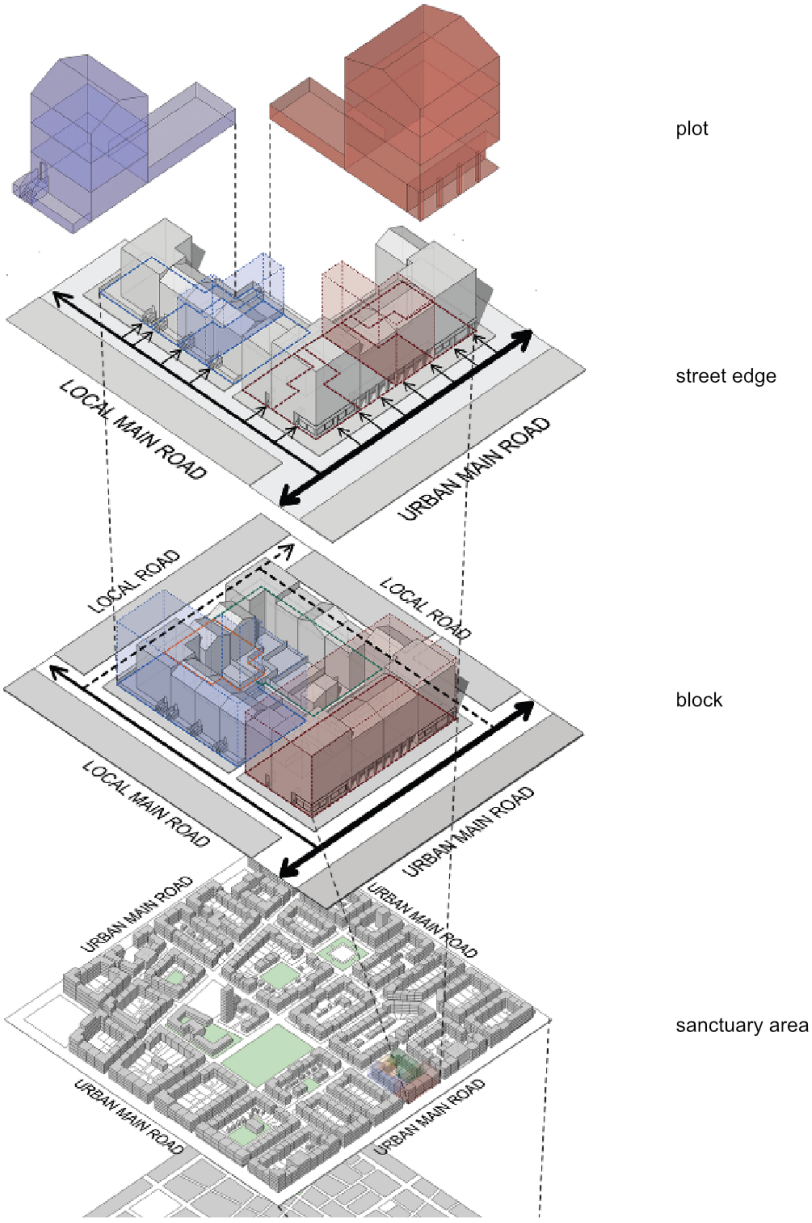


Figure 6.9: Hierarchical framework combining individual plots into street edges, blocks and sanctuary areas. Reproduced from Feliciotti (2018).

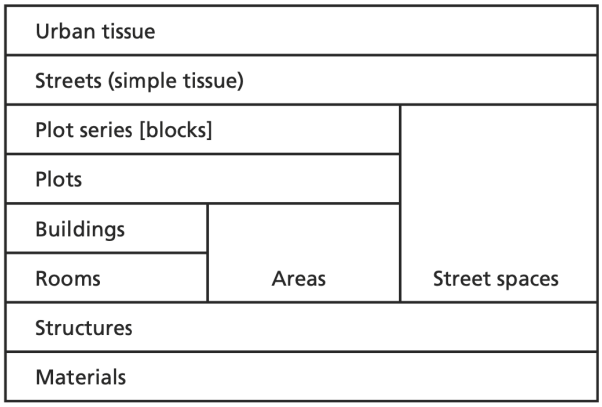


Figure 6.10: Multi-level diagram of built form proposed by Kropf (2017, figure. 6.39)

A structurally different approach is proposed by Alexander (1966). In his work *A City is not a Tree*, Alexander points out that city does not work in a tree-like hierarchy, which is reflected in both frameworks above, but as a semi-lattice of connections (figure 6.11). However, none of the analytical frameworks tends to reflect that. Furthermore, Alexander's own work on pattern language does not help in morphometric analysis.

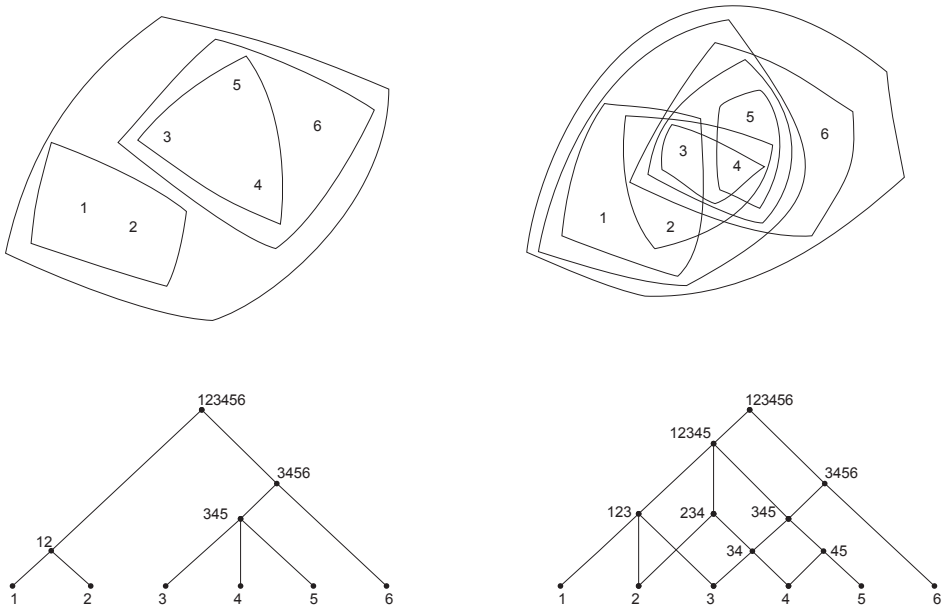


Figure 6.11: Diagrams comparing the tree-like hierarchical structure (left) and overlapping semi-lattice (right). Reproduced from Alexander (1966).

Some frameworks focus on specific elements of urban form, but elements from different approaches are usually not linked together. A typical example is network analysis based on nodes and edges and its relation to hierarchical frameworks mentioned above. The value of, say 'degree of a node', is not possible to reflect within the framework as the relation between a node and other elements is not captured.

While the literature shows that there is a broad spectrum of elements and their aggregations useful for morphometric analysis, no method links then into a singular overarching framework, which would, in turn, take a structure of semi-lattice instead of simple tree-like hierarchy. The key question here is how to build a framework which is not strictly hierarchical?

6.2 Theory of Relational framework of urban form

To put things forward, this research proposes a relational framework of urban form for urban morphometrics.

6.2.1 Relational analytical framework as an analytical tool

Relational analytical framework (RF) of urban form is based on two concepts - topology and inclusiveness. The framework acknowledges that there are identifiable relations between all elements of urban form and their aggregations. As such, it accommodates all analytical aggregations into a singular framework, linking all potential measurable characters to the smallest element. Furthermore, it employs topological relations in the way it generates location-based aggregations of fundamental elements.

Unlike frameworks above, relational analytical framework is analytical, not conceptual or structural. It does not try to propose a new theory of urban form; it has purely morphometric nature.

Within this research, relational analytical framework is operationalised based on morphological tessellation. That does not have to be the rule, but only one interpretation of the

principle. The same could be done with plots and different aggregation frameworks.

6.2.1.1 Tessellation-based relational framework

Tessellation-based relation framework starts from two hypotheses. First, that morphological tessellation can be the smallest spatial unit in the morphometric analysis. Second, that tessellation-based contiguity aggregation is better than any other location-based aggregation framework. Both hypotheses will be further tested before the application of relational framework.

Assuming both hypotheses hold, the key principles are as follows.

1. Urban form is represented as building footprints, street networks and footprint-based morphological tessellation.
2. There is an identifiable relationship between buildings and street networks, buildings and street nodes and buildings and tessellation cells.
3. Morphometric characters are measured on scales defined by topological relations between elements.
    - Element itself
    - Element and its immediate neighbours
    - Element and its neighbours within n topological steps, either in a constrained or an unconstrained way.
4. Therefore, we can define subsets of relational framework as measurable entities of urban form based on fundamental elements and topological scales.
5. Subsets are overlapping, reusing each element within all relevant relations.

Since the relation between all elements is preserved throughout the process of their combination, we can always link values measured on one subset to another. For example, due to the fixed relation between building and street node, we can attach a node's degree value to a building as an element. The constrained topological relation can identify traditional area-based aggregations like block (as a combination of all tessellation cells which topological relation does not cross a street). As such, they allow us to combine both area-based and location-based aggregations while minimising MAUP for each of them.

6.2.1.1.1 Subsets of elements Subsets are a combination of topological scales and fundamental elements. Overlap of morphometric characters derived from subsets, where each subset is representing a different structural unit, gives an overall characteristic of each duality building - cell, which can be later used for further analysis.

We can divide subsets into three topological scales: Small (or Single), Medium and Large.

Note that topological distance is possible to define within each layer (relations between buildings, relations between cells, relations between edges or nodes), but not as a combination of layers. The relation between building, its cell, its segment and its node is fixed and seen as a singular feature. That is why morphometric characters like covered area ratio of the cell are classified as a Small scale character.

6.2.1.1.1.1 Small/Single (S) Small scale captures fundamental elements themselves (topological distance is 0 - itself). In the case of building and tessellation cell, it captures the individual character of each cell. In the case of street segment and node, it captures value for segment or node, which is then applied to each cell attached to it.

We have four subsets within small scale:

- building
- tessellation cell
- street segment
- street node

Figure 6.12: Diagrams illustrating the subsets on the small/single scale.

6.2.1.1.1.2 Medium (M) The medium-scale reflects topological distance 1. It captures individual character for each element derived from the relation to its adjacent ele-

ments.

- adjacent buildings
- neighbouring cells
- neighbouring segments
- linked nodes

Figure 6.13: Diagrams illustrating the subsets on the medium scale.

6.2.1.1.1.3 Large (L) Large scale captures topological distance 2-n. In the case of cells, it captures individual character for each cell derived from the relation to cells within set topological distance. In the case of joined buildings and block, resulting measurable values are shared among all elements within such a structural unit. Block here is based on morphological tessellation and is defined as the contiguous portion of land comprised of cells which are normally bounded by streets or open space.

- joined buildings
- neighbouring cells of larger topological distance
- block
- neighbouring segments of larger topological distance
- linked nodes of larger topological distance

Figure 6.14: Diagrams illustrating the subsets on the large scale.

The resulting combination of all subsets is overlapping, following, in principle, Alexander's schema more than hierarchical frameworks.

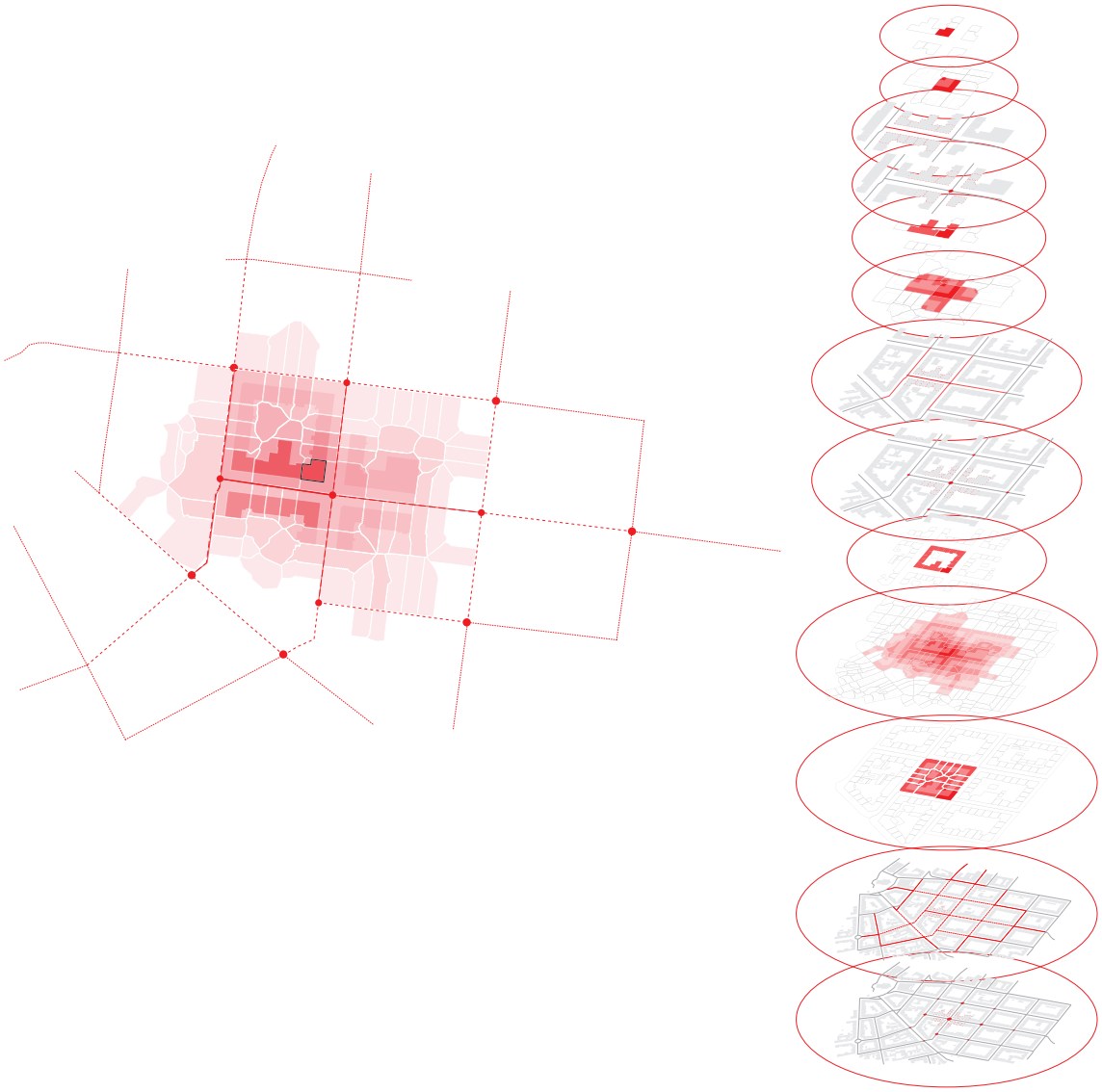


Figure 6.15: Diagrams illustrating the overlapping nature of the relational framework. The left diagram overlays all subsets on top of each other capturing the importance of each element for description of urban form around the indicated building. The darker the colour is, more times each element is used within various subsets. Diagram on the right shows all subsets aligned on top of each other describing the similar information while showing each subset directly.

6.3 Method of tessellation testing

Other sections of this chapter test the viability of morphological tessellation as the smallest unit based on its ability to reflect the similar information as the plot and topological aggregation compared to other options. The relational framework itself can be tested only in an application, which is left for chapters 7 and 8.

6.3.1 Tessellation as a unit

The first section tests whether tessellation holds as the smallest spatial unit, alongside the traditionally used plot, here represented by cadastral layer (i.e. following ownership-based definition of the plot). In the first part it presents a method of creation of morphological tessellation and in the second a method of assessment of resulting geometry in comparison to the plot.

6.3.1.1 Generation of Morphological Tessellation

Whenever observing a map or a satellite view of a city, the eye of the observer is caught by the existence of a fundamental relationship between buildings – their geometry and spatial configuration – and the plot pattern. This '*intuitive*' relationship is the reason why approaches based on Voronoi tessellation appear to *'make sense'* when applied to the urban form of cities: by partitioning the space into cells, they capture the way buildings relate to each other in space and, more precisely, give a spatial meaning to the "*morphological influence*" that each building exerts on its immediate spatial context (Usui and Asami, 2017). It, in turn, implicitly captures how spatial configuration affects visibility, light penetration, ventilation, movement, etc. around each and every building (Hamaina *et al.*, 2012).

The main advantage of methods based on Voronoi tessellation is the capacity to derive objective spatial partitions that are applicable to every type of urban tissue in a way independent from the researcher's subjective interpretation. In addition, most of these methods (Hamaina *et al.*, 2012, 2013) require minimum data input, as they fundamentally rely on the polygon that describes the footprint of a building. Similarly, the proposed

morphological tessellation method only requires a polygon layer representing building footprints (figure 6.16a). From this, morphological tessellation moves forward in five steps:

1. *Inward offset from building footprint* (figure 6.16b). The offset is necessary to avoid overlaps between boundaries of adjacent buildings and generate a gap between adjacent geometries which will later define the boundaries of the cell.
2. *Discretisation of polygons' boundaries into points* (figure 6.16c). As Voronoi tessellation can efficiently be generated only from point features, the polygonal shape of the building footprint needs to be approximated as series of points to be placed at regular intervals along its boundary, where generated points retain the ID of the building they belong to.
3. *Generation of Voronoi cells* (figure 6.16d). Voronoi cells are generated around each of the points representing the building footprint. Again, the original ID of the building is preserved in the resulting VC.
4. *Dissolution of Voronoi cells* (figure 6.16e). All Voronoi cells sharing the same building ID – and hence generated from the same building – are dissolved in unitary geometries. This step provides a preliminary boundary of the morphological cells.
5. *Clip of preliminary tessellation* (figure 6.16f). As a geometrical construct, Voronoi cells tends to infinity as the boundaries of each cell are only defined by proximity with adjacent *'seeds'*. However, when applied to the analysis of urban form, for obvious reasons, no cell can tend to infinity. To avoid this, it is necessary to limit the maximum spatial extent of the tessellation by setting defined study area boundaries.

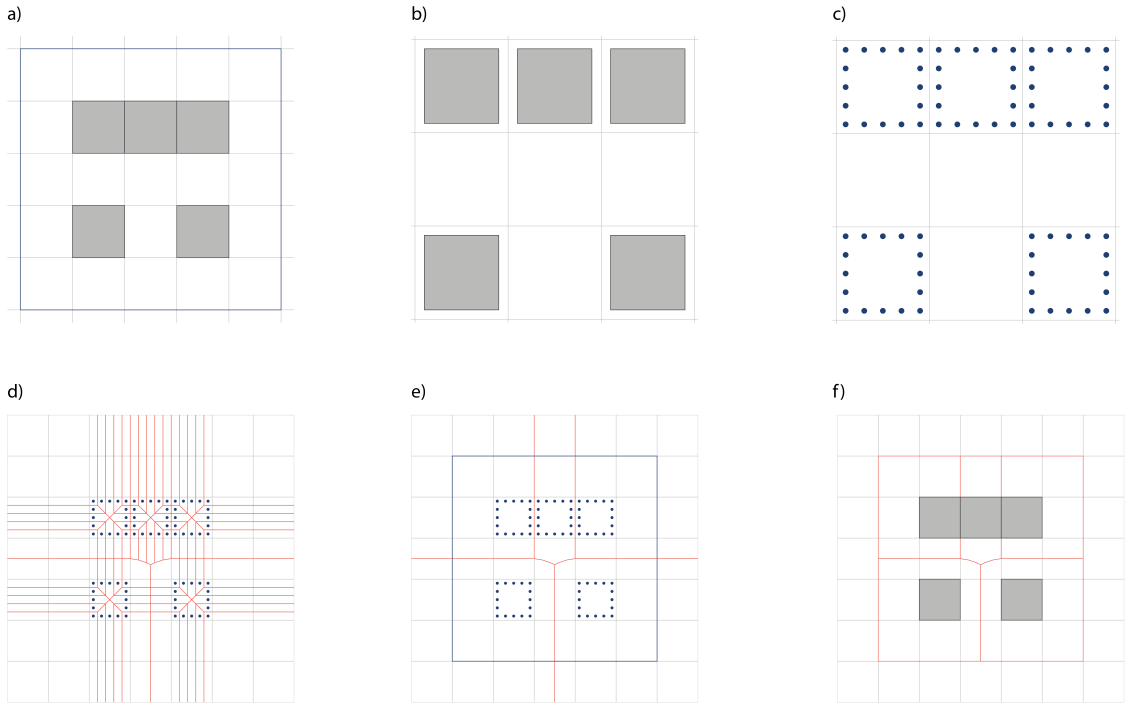


Figure 6.16: The proposed morphological tessellation method. Grey polygons represent building footprints, while red lines show the edges of tessellation at each step. 3a) Building shapes within the boundary of the study area (blue); 3b) inward offset from building footprint polygon; 3c) discretisation of boundaries of polygons into points; 3d) generation of VCs around points: at this stage, the edges of cells (red) tend to infinity; 3e) dissolution of Voronoi cells based on original building ID; 3f) clip of preliminary tessellation by study area.

Three of the five steps listed above, namely inward offset distance (step 1), discretisation interval (step 2) and clipping method (step 5), require setting parameters that can have a significant effect on the resulting tessellation. As such, these need to be evaluated in greater detail. More specifically, in the case of inward offset distance (step 1), the selection of too large values may cause the collapse of narrow parts of building shapes and loss of detail, while too small ones may generate unwanted "*saw-like*" geometries between adjacent buildings. Similarly, a large discretisation interval (step 2) may produce the same "*saw-like*" geometry issue, whilst the opposite would increase exponentially computational demand (figure 6.17). Additionally, since the two parameters are interlinked, their individual effect on the shape of each cell is not independent: as such, their combined effect needs to be balanced to generate geometries with insignificant shape deviation and minimum computational burden. Finally, the adoption of a clipping method for the tessel-

lation (step 5) also requires considerations in order to appropriately limit the focus of the analysis to the urbanised footprint and exclude large open un-built spaces while limiting potential MAUP effects (Openshaw, 1984). Due to the importance of correctly setting these parameters, section 6.3.1.1. will discuss the adopted method for the determination of inward offset distance (step 1), discretisation interval (step 2) and method.

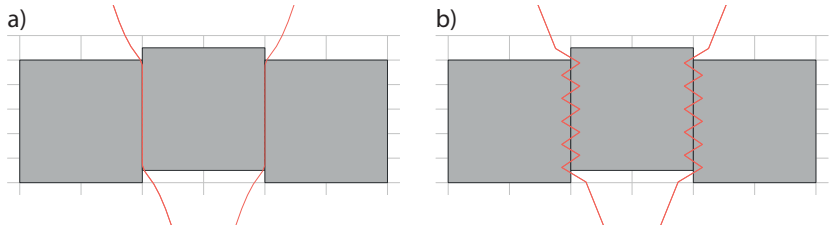


Figure 6.17: Illustration of the effect of an improper combination of inward offset distance and discretisation interval causing geometry on the boundary between adjacent buildings (b) compared to ideal combination (a).

The conceptual sequence described in this section was translated into a Python code, building its key parts on the capability of SciPy (Jones *et al.*, 2001), Shapely (Gillies and others, 2007) and GeoPandas (Jordahl *et al.*, 2020). Computation was run on Ubuntu Bionic 18.04 running at Amazon Web Services EC2. The resulting Python script is released as part of momepy.

6.3.1.2 Morphological Tessellation and plots: data and comparison method

6.3.1.2.1 The dataset Even though the rest of the thesis works primarily with Prague and Amsterdam as a case study, the following section focuses on the administrative area of Zurich, Switzerland (figure 6.18). It was chosen for its historically characterised and heterogeneous urban fabric as well as for the availability of the '*Amtliche Vermessung*' dataset, a freely-accessible resource containing high-quality information on cadastral plots and building footprints. Before generating the morphological tessellation, data was cleaned as follows:

1. From the *cadastral layer*, which covers the 100% of the study area, all features not containing buildings (e.g. streets or large open spaces) were removed, as they do not represent built-up form;

2. From the *building layer*, features smaller than 30 m^2 were filtered out, as such smaller objects are likely ancillary structures rather than actual buildings.

Figure 6.18: The selected study area, defined by the administrative boundary of the Zurich Kanton (left); the Langstrasse area in Zurich (right) was selected for testing the tessellation algorithm parameters: the red-line boundaries follow the street centerlines.

6.3.1.2.2 Definition of morphological tessellation parameters: inward offset distance, discretisation interval and clipping method To determine the optimal setting for inward offset distance (step 1) and discretisation interval (step 2), a test was run on a portion of the Langstrasse area in Zurich (Figure 5), a heterogeneous fabric predominantly characterised by adjacent buildings (significantly more prone to error than isolated buildings) limited by the street network. The test considered several combinations of inward offset (from 0.1 to 1 meter) and discretisation interval (from 0.05 to 5 meters) and evaluated them against the most precise setting (0.1 / 0.05), which provides the highest-resolution tessellation with minimal effect on the building shape. The test then assessed deviation of cell perimeter and area values for each combination, as well as its computational demand: the latter is a function of the number of discretisation points, as these directly impact on memory and processing demand. The result of this test is presented in section 6.4.1.1. Based on it, the optimal combination of the two parameters was adopted to generate the morphological tessellation in later stages.

Finally, in order to clip the tessellation (step 5), the test adopted a definition of urban footprint aligned with Angel et al. (2007; 2018), and limited the study area extent by setting a 100m buffer from the built-up area. However, to test the robustness and stability of the buffer and avoid arbitrary selection, it also tested 14 other buffers, ranging from 10 to 300 metres. The stability of the 15 buffer distances is discussed alongside the comparative analysis in section 6.4.1.2.

6.3.1.2.3 The informational value of morphological tessellation vs plots: the 12 morphometric characters To test the informational value of the morphological tessellation compared to plots, 12 morphometric characters (table 6.1) grouped into the six morphometric categories (dimension, shape, spatial distribution, intensity, connectivity and diversity) proposed in chapter 4, are selected and measured on both the cadastral layer and morphological tessellation layer, at the 15 buffer distances. With the exclusion of Reach Centrality, which is measured using the *Urban Network Analysis (UNA) Toolbox* (Sevtsuk and Mekonnen, 2012), all characters are computed using Python scripts released as part of the momepy package.

Table 6.1: Selection of morphometric characters used for comparison

Category	Character	Formula
Dimension	Area	area
	Longest Axis Length (LAL)	$max\{d_1,\ d_2,\ \ldots,\ d_n\ \}$
Shape	Circular Compactness	$\frac{\text{area}}{\text{area of enclosing circle}}$
	Shape Index	$\frac{\sqrt{\frac{\text{area}}{\pi}}}{0.5*\text{LAL}}$
	Rectangularity	$\frac{\text{area}}{\text{area}_{\text{MBR}}}$
	Fractal Dimension	$\frac{2log(\frac{\text{perimeter}}{4})}{\log(\text{area})}$
Spatial Distribution	Orientation	$\begin{cases} \text{azimuth}_{\text{MBR}}, & \text{azimuth}_{\text{MBR}} < 45^\circ \\ \text{azimuth}_{\text{MBR}} - 2(\text{azimuth}_{\text{MBR}} - 45^\circ), & \text{azimuth}_{\text{MBR}} \geq 45^\circ \end{cases}$
Intensity	Frequency	$\sum_{dist=1}^{400}$ element
	Coverage Area Ratio (CAR)	$\frac{\text{area}_{\text{building}}}{\text{area}}$
Diversity	Gini Index of Area	$G = \frac{\sum_{i=1}^{n}(2i-n-1)\text{area}_i}{n\sum_{i=1}^{n}\text{area}_i}$
	Gini Index of CAR	$G = \frac{\sum_{i=1}^{n}(2i-n-1)\text{CAR}_i}{n\sum_{i=1}^{n}\text{CAR}_i}$
Connectivity	Reach Centrality	$R^r[i] = \parallel \{j \in G - \{i\} : d[i,j] \leq r\} \parallel$

$d_1 \ldots d_n$ are diagonals of convex hull of element.

MBR is minimum bounding rectangle.

Azimuth is defined as orientation of axis between 1st and 3rd quadrant.

"The reach centrality, $R^r[i]$, of a building i in a graph G describes the number of other buildings in G that are reachable from i at a shortest path distance of at most r." (Sevtsuk and Mekonnen, 2012, p.9).

Once all morphometric characters are calculated for cadastral plots and the 15 morphological tessellation layers (at each buffer distance), the similarity of the resulting values for the two datasets is evaluated using three methods: 1) Spearman's rank correlation; 2) Normalised root squared mean deviation (NRSMD) and 3) Accuracy of significant patterns defined by local Moran's I indicator of spatial autocorrelation (LISA) (Anselin, 2010). Spearman's rank correlation is "*a measure of the correlation between ranks, calculated by using the ranks in place of the actual observations in the formula for the correlation coefficient r*" (Kokoska and Zwillinger, 2000, p. 372) (see Equation 1) and was used due to non-normality of distribution of measured values. It ranges from -1 (negative correlation) to 1 (positive correlation), with values > 0.5 or < -0.5 indicate moderately significant positive or negative correlation (Hinkle *et al.*, 2003).

$$(1)\quad s = 1 - \frac{6\sum d^2}{n\left(n^2-1\right)},$$

where $d_i = \mathrm{rg}\left(X_i\right) - \mathrm{rg}\left(Y_i\right)$ is the difference between the rank of observed and expected value and n is the number of observations (Kokoska and Zwillinger, 2000).

NRSMD is a frequently used measure of "*an estimate of the standard deviation of residuals from the model*" (Alexander *et al.*, 2015, p. 5) normalised by the range (see Equation 2), and it is used to measure the difference between the expected and observed values, normalised by the range. As a ratio of deviation, it ranges from 0 to 1, where 0 means no deviation and 1 means deviation equal to the range of values. As the range is sensitive to outliers, NRSMD might not be relevant for characters of *Dimension* category.

$$(2)\quad \mathrm{NRMSD}(y, \hat{y}) = \frac{\sqrt{\mathrm{MSE}(y,\hat{y})}}{y_{\max} - y_{\min}},$$

where $\text{MSE}(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2$ where $y, \hat{y}$ are observed and expected values.

Accuracy is "*closeness of computations or estimates to the exact or true values that the statistics were intended to measure*" (OECD, 2006). It is here used to measure the similarity of significant spatial clusters identified from the cadastral layer and those identified from each version of the tessellation (see Equation 3). Since studies in Urban Morphometrics are more interested in uncovering *recurrent patterns* in urban form rather than *actual values* (Feliciotti, 2018), this method is probably the most relevant of the three. In fact, it measures whether corresponding features from both datasets (cadastral plots and morphological cells in this case) significantly fall within the same cluster (i.e. p <= 0.05), with values ranging from 0 (no match) to 1 (perfect match).

(3) $\text{aLISA} = \frac{SC_{\text{match}}}{SC_{\text{max}}}$,

where SC_{match} is the number of the elements belonging to the same significant spatial cluster in both $y, \hat{y}$ and SC_{max} is the number of the elements $\hat{y}$ belonging to any significant cluster. The adjacency matrix used for LISA represents 200 metres Euclidean distance from each building.

It must be noted that, for the statistical comparison of selected morphometric characters across the morphological tessellation layers and the cadastral layer, these must correspond perfectly. However, whilst there is a 1:1 match between morphological cells and *buildings*, the same does not apply to morphological cells and plots, as the latter may contain one building (single-building plots) or more than one (multi-building plots). To resolve this issue, the building layer is used as a proxy between tessellation and cadastre and, therefore, all morphometric characters computed on both morphological cells and plots are associated to the building layer (i.e. each building is linked to the value of its morphological cell and of the plot it sits on). However, to better understand the impact of '*single-building*' and '*multi-building*' plots (79% and 21% of all plots respectively), the three methods described above are applied to the whole dataset and, separately, for *single-building* and *multi-building* plots. In particular, we expected that *multi-building* plots, although important for their effect on the overall analysis, would hold limited comparative value for most of the assessed morphometric characters (perhaps with the only exclusion of covered area ratio and Gini index of CAR, which capture compatible concepts).

6.3.2 Topological contiguity of tessellation as an aggregation framework

The second test of tessellation within this chapter focuses on its ability to derive topological location-based aggregations. The definition of aggregated analytical units via the topology of morphological tessellation can overcome issues of the three methods described above and provide a more consistent way to understand the relationship between adjacent elements of urban form (in the case of buildings or morphological cells). This section is testing this hypothesis in the case of Prague, Czechia.

6.3.2.1 Comparing aggregation methods

The methodology of this research follows a twofold approach, analysing both small scale case studies and urban scale statistical data. Small scale case studies examine the difference between three methods extracted from literature (Euclidean distance, metric reach, K-nearest neighbour) and unconstrained topology of morphological tessellation in different types of urban form; large scale statistical analysis examines the parameters of these methods of aggregation across the whole of Prague.

The method compares how each of the tested methods aggregate tessellation cells (being smallest spatial unit) within two scales: one achieved by nine topological steps, the equivalent of approximately 400 metres used in the morphological analysis to represent a walking distance of 5 minutes; and one achieved by 4 topological steps, representing roughly 200 metres. The number of neighbours for KNN is then derived from the mean number of neighbours captured by each of the topological distance and metric reach, to keep the dimension comparable (table 6.2).

Table 6.2: Default topological distances and their equivalents. Values are derived from the summative analysis of topology-based aggregations defined around each morphological cell on Prague.

Topology of MT	Euclidean	Metric reach	KNN
4 steps	200 metres	200 metres	70 neighbours
9 steps	400 metres	400 metres	320 neighbours

The aim of the small-scale analysis is to understand how each of the four methods identifies and represent the same information across 5 types of urban form - medieval organically grown, 19th century compact perimeter blocks, 20th century mixed single and multi-family villas, 20th century modernist housing, and 20–21th century industrial estates. It should allow ascertaining how consistently each method distinguishes variations in form and morphological behaviours.

At the urban scale is conducted a statistical analysis to compare the distribution of values derived from the whole of Prague. The statistical distribution of data across the whole urban area describes the spread and variance of values, which can be used to assess the ability of each method to capture the intended information across types of urban tissue. To understand the different performance of each method, the method compares distributions of two descriptive variables as a proxy for the performance assessment – number of neighbours and covered area.

The first variable is the number of neighbours captured. Neighbours represented by buildings and related tessellation cells capture most of the morphological information. For that reason, it is desirable to use the method which will identify the somewhat similar number of neighbours no matter the urban tissue to keep the similar essence and amount of information to maximise comparability of values. It means that the distribution of such values should have a relatively small standard deviation and be close to symmetrical distribution to have a similarly positive and negative deviation from the mean.

The second variable should represent the concept of the geographical extent of aggregation, as it bears the information of the scale of each type of urban tissue and therefore could describe the ability of each method to adapt to the scale. Amongst the possible measurable variables are mean distance to neighbours, maximum distance to neighbours and area covered by aggregation. Because they all represent the scale and extent of aggregation of elements (buildings, tessellation cells), we use the only area covered to represent them all as it is the most straightforward one and easy to understand. The statistical distribution of the covered area should represent the adaptability of the aggregation method. Hence, the ideal outcome should have a high standard deviation and high range of values, meaning that many different options (levels of granularity of urban form) are all captured.

6.4 Results

The following section presents results of the analysis, first assessing the tessellation as a unit in the case of Zurich and then analysing its aggregation ability in the case of Prague.

6.4.1 Tessellation as a unit

Test of tessellation as a unit focused on two aspects - determination of optimal parameters of the algorithm generating tessellation from building footprints and comparison of morphometric values between the cadastral layer and tessellation.

6.4.1.1 Determination of optimal parameters of the MT algorithm

The test performed on the selected inward offset ranges (from 0.1 to 1 meter) and discretisation intervals (from 0.05 to 5 meters) allowed to assess computational demand (i.e. a number of discretisation points) and deviation of cell perimeter and area for each combination. In terms of computational demand, as shown in figure 6.19, it appears that the discretisation segment length has an exponential effect on the number of generated points. For values below the mean (tail of the distribution), computational demand remains relatively stable, whilst for higher values (head of the distribution) it grows sharply, more than doubling at each step. Discretisation intervals 0.5m are therefore preferred as more computationally effective.

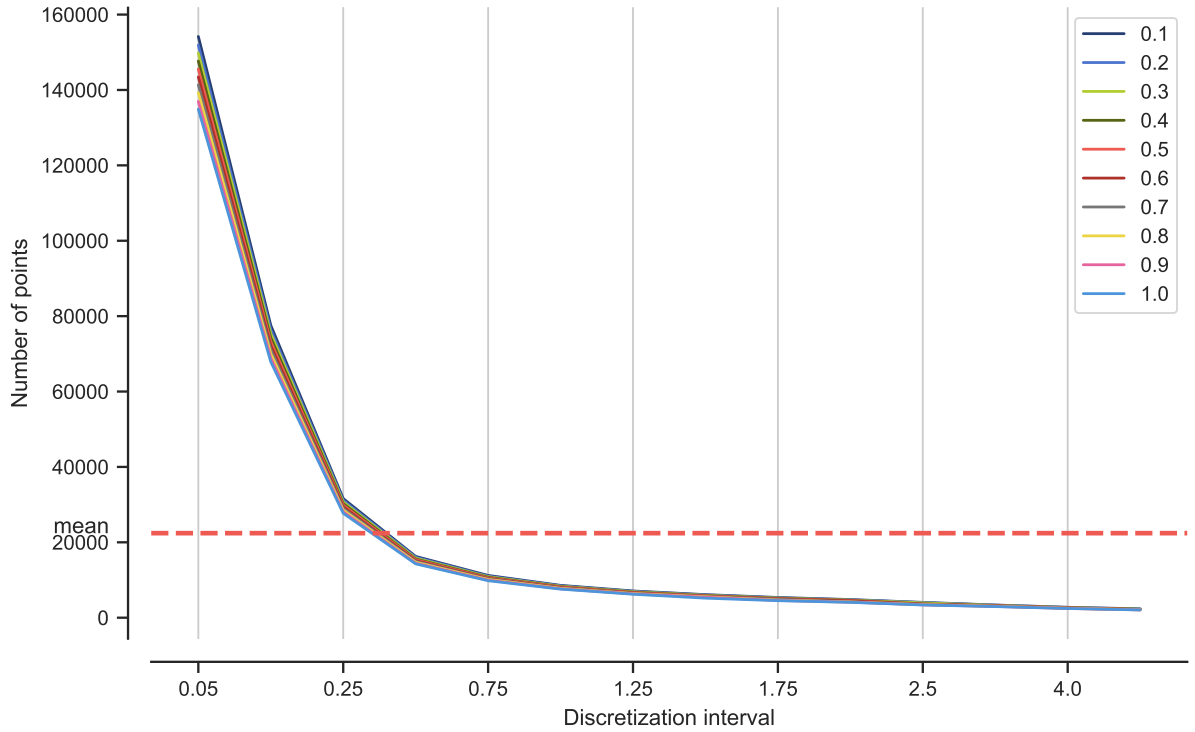


Figure 6.19: Relation of discretisation segment length and number of points generated. The red line illustrates the mean value above which the number of points more than doubles at each step

The effect of negative buffer and discretisation interval on the deviation of the morphological cell's area compared to the high-resolution tessellation is insignificant for all tested combinations (0.00 and 0.01%), showing that, no matter the parameters, results are stable. In turn, the same effect on the morphological cell's perimeter is more pronounced (figure 6.20) due to the aforementioned phenomenon of "*saw-like*" geometries (see the section 6.3.1.1) with per cent deviation ranging from 0.05% to 7.4%. Focusing on the 0.5 metres discretisation interval, providing the balance between the morphological cell shape detail and computational demands, deviation values range from 0.47% to 3.1% (figure 6.21). It suggests that the combination of 0.5m metres discretisation interval and 0.4m inward offset distance provides the optimal balance in terms of the effectiveness of computation and minimisation of error. These values are hence adopted as parameters in the computation of the morphological tessellation in the next stages.

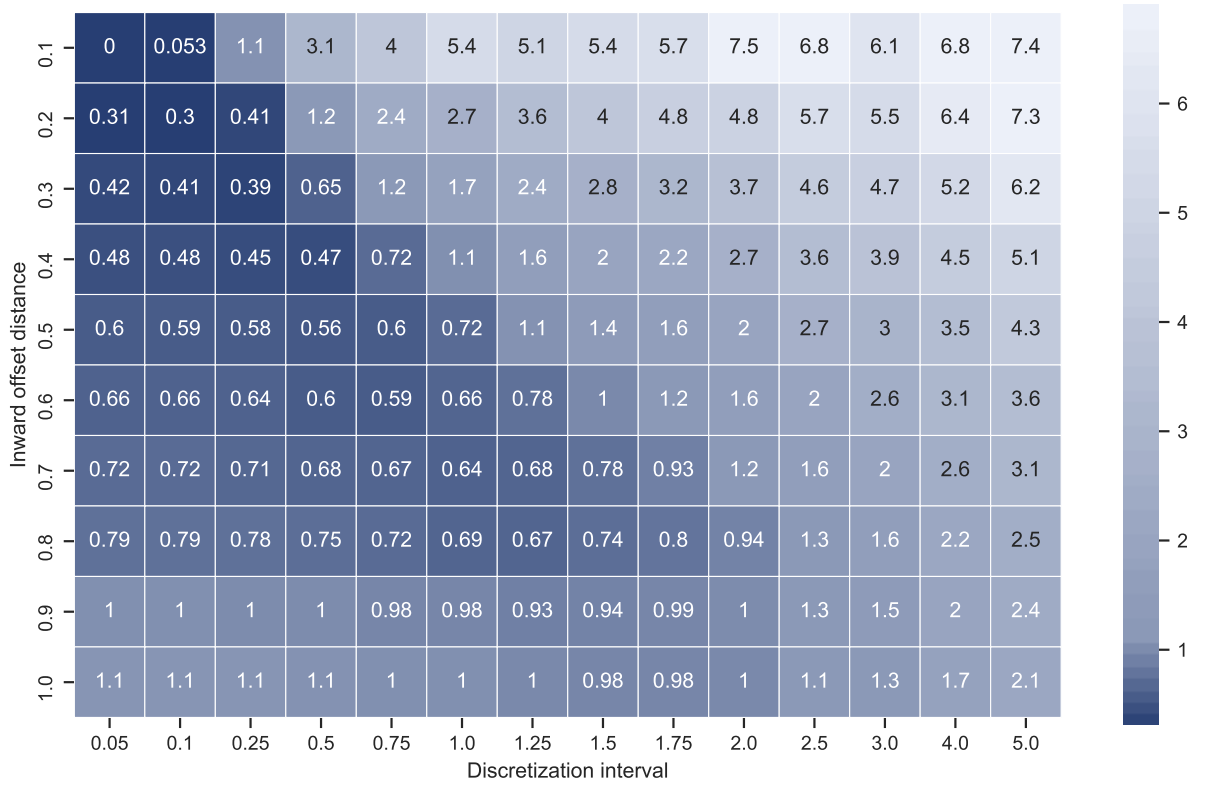


Figure 6.20: The mean deviation in percents of perimeter of each cell for each combination of inward offset distance (vertical axis) and discretisation interval (horizontal axis).

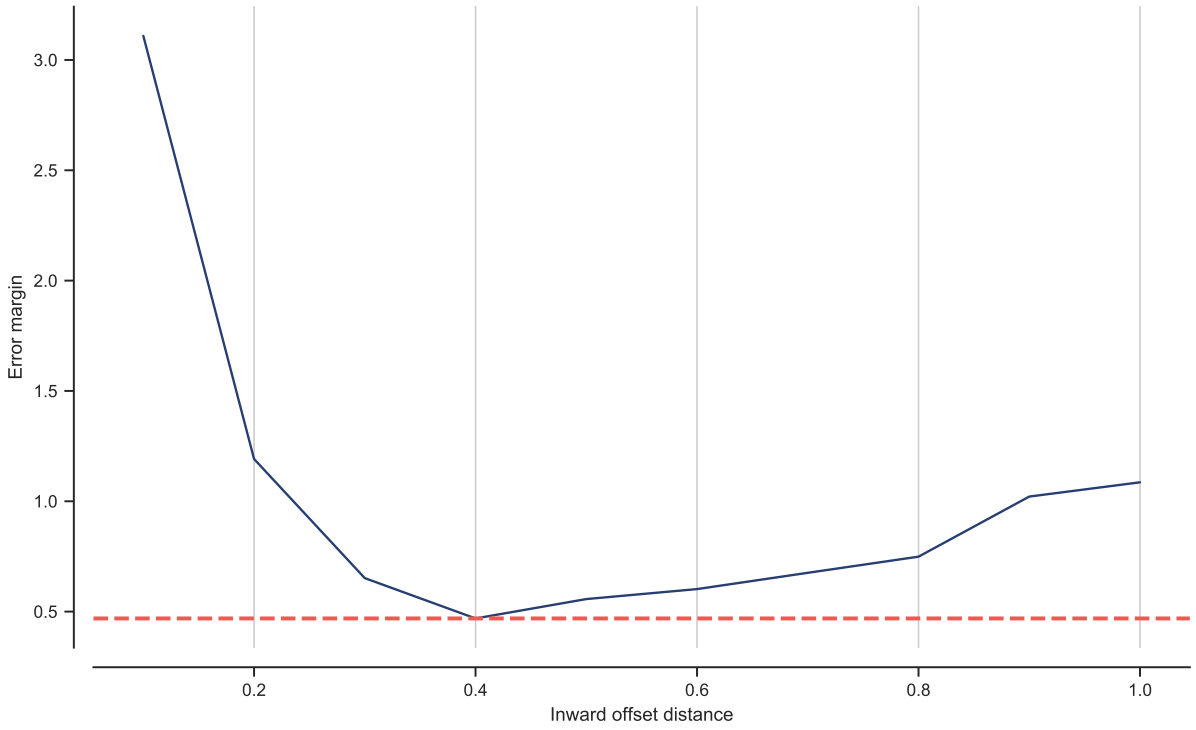


Figure 6.21: Relation of inward offset distance and error margin, showing that for 0.4 meter, the error margin reaches its minimum.

6.4.1.2 Comparison between the cadastral layer and morphological tessellation

Having determined the optimal combination of the tessellation parameters (inward offset = 0.4m and discretisation intervals = 0.5m), the morphological tessellation for Zurich is computed using momepy. From a first visual inspection of the generated layer, it is already possible to appreciate how the morphological tessellation is able to capture variations in size nicely, grain and compactness of buildings (figure 6.22), not dissimilar what observed in a typical cadastral layer. The method subsequently calculates the 12 morphological characters in table 6.1 for cadastral plots and tessellation cells. In the next section, their correlation at each buffer of tessellation is studied.

Figure 6.22: Morphological tessellation cells as generated across four different areas of Zurich; 4a) organic tissue of Niederdorf; 4b) compact tissue of Langstrasse; 4c) detached villas of Hottingen; 4d) mixed post-war development of Friesenberg.

6.4.1.2.1 Spearman's rank correlation Using Spearman's rank correlation, results show that correlation of measured characters ranges between 0.25 (fractal dimension) to 0.89 (reach), with differences between morphometric categories and between *single-* or *multi-building* plots. Characters in the *Shape* category exhibit the worst performance, with insignificant correlation for the whole sample (~0.27) and multi-building plots (~0.09) and low significance for single-building plots (~0.42). This result was expected, due to the intrinsically different geometry of the two spatial units (morphological cells and plots)

and to the existence of multi-building plots.

Dimension characters inherently differ between *multi-* and *single*-building plots, showing only low significance for the former (~0.35, ~0.4) and high significance for the latter (~0.83, ~0.7). Remaining characters show moderate or high significance for all samples, with higher values for single-building plots (figure 6.23 and Table 2). Results for all buffers are relatively consistent, with fluctuations observed only at smaller distances (< 50m), indicating the stability of the selected value of 100m.

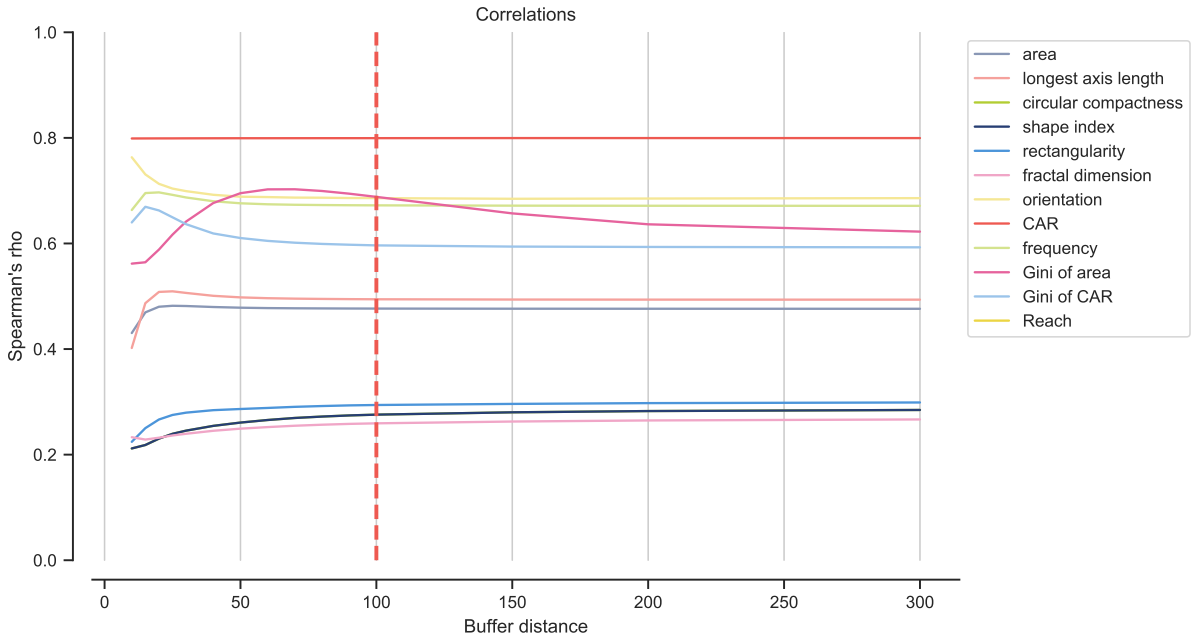


Figure 6.23: Spearman's rho rank correlation between cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi-building plots are found in Appendix 6.1).

Table 6.3: Spearman's rank correlation of the whole dataset, single-building plots and multi-building plots at 100m buffer (emphasis reflects significance of correlation).

Category	Character	All	Single	Multi
Dimension	Area	*0.4767*	***0.8273***	*0.3583*
	Longest Axis Length	*0.4943*	***0.7055***	*0.4073*
Shape	Circular Compactness	0.2758	*0.4203*	0.0864
	Shape Index	0.2758	*0.4203*	0.0864
	Rectangularity	0.2940	*0.4040*	0.1214

Table 6.3: Spearman's rank correlation of the whole dataset, single-building plots and multi-building plots at 100m buffer (emphasis reflects significance of correlation).

Category	Character	All	Single	Multi
	Fractal Dimension	0.2593	*0.4407*	0.0360
Spatial Distribution	Orientation	**0.6859**	***0.7985***	**0.5713**
Intensity	Frequency	***0.7995***	***0.9103***	***0.7093***
	Coverage Area Ratio	**0.6721**	***0.7649***	**0.5567**
Diversity	Gini Index of Area	**0.6882**	***0.7291***	**0.6312**
	Gini Index of CAR	**0.5963**	**0.6263**	**0.5551**
Connectivity	Reach	***0.8851***	***0.9371***	***0.8282***

6.4.1.2.2 Normalised RMSD Overall, the RMSD test indicates a high level of similarity between datasets (figure 6.24 and Table 3), excluding *Dimension* characters which, as mentioned in Section 3.2 are heavily skewed by large outliers, hence not comparable with the rest of the data. Apart from *Orientation*, which is the worst-performing character in the set (~0.22 for the whole dataset, ~0.26 for multi-building plots and ~0.18 for single-building plots), all other characters score RMSD values lower than 0.2 (~0.15 for single-building plots and ~0.18 for multi-building plots). It suggests that, even though the spatial coverage of the morphological tessellation is different from plots, this difference is, in terms of information, only minor. Even the poorer performance of *Orientation* depends more on the way this is measured than on dissimilarity between datasets: unlike other metrics, *Orientation* is calculated as a deviation of the orientation of the longest axis of minimum bounding rectangle (sometimes called oriented envelope) from cardinal directions in degrees and, as such, it ranges from 0 to 45º. Hence, a deviation of 0.2 corresponds to a difference only of 9º. It is worth noting that for smaller buffers (15 to 40m) results show high instability, where some characters exhibit the highest correlation values and others the lowest: this confirms that smaller buffers are unsuitable as parameters to limit the tessellation. In turn, the 100-metre buffer is confirmed as robust and stable across all characters.

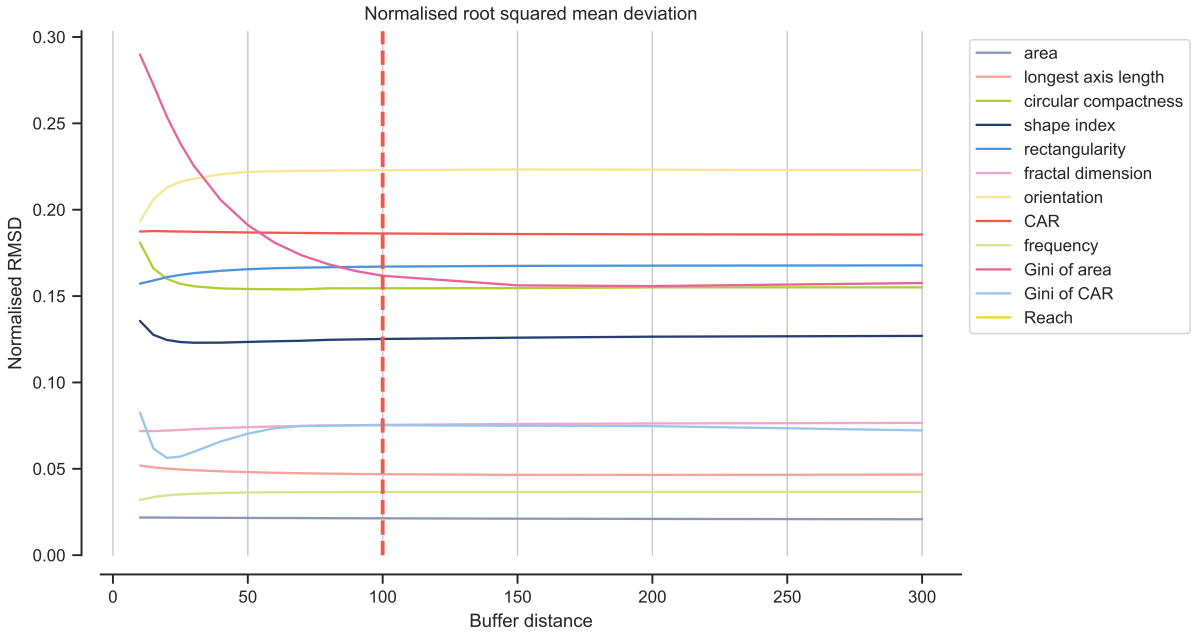


Figure 6.24: NRMSD of cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi-building plots are found in Appendix 6.2).

Table 6.4: NRMSD for the whole dataset, single-building plots and multi-building plots at 100m buffer.

Category	Character	All	Single	Multi
Dimension	Area	0.0213	0.0075	0.0326
	Longest Axis Length	0.0469	0.0162	0.0645
Shape	Circular Compactness	0.1545	0.1270	0.1788
	Shape Index	0.1252	0.1000	0.1479
	Rectangularity	0.1671	0.1563	0.1773
	Fractal Dimension	0.0754	0.0566	0.0970
Spatial Distribution	Orientation	0.2229	0.1775	0.2601
Intensity	Frequency	0.1862	0.1507	0.2163
	Coverage Area Ratio	0.0366	0.0432	0.1224
Diversity	Gini Index of Area	0.1618	0.1509	0.1724
	Gini Index of CAR	0.0752	0.0691	0.0838
Connectivity	Reach	0.1685	0.1528	0.1828

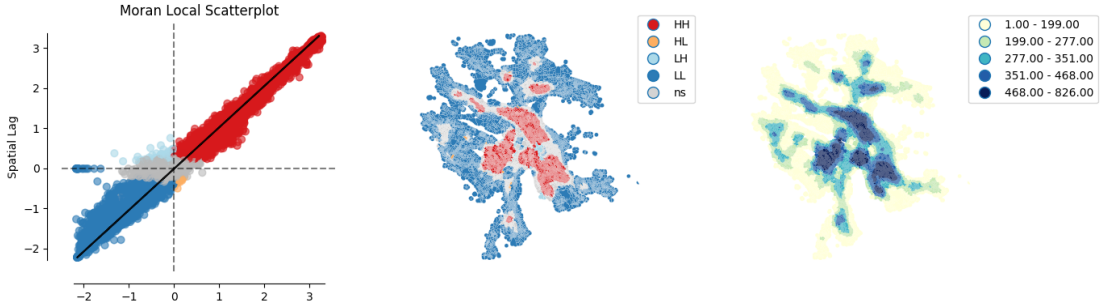


Figure 6.25: Example of LISA patterns of Frequency measured on the MT shows Moran Local Scatterplot LISA clusters and actual distribution of values.

6.4.1.2.3 Recognition of significant patterns using LISA The analysis of patterns with LISA (figure 6.25) captures differences across measured characters; however, given the dissimilarity of the datasets due to multi-building plots, the accuracy scores are not expected to reach values close to 1. The highest pattern similarity is recognised for the *Frequency* character (~0.78, corresponding to an almost 80% match) (figure 6.26), followed by *Orientation* and *Diversity* characters (Gini Index of Area and Gini Index of CAR) and CAR (0.74 – 0.66), while *Dimension* characters are around ~0.5 depending on the sample considered (single- or multi-building plots). *Shape* characters are consistently the ones providing lowest accuracy, apart from *Reach Centrality*, due to the single-building vs. multi-building deviation in the datasets. Overall, the difference between samples is relatively consistent, with single-building plots reaching values between 0.1 and 0.2 higher than multi-building plots. While none of the values indicates an equality of both datasets, some are close enough to be considered as proxies of each other. The effect of buffer distance confirms already observed pattern and the stability of the 100m buffer.

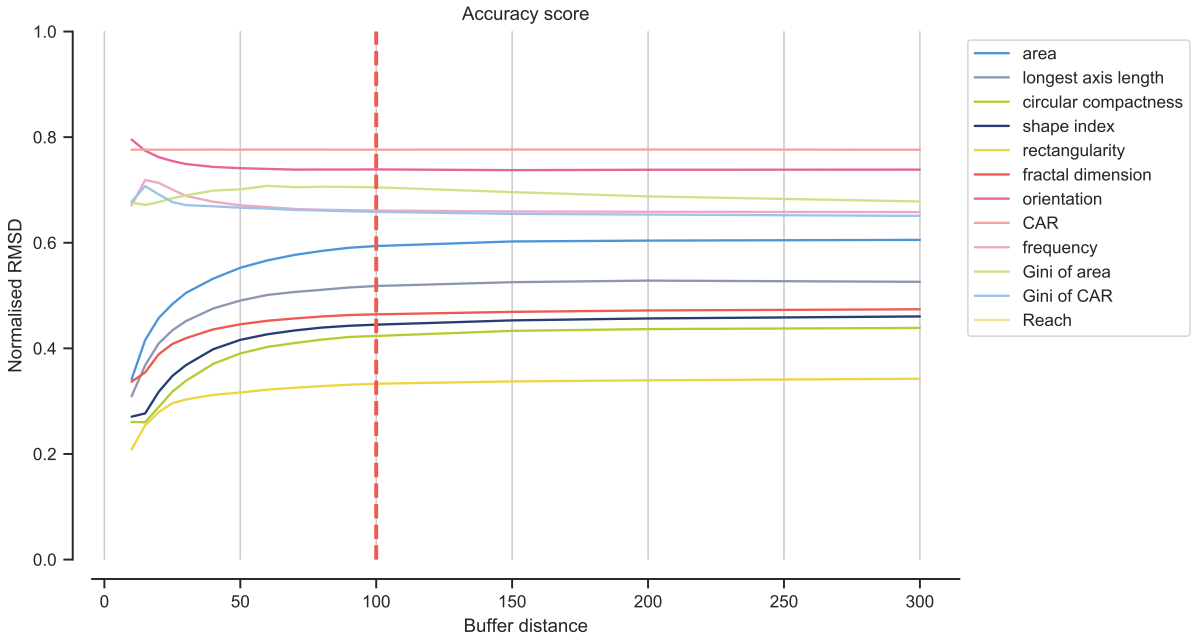


Figure 6.26: LISA accuracy of cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi-building plots are found in Appendix 6.3).

Table 6.5: aLISA for the whole dataset, single-building plots and multi-building plots at 100m buffer.

Category	Character	All	Single	Multi
Dimension	Area	0.5938	0.6447	0.5090
	Longest Axis Length	0.5181	0.6138	0.4028
Shape	Circular Compactness	0.4235	0.5061	0.3319
	Shape Index	0.4449	0.5312	0.3475
	Rectangularity	0.3330	0.3930	0.2761
	Fractal Dimension	0.4644	0.5652	0.3489
Spatial Distribution	Orientation	0.7389	0.8055	0.6711
Intensity	Frequency	0.7763	0.8240	0.7318
	Coverage Area Ratio	0.6610	0.7313	0.5908
Diversity	Gini Index of Area	0.7050	0.7333	0.6759
	Gini Index of CAR	0.6585	0.6742	0.6423
Connectivity	Reach	0.4007	0.3363	0.4644

6.4.2 Tessellation contiguity as an aggregation method

At the small scale, the method studied how five urban types in Prague are represented by the four methods of aggregation. The Old Town, a tissue of a medieval origin which has grown organically, shows few differences between the four methods, with slightly larger footprints of aggregations defined by Euclidean distance and a morphological tessellation topology at both 200 (figure 6.27a) and 400 (figure 6.28a) metres distances. Numerically, the difference is clear, but for pattern-detection this difference is not substantial, suggesting that all methods are relatively equal in this tissue for both 200 and 400 metres (and equivalents). Differences might be explained by the high granularity of the tissue, with many elements on a relatively small area (the reason for KNN being the smallest) and complex configuration amongst them (buildings have many neighbours, expressed by more extensive topological-based aggregation).

Figure 6.27: Comparison of boundaries of aggregations defined by each of the tested method for 4 topological steps and equivalents (200 metres, 70 neighbours). a) Old Town, b) Vinohrady, c) Hanspaulka, d) Jižní Město, e) Malešice

In the second case, the urban tissue of 19th century of compact perimeter blocks in the Vinohrady neighbourhood results almost match the previous case. Due to the high granularity of this urban tissue, purely Euclidean distance-based area is the largest, while K-nearest neighbour the smallest (figure 6.27b, figure 6.28b).

Overall, in these two historic tissues, the difference is not significant to conclude that one method is better than the other - simple visual comparison shows that boundaries almost overlaps for both tested distances.

The first crucial differences are noticeable in Hanspaulka, an area of 20th century mixed

single and multi-family villas, where the street network is becoming less dense and less connected than in the city centre, leading to the difference between the area captured by metric reach (smaller) and the other three methods, which almost overlap without any substantial distinction (figure 6.27c, figure 6.28c). This indicates that the street network plays a crucial role in the applicability of reach-based methods.

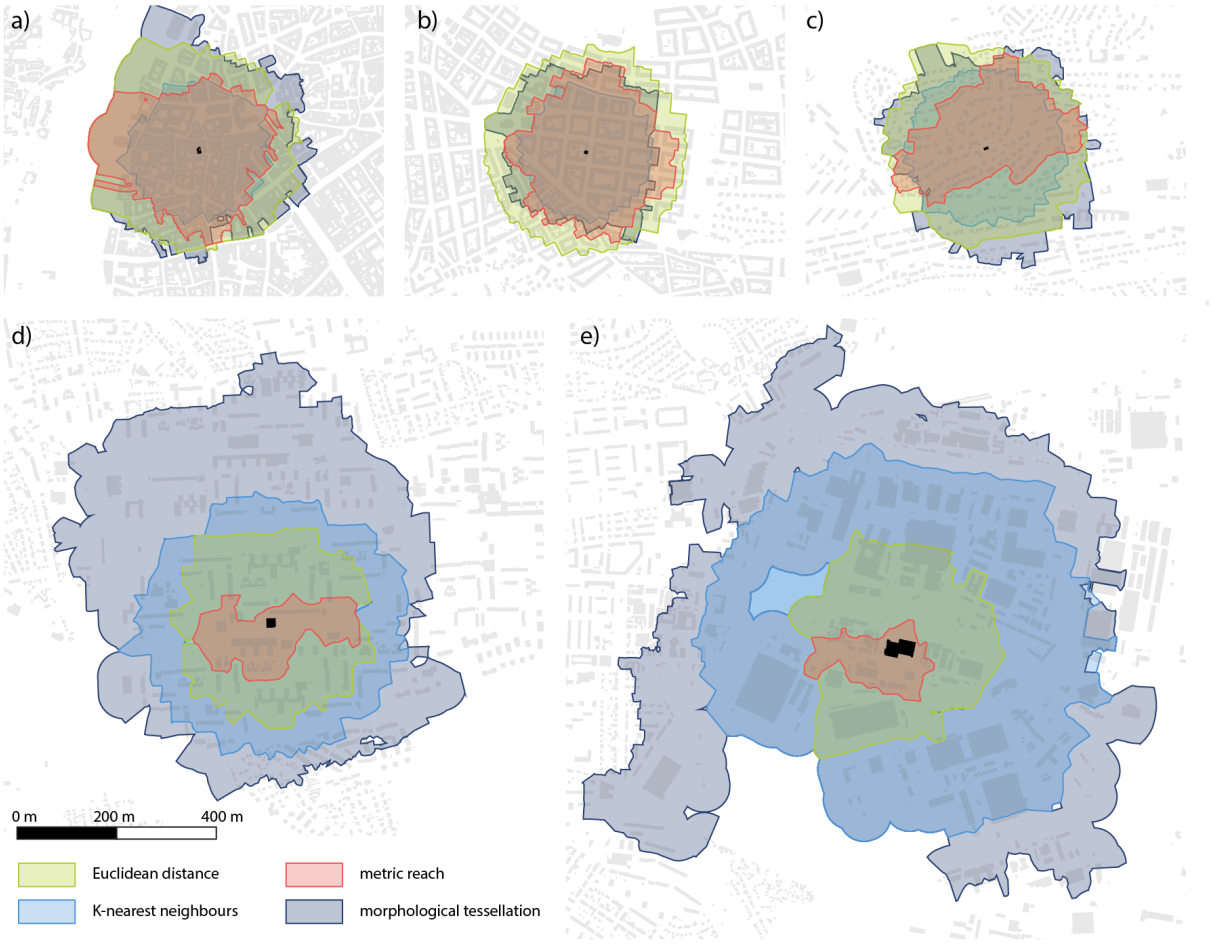


Figure 6.28: Comparison of boundaries of aggregations defined by each of the tested method for 9 topological steps and equivalents (400 metres, 320 neighbours). a) Old Town, b) Vinohrady, c) Hanspaulka, d) Jižní Město, e) Malešice.

The twentieth-century modernist housing of Jižní Město is the first example of a post-WW2 urban tissue. The planning ideology behind it comes with the radical change of scale and a distinctive approach to streets and their connectivity. This change is reflected in how each of our tested methods captures the space: while the area defined by Euclidean

distance remains mostly the same as in pre-WW2 tissues, the area captured by metric reach shrinks due to the convoluted street network. In theory, both topological and KNN definitions of aggregation should be able to capture the difference in scale and up to a certain level they do. However, KNN, even though being larger than metric-based methods, lacks the ability to deal with large pavilion-like buildings with many direct neighbours, unlike the topological definition which correctly reacts to the abrupt change of scale of the granularity of urban tissue and captures the relationship between high-rise buildings and their low-rise pavilion counterparts by acknowledging that they are neighbouring (figure 6.27d, figure 6.28d).

Whilst industrial type tissues are generally not the concern of urban morphology, as classified as specialist and treated differently than more ordinary fabric, they are nonetheless large, therefore important parts of our cities and as such deserve to be studied using the same approach as the more conventional ones. Their scale is radically different. Buildings are of the size of the traditional block or larger, the plot structure is mostly unorganised, and the street network is utilitarian only, following different principles than in residential or mixed-use parts of the city. These differences are captured through the application of our four methods. The network-based method is unreliable on this tissue, capturing the only a minor area around the building due to the major drop in a granularity and connectivity. The Euclidean distance of 200 or 400 metres, which seems to capture enough information in more granular urban tissues lacks the same capacity in this case. K-nearest neighbour analysis struggles to capture the peculiarities of this particular urban tissue, which is characterised by a large amount of additional built-up structures to main buildings, leading to the identification of smaller area that makes a comparison with the other cases confused. The topological definition achieved by the morphological tessellation seems to tackle all issues of the other methods, whilst capturing a similar amount of information as it did in previous cases (figure 6.27e, figure 6.28e).

Overall, the differences between methods in defining aggregation are heavily dependent on the type of urban tissue analysed. More traditional (from a European perspective) urban tissues like medieval (Old Town) or perimeter blocks (Vinohrady) indicate that in these contexts the choice of the method is purely the matter of opinion and that the resulting value offered by the four methods is mostly similar. However, once we start focusing on post-WW2 development, we often observe a change of scale of urban patterns, which makes distance-based methods (Euclidean, metric reach) unable to react to such change.

The information captured is consequently different in pre-and post-WW2 urban tissues, complicating the further comparability, whilst we seek similar and consistent data. If the urban patterns change their scale, the method of capturing such an extent needs to be able to adapt to it. Our results of the small-scale analysis indicate that topology of morphological tessellation is the method able to fulfil this condition adequately. Whilst small scale analysis illustrated the capacity of the four selected methods to provide stable information, it is only at an urban scale, through statistical analysis, that we can show a full overview of how the four methods perform.

As mentioned, neighbouring elements are bearing the primary information about urban patterns. For this reason, researchers aim to use methods capturing an equal number of neighbours across contexts. Such a method might be K-nearest neighbour, but due to the variety of urban configurations, a method needs a certain level of adaptability (which KNN with a fixed number cannot provide). As figure 6.29(a, b) shows, the statistical distribution of the number of neighbours captured is the most stable for the topology of morphological tessellation, being almost perfect Gaussian distribution (the deviation in the number of neighbours is the same in both directions from the mean), with the smallest standard deviation (σ). The metric reach method to provides right-skewed distribution and Euclidean distance high deviation, which are both undesirable features in terms of stability of information.

Then, the comparison of distributions of covered area aims to test the adaptability of each method. As mentioned, the changing scale of urban patterns means that the same level of information is spread to larger areas. Therefore, an ideal method should show high flexibility (the distribution of values should have large range and high standard deviation) in the area captured to fit all patterns possible. The results as shown in figure 6.29(c, d) indicate that topology of morphological tessellation offers by a large margin the highest standard deviation out of tested methods, indicating that the change of the scale is captured successfully. Metric methods (Euclidean distance, metric reach) are the least flexible in this sense, while K-nearest neighbour might offer desired value alongside with morphological tessellation topology.

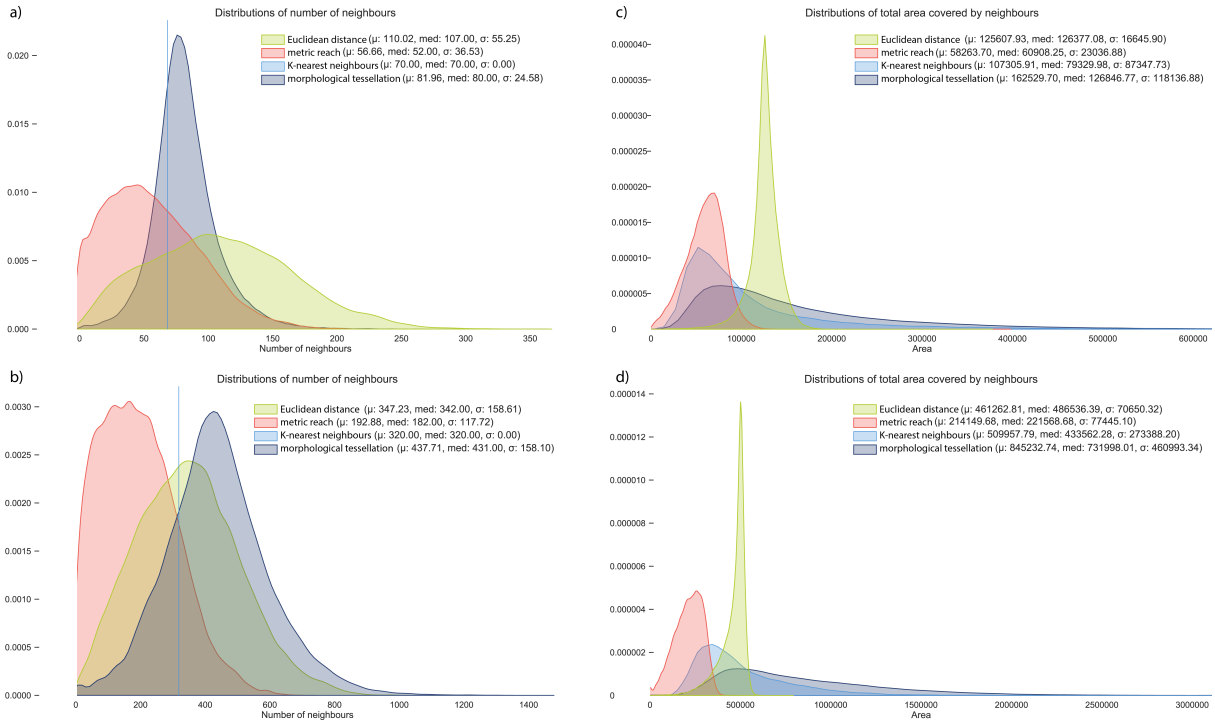


Figure 6.29: Statistical distributions of number of neighbours a) 4 steps and equivalents, b) 9 steps; and total covered area c) 4 steps and equivalents, d) 9 steps.

Even though there are differences between smaller and larger distances (4 steps / 9 steps and its equivalents), topologically defined aggregation seems to reflect desired outcomes (i.e. stability in a number of elements captured and flexibility in metric values) in both statistical comparisons better than other tested methods. This finding is in line with the one we drew from small-scale case studies, indicating that topology of morphological tessellation is a valuable approach to be employed in morphological analysis.

6.5 What is the value of tessellation?

Results of both experiments show that there is a high potential in using morphological tessellation as a spatial unit, both as a unit of analysis and as a core of the aggregation technique.

6.5.1 Tessellation in relation to plot

The results suggest that the proposed method contributes to resolving some of the limitations associated with using the plot as a unit of morphometric analysis. However, the picture resulting from testing the similarity between cadastral plots and morphological cells is rather complex. Notably, the significance of the similarity between plots and morphological cells varies considerably depending on the morphometric character selected: this is generally high for all Intensity characters (*Frequency, CAR*, whilst Shape characters (*Rectangularity, Circular Compactness, Shape Index*, and *Fractal Dimension*) report a comparatively lower performance and a higher deviation. It means that if, for several of the morphometric characters assessed, morphological tessellation is able to retain plot-level information which is comparable to that provided by the cadastral layer, for other characters morphological cells are less efficient proxies of plots and capture comparatively different information.

It is also evident that the similarity of datasets is higher across all measured characters for single-building plots compared to multi-building plots. To consider that the former ones are predominantly found in pre-industrial urban tissues. At the same time, the latter is more typical of modern and contemporary development, and it is suggested that the morphological cell might capture similar information as the plot in the context of traditional fabrics better than in modernist and contemporary ones.

Overall, there appears to be a scope for the morphological cell to be utilised as the basic unit of morphometric analysis, given its ability to capture meaningful patterns of urban form at the plot scale, the degree of reliability and universality of the underlying method and the wide accessibility of the data required to generate it. Indeed, while the recognition of plots can be very troublesome and resource-intensive, morphological tessellation is consistent throughout, since it is only based on building footprint information which is equally present in all kinds of urban areas. Moreover, by using morphological tessellation instead of traditional methods relying on buildings, street networks and plots, data dependency is reduced by a third as the tessellation is generated from the building layer alone.

On the other hand, the morphological tessellation cannot fully substitute the plot. The similarity of information indicates only partial overlap, and there are aspects of form

which cannot be derived from morphological tessellation conceptually, like the relationship between public and private or ownership-based analysis. That said, the morphological tessellation has the potential to work as an analytical spatial unit within the morphometric assessment, acknowledging that the information obtained from it is not equal to that derived from plots (no matter the definition).

6.5.2 Location-based aggregation using tessellation

Existing location-based methods of aggregation of elements into larger analytical units all face some issues limiting their applicability and reliability. The alternative method presented is based on the topology of space as captured by the morphological tessellation. Such a method of partitioning space reflects the influence of each building on the space around it to overcome existing challenges and provide a context-sensitive method. Initial results of the twofold analysis of the topological ability of the morphological tessellation indicate that the type of urban tissue influences the outcome of morphological analysis and that in the case of pre-WW2 traditional European-like urban tissues, all currently available methods of definition of aggregation are relevant and almost interchangeable. However, this is not the case with post-WW2 urban developments, as in them there has been a significant change in the scale of form's granularity. In these cases, urban morphology needs to employ methods which are sensitive to the scale and configuration of urban form and at the same time can detect its granularity. The morphological tessellation and the topology derived from the analysis of its structure seem to be the most successful, sensitive method, suitable for general analysis. All of the methods that have been tested partially solve one of the key issues identified in spatial analysis (MAUP), as data are aggregated independently for each element, and there are no preselected boundaries in play.

No matter the results of the presented analysis, methods extracted from the literature have their role in morphological analysis. However, Euclidean definition and metric reach should be used in specific situations only, due to their limitations, as mentioned above. It is either in stable environments without abrupt changes of granularity or in a definition of larger-scale aggregations, where multiple urban tissues are included. In that case, the main benefits of the morphological tessellation — following the spatial configuration of urban patterns — is not so crucial and from certain scale does not even provide added value.

For the analysis done on the small scale (scale of urban tissue and smaller), Euclidean definition and metric reach do not provide stable information, unlike KNN (which always captures the same number of elements of similar informational value, but not the same relationship) and the topology of morphological tessellation. Moreover, aggregation defined via the topology of morphological tessellation may be used even on the smallest scales of one or two steps as it will always capture intended comparable information based on the relationship of elements.

6.6 Summary

The method of morphometric analysis presented in this thesis is designed with applicability in mind. To ensure that the paucity of viable input data does not limit it, the analytical framework is based on fundamental morphometric elements only. Those reflect fundamental elements known in urban morphology - building, street and, to a degree, the plot. Plots are commonly seen as the ideal spatial division for morphological analysis, but they also have their drawbacks, causing the limited applicability of plot-based methods and, more importantly, the reduced reliability of results obtained by employing them. This chapter tries to address some of the issues characterising the definition of the plot and plot boundaries, the availability and accessibility of plot data and the labour intensiveness of manually extracting reliable plot-level information, aspects that limit the potential of urban morphometrics. The need to objectively define a unit of analysis able to capture the smallest and arguably most fundamental level of spatial subdivision, and to develop a reliable and replicable method to generate and measure it, is the rationale behind the *morphological cell* unit and the *morphological tessellation* method.

The universal and algorithmic nature of the proposed morphological tessellation has the potential to scale up morphometric analysis with minimum effort to the large scale, while significantly reducing the interpretative input of the analysts along the process. This latter property of morphological tessellation appears to be particularly relevant to making large scale morphometric analysis viable and take full advantage of big data in the GIS area. The robustness of the proposed method and the validity of the proposed spatial unit of analysis is verified through the assessment of 12 representative morphometric characters and the application of three different quantitative comparative methods, Spearman's cor-

relation, NRMSD and accuracy of LISA, aimed at evaluating the similarity of information between morphological cells and cadastral plots.

The morphological tessellation, as tested and presented in this chapter, offers a different approach to spatial division whilst still capturing a level of quality of information on urban form that is, to a degree, similar to that conveyed by the plot. Findings presented in this chapter indicate that there is a partial overlap between the information derived from cadastral plots and the one derived from morphological tessellation. The degree of this overlap depends on the category of morphometric characters and the type of urban context, but for certain types of morphologic analysis, it is large enough to consider the information comparable to the one derived from plots. At the same time, it is important to keep in mind that morphological cells cannot fully *replace* plots in the understanding and analysis of urban form patterns.

However, the morphological tessellation is a step towards achieving consistency in urban morphology in both definitions of the smallest spatial unit and analytical aggregation. The advantage of morphological tessellation is that it limits the data dependency as it is based on building footprints only and allows the elimination of subjectivity in the partitioning of space. Most importantly, it is context-sensitive, allowing the researcher to use the same method across different types of urban tissues whilst still get comparable information, much needed for reliable results of any statistical analysis.

On top of an application of tessellation itself, this chapter proposes the tessellation-based relational framework for morphometric analysis of urban form based on the idea of overlapping elements and their aggregations. The resulting description is then based on the semi-lattice of relationships between individual subsets of measurable features. Since tessellation seemingly holds as a spatial unit, the proposed tessellation-based relational framework will be used and tested in the next chapters as a basis of morphometric analysis.

Finally, the overview of morphological aggregations and their linkage to a mixture problem of identification of OTU resulted in a working hypothesis of urban tissue types as OTU for numerical taxonomy of urban form. To which degree can we operationalise this idea and how to use morphometrics to delineate tissue types prior taxonomy are questions left for the next chapter.

Chapter 7

Identification of tissue types through urban morphometrics

The previous chapter defined the building blocks for morphometric analysis and proposed a conceptualisation needed for a comprehensive description of urban form, namely via relational framework of urban form. That can be now implemented in the next step towards a taxonomy of urban form - the algorithmic identification of urban tissue types as operational taxonomic units.

This chapter aims to provide theoretical and practical grounds to the method of automatic detection of distinct types of urban tissues. While similar research has been done before (Schirmer and Axhausen, 2015; Araldi and Fusco, 2019; Berghauser Pont *et al.*, 2019; Bobkova *et al.*, 2019), it was never linked to the coherent theory of morphometrics and numerical taxonomy, nor it was both inclusive in terms of a spectrum of characters used within a model and the spatial extent (see Chapter 2). Following pages present a method which aims to be inclusive in these terms and at the same time automatised and efficient to allow for examination of large datasets spanning across metropolitan regions.

Structurally, this chapter is divided into two main sections. The first one is focusing on the methodological propositions and briefly discuss theoretical grounds of the whole approach based on numerical taxonomy (section 7.1.1), following with specification of two types of morphometric characters used within this research - primary (section 7.1.2.1) and contextual (section 7.1.2.2). The second part of methods (section 7.1.3) outlines the

cluster analysis and introduces a Gaussian Mixture Model clustering together with the issues of the selection of the optimal number of clusters and scalability of the method.

The second major section (7.2) applies the methods on the case study of Prague, Czechia and presents the results of all steps, eventually presenting the urban tissue types detected using cluster analysis based on morphometric assessment.

7.1 Methodological proposition

The automatic detection of urban tissue types consists of multiple procedural steps detailed in the following section. It first requires specification of the principle of the recognition itself, followed by the design of actual methodological steps, starting from identification of morphometric characters for individual elements, finishing with the clustering algorithm detecting clusters. The structure of the method is reflected in the structure of the following sections.

7.1.1 Tissue type as a homogenous cluster

Building on the propositions outlined in chapters 5 and 6, it is possible to define the hypothesis of urban tissue type recognition based on relational framework presented in the previous chapter. The working hypothesis hence stands as follows:

Urban tissue types can be recognised by empirical measuring of the physical structure of urban fabric represented by *the relational analytical framework of urban form* in the form of homogenous clusters.

The concept of urban tissue discussed in the previous chapter is fundamental for the understanding of the structure of cities we live in but at the same time a bit elusive in what *distinct* in the definition means. How much distinct two parts of the urban fabric needs to be to become different tissues? Who makes the decision, and based on what ground? While some have partial answers to these questions (Kropf, 1996), one remains unanswered. How to consistently identify urban tissues across metropolitan areas in an automatised, algorithmic way?

This research attempts to answer that question using the recognition of clusters based on the principles of numerical taxonomy (see chapter 3). Whilst previous chapters identified urban tissue as an OTU of urban form; we first have to identify those. For the cluster analysis recognising tissue types, we focus on a dual feature building-tessellation cell as the smallest entity of urban form. The whole cluster recognition is then based on the assumption that features recognised as a part of the same cluster (*species*) are, in fact, elements of the single urban tissue (where continuous) or of multiple individuals of the same kind of urban tissue (where discontinuous).

The urban form is full of exceptions from the pattern. Individual plots follow the different development process and are, in some cases, amalgamated or split (Conzen, 1960). That does not happen to the rest of the same tissue at the same time (while it might or might not later), causing the constant emergence of exceptions from the pattern. The proposed method is working with two kinds of characters - primary and contextual to overcome the issue of exceptions.

The primary characters are those focusing on the individual elements and their relationships as identified in the relational framework (Chapter 6). These are mostly following what a character would be in biology. A typical example could be building height or area. Both are specific to each building. In the context of plots with subsequent internal development, buildings in the head and the tail of the plot might have significantly different values, i.e. exceptions from a continuous pattern.

As primary characters, by definition, do not describe the pattern but rather its elements, they are not optimal input for pattern detection algorithms. The second type of characters, contextual, has been designed specifically to turn values captured by primary characters into values describing the characters' tendency in the area - to describe the pattern. As such, values are spatially lagged and can be used as an input for the cluster analysis. In the end, the data captured by contextual characters are used to cluster individual building-tessellation cell entities to statistically homogenous clusters, each capturing distinct kind of urban tissue.

The following section will detail the use of primary characters, contextual characters and the clustering method itself.

7.1.2 Morphometric characters

The main scope of this research is not to develop new morphometric characters (even though there are some to fill the gaps), but to use existing knowledge in urban morphometrics and combine it in a systematic framework providing a comprehensive description of urban form. Chapter 4 mapped in detail the existing characters used across the field and the resulting database (Appendix A4) and classification is the basis for selection and definition of primary characters and to some extent, even contextual characters.

7.1.2.1 Primary characters

Primary characters describe different elements and their relationships as are identified within the relational framework of urban form. Building on the definition[1] of the term *primary* from Oxford English Dictionary (Oxford English Dictionary, 2020b), we can define primary characters within the context of this cluster analysis as *characters occurring first in a sequence of methodological steps capturing individual features of urban form elements and their fundamental relations*. The link to the relational framework is crucial here as it defines which relations are meant and later reflected in the whole recognition model.

The choice of characters affects the result of cluster analysis; therefore, a potential embedded bias can cause a distortion of clustering results towards aspects occurring multiple times or other adverse effects. For that reason, specific principles of characters selection were defined.

7.1.2.1.1 Principles of character selection and definition The idea of morphometric recognition of tissue types is based on mixture problem, and the selection of morphometric characters then build on the principles used within the selection of taxonomic characters in biology, as defined by Sneath and Sokal (1973). Building on the biological experience brings methodological grounds to the selection, and it is expected that a final set of characters selected according to these rules will describe an urban form suitable for

[1] " Occurring or existing first in a sequence of events; belonging to the beginning or earliest stage of something; first in time." (Oxford English Dictionary n.d.b)

recognition of clusters. However, the validity of the set is still only hypothetical, unlike the validity of individual characters which is tested throughout the selection process.

Selection strategy is tied to the classification of morphometric characters into categories as defined in chapter 4 and, more importantly, to the relational framework of urban form. There are three top-level aims for the selection of primary characters. The set should:

1. **Capture structural complexity of urban form by covering all categories of morphometric characters:**
    - dimension
    - shape
    - spatial distribution
    - intensity
    - connectivity
    - diversity

Each category captures different aspects of urban form, which all should be incorporated to derive a complex description of urban patterns. However, as different categories tend to focus on different scales and elements (see chapter 3), not all are likely to be equally represented. That is not an issue, rather a consequence of the nature of characters and the aim of the recognition model.

2. **Capture all fundamental elements of urban form**

In this case, in the context of the relational framework, these are:

- building
- street network
- morphological cell

Urban form is composed of multiple elements. Hence all fundamental ones should be captured. Here the attempt is to use as little of input data as possible, to extend the applicability of the whole model. Other elements (e.g., plot, open space, greenery) could

be included, and the resulting model would likely be more precise, but the availability of such data is limited. This research uses only the three elements of urban form defined in the relational framework (coming from two data sources as MT is generated); hence this aim is focused on these only.

3. **Capture cross-scale complexity of urban form by covering all meaningful topological scales**

The relational framework defines three topological scales:

- single/small
- medium
- large

For tissue type recognition, not all scales are equally meaningful, as the spatial extent of tissues is usually restricted. However, S, M and L are all relevant for the scale of urban tissue and should all be represented. The city and its urban form are composed of complexities occurring on different scales. Capturing them all together within the single model allows the description of cross-scale complexity needed for systematic morphometric characteristics of built-up patterns.

To fulfil the aims, the relational framework comes to help with defined subsets as a combination of elements and scales, combining second and third aim into a single solution. Each of the subsets represents specific relations between specific elements, hence covering all subsets will help the pursuit of complex description. Then, having subsets, meaningful characters for each subset should be identified. The following procedure directly builds on the Sneath and Sokal (1973) to determine a methodical approach to the selection of the final set of morphometric characters. Steps of selection and elimination should follow this sequence:

1. **Extract all characters used in relevant literature**

The starting point should be a wide range of characters used within relevant literature, as such characters are already tested, and it is expected that they bear significant meaning

in the description of urban form. This extraction has already been done in Chapter 4, so the resulting database of morphometric characters can be directly used. This database works as the principal source of characters. Due to its extent, it is expected that the majority of possible characters is included (keeping in mind identified gaps).

2. **Select characters using data intended to be used within each subset**

Not all characters are based on the same data sources used within this research and relational framework (see Chapter 6 for details). Some can be adapted (e.g., a morphological cell can be, in some cases, used as a spatial unit where plot is normally used), but some are based on the different sources of data. Characters which could not be used within subsets of the relational framework are then excluded from the initial selection.

3. **Adapt characters to fit the framework**

Those characters which are applicable, but are not readily available to be used within the relational framework should be adapted to fit the framework. It comprises mostly translation of plot-based characters to cell-based and metric-based characters into topology-based. Adaptation should be made with a sense of the meaning of each character which should not be significantly changed. Otherwise, its foundation in literature would be questionable, and a character should be seen as a newly developed one.

4. **Eliminate logical correlations**

Logically correlated characters should be omitted. Otherwise, the feature which is causing the correlation could distort the results of the clustering. Fully correlated characters caused by the causality (because A equals 1, B will be 1) have to be excluded, and only one should be kept. Partial logical correlation depends on the nature of other factors that are affecting character. If they reflect variation, we can include them. Also, "*characters that are tautological - those that are true by definition as well as those that are based on properties known to be obligatory - should not be included.*" (Sneath and Sokal, 1973, p. 104)

5. **Eliminate ineffective characters**

Due to the nature of the analysis, working with large-scale data, the process of measuring has to be computationally efficient. Some of the characters are not easily measurable, and it has to be evaluated whether the value of the characters would balance the difficulty of implementation and/or computational demand. Examples of such characters could be those based on expensive generative elements like axial maps (Hillier, 1996) or topological skeleton (Schirmer and Axhausen, 2015), or characters which implementation details are improperly documented in the literature.

6. **Add characters to minimise gaps in subsets**

The database of characters showed an imbalance of different categories and pointed out gaps, especially in the measuring of diversity. Moreover, the relational framework brings some subsets, which are often overlooked in the existing literature. While the overall balance between subsets and categories is only theoretical and would not reflect different nature and importance of different subsets, each of them should be sufficiently covered to capture the complex phenomenon of urban form structure. That may involve the development of new characters.

7. **Exclude invariant characters**

Some characters might be invariant over the entire sample. Those should not be included as they are not bearing any morphometric value. However, this exclusion is an ongoing process because it depends on actual measured values. Moreover, the invariance in one sample of data does not mean that the character overall is not valuable.[2]

8. **Limit empirical correlation**

[2]Consider an example of a courtyard area within a building. That will likely show variance in Mediterranean historical context, but invariance in the US sprawled urban tissues. If the study does not aim to be comparable across different contexts, characters like this should be excluded. However, if the study expects later inclusion of additional data, it may be more complicated to select those which should be eliminated. Such a decision needs to be done base don the complete data which will be used within the study.

When we have the evidence that more than one factor affects two correlated characters within a study, regardless of whether this evidence comes from within a study or from outside, we would include both characters; otherwise, we would employ only one. We assume that at least some independent sources of the variation in any empirical correlation unless we have reason to believe otherwise.

9. **Exclude characters which cannot capture patterns**.

Some characters may show random-like spatial distribution, meaning that the geographical location has no relationship to the actual value. These characters do not have the value within this framework, as they are not able to describe a spatial pattern. To test the capability of each character to capture such patterns is used spatial autocorrelation analysis based on global Moran's I (Moran, 1950). Those characters without a significant autocorrelation should be excluded as they do not bear any value in the process of identification of tissue types.

10. **Balance scales and uniqueness of values**.

The set of taxonomic characters has to be balanced regarding the scale as well as *uniqueness* of values. Some of the initially identified characters are possible to measure on different topological scales. Due to the logical correlations between them, only one has to be used. The selection is trying to use the most appropriate in terms of the meaning of the character (which might be more suitable to the street edge than a block, for example). It also aims to limit the characters with limited uniqueness of values. Because the values are always stored on the smallest scale, the values of characters measured on the block scale are shared among all elements in the block. The intention is to limit those characters to a minimum.

The data on selection itself, starting from the database retrieved from chapter 4 is available as Appendix 7.1. It illustrates the selection process determining which characters should be part of the final set. The following section describes the final set of 74 primary characters only.

7.1.2.1.2 Identified set of primary characters Based on the principles described in the section above, the following morphometric characters compose the final set of primary characters.[3]

1. **Area of a building** is denoted as

(4) a_{blg}

and defined as an area covered by a building footprint in m^2 .

2. **Height of a building** is denoted as

(5) h_{blg}

and defined as building height in m measured optimally as weighted mean height (in case of buildings with multiple parts of different height). It is a required input value not measured within the morphometric assessment itself.

3. **Volume of a building** is denoted as

(6) $v_{blg} = a_{blg} \times h_{blg}$

and defined as building footprint multiplied by its height in m^3.

4. **Perimeter of a building** is denoted as

(7) p_{blg}

and defined as the sum of lengths of the building exterior walls in m.

[3]For the implementation details, please refer to the original referred work and to the documentation and code of momepy, which contains Python-based implementation of each character. On top of the definition and related formulas below, classification of characters and references are in Appendix 7.2.

5. **Courtyard area of a building** is denoted as

(8) a_{blg_c}

and defined as the sum of areas of interior holes in footprint polygons in m^2.

6. **Form factor of a building** is denoted as

(9) $FoF_{blg} = \frac{a_{blg}}{v_{blg}^{\frac{2}{3}}}$.

It captures three-dimensional unitless shape characteristic of a building envelope unbiased by the building size (Bourdic *et al.*, 2012).

7. **Volume to façade ratio of a building** is denoted as

(10) $VFR_{blg} = \frac{v_{blg}}{p_{blg} \times h_{blg}}$.

It captures the aspect of the three-dimensional shape of a building envelope able to distinguish building types, as shown by Schirmer and Axhausen (2015). It can be seen as a proxy of volumetric compactness.

8. **Circular compactness of a building** is denoted as

(11) $CCo_{blg} = \frac{a_{blg}}{a_{blgC}}$

where a_{blgC} is an area of minimal enclosing circle. It captures the relation of building footprint shape to its minimal enclosing circle, illustrating the similarity of shape and circle (Dibble *et al.*, 2017).

9. **Corners of a building** is denoted as

(12) $Cor_{blg} = \sum_{i=1}^{n} c_{blg}$

where c_{blg} is defined as a vertex of building exterior shape with an angle between adjacent line segments ≤ 170 degrees. It uses only external shape (`shapely.geometry.exterior`), courtyards are not included. Character is adapted from (Steiniger *et al.*, 2008) to exclude non-corner-like vertices.

10. **Squareness of a building** is denoted as

$$Squ_{blg} = \frac{\sum_{i=1}^{n} Dc_{blg_i}}{n} \quad (13)$$

where D is the deviation of angle of corner c_{blg_i} from 90 degrees and n is a number of corners.

11. **Equivalent rectangular index of a building** is denoted as

$$ERI_{blg} = \sqrt{\frac{a_{blg}}{a_{blgB}}} * \frac{p_{blgB}}{p_{blg}} \quad (14)$$

where a_{blgB} is an area of a minimal rotated bounding rectangle of a building (MBR) footprint and p_{blgB} its perimeter of MBR. It is a measure of shape complexity identified by Basaraner and Cetinkaya (2017) as the shape characters with the best performance.

12. **Elongation of a building** is denoted as

$$Elo_{blg} = \frac{l_{blgB}}{w_{blgB}} \quad (15)$$

where l_{blgB} is length of MBR and w_{blgB} is width of MBR. It captures the ratio of shorter to the longer dimension of MBR to indirectly capture the deviation of the shape from a square (Schirmer and Axhausen, 2015).

13. **Centroid - corner distance deviation of a building** is denoted as

$$CCD_{blg} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(ccd_i - \bar{ccd}\right)^2} \quad (16)$$

where ccd_i is a distance between centroid and corner i and $\bar{ccd}$ is mean of all distances. It captures a variety of shape. As a corner is considered vertex with angle $< 170^{o}$ to reflect potential circularity of object and topological imprecision of building polygon.

14. **Centroid - corner mean distance of a building** is denoted as

(17) $CCM_{blg} = \frac{1}{n}\left(\sum_{i=1}^{n} ccd_i\right)$

where ccd_i is a distance between centroid and corner i. It is a character measuring a dimension of the object dependent on its shape (Schirmer and Axhausen, 2015).

15. **Solar orientation of a building** is denoted as

(18) $Ori_{blg} = |o_{blgB} - 45|$

where o_{blgB} is an orientation of the longest axis of bounding rectangle in a range 0 - 45. It captures the deviation of orientation from cardinal directions. There are multiple ways of capturing orientation of a polygon. As reported by Yan *et al.* (2007), Duchêne *et al.* (2003) assessed five different options (longest edge, weighted bisector, wall average, statistical weighting, bounding rectangle) and concluded a bounding rectangle as the most appropriate. Deviation from cardinal directions is used to avoid sudden changes between square-like objects.

16. **Street alignment of a building** is denoted as

(19) $SAl_{blg} = |Ori_{blg} - Ori_{edg}|$

where Ori_{blg} is a solar orientation of the building and Ori_{edg} is a solar orientation of the street edge. It reflects the relationship between the building and its street, whether it is facing the street directly or indirectly (Schirmer and Axhausen, 2015).

17. **Cell alignment of a building** is denoted as

(20) $CAl_{blg} = |Ori_{blg} - Ori_{cell}|$

where Ori_{cell} is a solar orientation of tessellation cell. It reflects the relationship between a building and its cell.

These seventeen characters are capturing aspects of individual building (topological context 0). Following are measuring aspects of tessellation cells on the same level.

18. **Longest axis length of a tessellation cell** is denoted as

(21) $LAL_{cell} = d_{cellC}$

where d_{cellC} is a diameter of the minimal circumscribed circle around the tessellation cell polygon. The axis itself does not have to be fully within the polygon. It could be seen as a proxy of plot depth for tessellation-based analysis.

19. **Area of a tessellation cell** is denoted as

(22) a_{cell}

and defined as an area covered by a tessellation cell footprint in m^2.

20. **Circular compactness of a tessellation cell** is denoted as

(23) $CCo_{cell} = \frac{a_{cell}}{a_{cellC}}$

where a_{cellC} is an area of minimal enclosing circle. It captures the relation of tessellation cell footprint shape to its minimal enclosing circle, illustrating the similarity of shape and circle.

21. **Equivalent rectangular index of a tessellation cell** is denoted as

(24) $ERI_{cell} = \sqrt{\frac{a_{cell}}{a_{cellB}}} * \frac{p_{cellB}}{p_{cell}}$

where a_{cellB} is an area of the minimal rotated bounding rectangle of a tessellation cell (MBR) footprint and p_{cellB} its perimeter of MBR. It is a measure of shape complexity identified by Basaraner and Cetinkaya (2017) as a shape character of the best performance.

22. **Solar orientation of a tessellation cell** is denoted as

$$Ori_{cell} = |o_{cellB} - 45| \tag{25}$$

where o_{cellB} is an orientation of the longest axis of bounding rectangle in a range 0 - 45. It captures the deviation of orientation from cardinal directions.

23. **Street alignment of a building** is denoted as

$$SAl_{cell} = |Ori_{cell} - Ori_{edg}| \tag{26}$$

where Ori_{cell} is a solar orientation of tessellation cell and Ori_{edg} is a solar orientation of the street edge. It reflects the relationship between tessellation cell and its street, whether it is facing the street directly or indirectly.

24. **Coverage area ratio of a tessellation cell** is denoted as

$$CAR_{cell} = \frac{a_{blg}}{a_{cell}} \tag{27}$$

where a_{blg} is an area of a building and a_{cell} is an area of related tessellation cell (Schirmer and Axhausen, 2015). Coverage area ratio (CAR) is one of the commonly used characters capturing *intensity* of development. However, the definitions vary based on the spatial unit.

25. **Floor area ratio of a tessellation cell** is denoted as

$$FAR_{cell} = \frac{fa_{blg}}{a_{cell}} \tag{28}$$

where fa_{blg} is a floor area of a building and a_{cell} is an area of related tessellation cell. Floor area could be computed based on the number of levels or using an approximation based on building height.

26. **Length of a street segment** is denoted as

(29) l_{edg}

and defined as a length of a `LineString` geometry in metres (Gil *et al.*, 2012; Dibble *et al.*, 2017).

27. **Width of a street profile** is denoted as

(30) $w_{sp} = \frac{1}{n} \left(\sum_{i=1}^{n} w_i \right)$

where w_i is width of a street section i. The algorithm generates street sections every 3 meters alongside the street segment, and measures mean value. In the case of the open-ended street, 50 metres is used as a perception-based proximity limit (Araldi and Fusco, 2019).

28. **Height of a street profile** is denoted as

(31) $h_{sp} = \frac{1}{n} \left(\sum_{i=1}^{n} h_i \right)$

where h_I is mean height of a street section i. The algorithm generates street sections every 3 meters alongside the street segment, and measures mean value (Araldi and Fusco, 2019).

29. **Height to width ratio of a street profile** is denoted as

(32) $HWR_{sp} = \frac{1}{n} \left(\sum_{i=1}^{n} \frac{h_i}{w_i} \right)$

where h_I is mean height of a street section i and w_i is the width of a street section i. The algorithm generates street sections every 3 meters alongside the street segment, and measures mean value (Araldi and Fusco, 2019).

30. **Openness of a street profile** is denoted as

$$Ope_{sp} = 1 - \frac{\sum hit}{2\sum sec} \tag{33}$$

where $\sum hit$ is a sum of section lines (left and right sides separately) intersecting buildings and $\sum sec$ total number of street sections. The algorithm generates street sections every 3 meters alongside the street segment.

31. **Width deviation of a street profile** is denoted as

$$wDev_{sp} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(w_i - w_{sp}\right)^2} \tag{34}$$

where w_i is width of a street section i and w_{sp} is mean width. The algorithm generates street sections every 3 meters alongside the street segment.

32. **Height deviation of a street profile** is denoted as

$$hDev_{sp} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(h_i - h_{sp}\right)^2} \tag{35}$$

where h_i is height of a street section i and h_{sp} is mean height. The algorithm generates street sections every 3 meters alongside the street segment.

33. **Linearity of a street segment** is denoted as

$$Lin_{edg} = \frac{l_{eucl}}{l_{edg}} \tag{36}$$

where l_{eucl} is Euclidean distance between endpoints of a street segment and l_{edg} is a street segment length. It captures the deviation of a segment shape from a straight line. It is adapted from Araldi and Fusco (2019).

34. **Area covered by a street segment** is denoted as

$$a_{edg} = \sum_{i=1}^{n} a_{cell_i} \quad (37)$$

where a_{cell_i} is an area of tessellation cell i belonging to the street segment. It captures the area which is likely served by each segment.

35. **Buildings per meter of a street segment** is denoted as

$$BpM_{edg} = \frac{\sum blg}{l_{edg}} \quad (38)$$

where $\sum blg$ is a number of buildings belonging to a street segment and l_{edg} is a length of a street segment. It reflects the granularity of development along each segment.

36. **Area covered by a street node** is denoted as

$$a_{node} = \sum_{i=1}^{n} a_{cell_i} \quad (39)$$

where a_{cell_i} is an area of tessellation cell i belonging to the street node. It captures the area which is likely served by each node.

37. **Shared walls ratio of adjacent buildings** is denoted as

$$SWR_{blg} = \frac{p_{blg_{shared}}}{p_{blg}} \quad (40)$$

where $p_{blg_{shared}}$ is a length of a perimeter shared with adjacent buildings and p_{blg} is a perimeter of a building. It captures the amount of wall space facing the open space (Hamaina *et al.*, 2012).

38. **Alignment of neighbouring buildings** is denoted as

$$Ali_{blg} = \frac{1}{n} \sum_{i=1}^{n} |Ori_{blg} - Ori_{blg_i}| \quad (41)$$

where Ori_{blg} is the solar orientation of a building and Ori_{blg_i} is the solar orientation of building i on a neighbouring tessellation cell. It calculates the mean deviation of solar orientation of buildings on adjacent cells from a building. It is adapted from Hijazi *et al.* (2016).

39. **Mean distance to neighbouring buildings** is denoted as

$$NDi_{blg} = \frac{1}{n} \sum_{i=1}^{n} d_{blg,blg_i} \quad (42)$$

where d_{blg,blg_i} is a distance between building and building i on a neighbouring tessellation cell. It is adapted from Hijazi *et al.* (2016). It captures the average proximity to other buildings.

40. **Weighted neighbours of a tessellation cell** is denoted as

$$WNe_{cell} = \frac{\sum cell_n}{p_{cell}} \quad (43)$$

where $\sum cell_n$ is a number of cell neighbours and p_{cell} is a perimeter of a cell. It reflects granularity of morphological tessellation.

41. **Area covered by neighbouring cells** is denoted as

$$a_{cell_n} = \sum_{i=1}^{n} a_{cell_i} \quad (44)$$

where a_{cell_i} is area of tessellation cell i within topological distance 1. It captures the scale of morphological tessellation.

42. **Reached cells by neighbouring segments** is denoted as

$$RC_{edg_n} = \sum_{i=1}^{n} cells_{edg_i} \quad (45)$$

where $cells_{edg_i}$ is number of tessellation cells on segment i within topological distance 1. It captures accessible granularity.

43. **Reached area by neighbouring segments** is denoted as

(46) $a_{edg_n} = \sum_{i=1}^{n} a_{edg_i}$

where a_{edg_i} is an area covered by a street segment i within topological distance 1. It captures an accessible area.

44. **Degree of a street node** is denoted as

(47) $deg_{node_i} = \sum_{j} edg_{ij}$

where edg_{ij} is an edge of a street network between node i and node j. It reflects the basic degree centrality.

45. **Mean distance to neighbouring nodes from a street node** is denoted as

(48) $MDi_{node} = \frac{1}{n} \sum_{i=1}^{n} d_{node,node_i}$

where $d_{node,node_i}$ is a distance between node and node i within topological distance 1. It captures the average proximity to other nodes.

46. **Reached cells by neighbouring nodes** is denoted as

(49) $RC_{node_n} = \sum_{i=1}^{n} cells_{node_i}$

where $cells_{node_i}$ is number of tessellation cells on node i within topological distance 1. It captures accessible granularity.

47. **Reached area by neighbouring nodes** is denoted as

(50) $a_{node_n} = \sum_{i=1}^{n} a_{node_i}$

where a_{node_i} is an area covered by a street node i within topological distance 1. It captures an accessible area.

48. **Number of courtyards of adjacent buildings** is denoted as

(51) $NCo_{blg_{adj}}$

where $NCo_{blg_{adj}}$ is a number of interior rings of a polygon composed of footprints of adjacent buildings (Schirmer and Axhausen, 2015).

49. **Perimeter wall length of adjacent buildings** is denoted as

(52) $p_{blg_{adj}}$

where $p_{blg_{adj}}$ is a length of an exterior ring of a polygon composed of footprints of adjacent buildings.

50. **Mean inter-building distance between neighbouring buildings** is denoted as

(53) $IBD_{blg} = \frac{1}{n} \sum_{i=1}^{n} d_{blg,blg_i}$

where d_{blg,blg_i} is a distance between building and building i on a tessellation cell within topological distance 3. It is adapted from Caruso *et al.* (2017). It captures the average proximity between buildings.

51. **Building adjacency of neighbouring buildings** is denoted as

(54) $BuA_{blg} = \frac{\sum blg_{adj}}{\sum blg}$

where $\sum blg_{adj}$ is a number of joined built-up structures within topological distance three and $\sum blg$ is a number of buildings within topological distance 3. It is adapted from Vanderhaegen and Canters (2017).

52. **Gross floor area ratio of neighbouring tessellation cells** is denoted as

$$GFAR_{cell} = \frac{\sum_{i=1}^{n} FAR_{cell_i}}{\sum_{i=1}^{n} a_{cell_i}} \quad (55)$$

where FAR_{cell_i} is a floor area ratio of tessellation cell i and a_{cell_i} is an area of tessellation cell i within topological distance 3. Based on Dibble *et al.* (2017).

53. **Weighted reached blocks of neighbouring tessellation cells** is denoted as

$$WRB_{cell} = \frac{\sum blk}{\sum_{i=1}^{n} a_{cell_i}} \quad (56)$$

where $\sum blk$ is a number of blocks within topological distance three and a_{cell_i} is an area of tessellation cell i within topological distance three.

54. **Area of a block** is denoted as

$$a_{blk} \quad (57)$$

and defined as an area covered by a block footprint in m^2.

55. **Perimeter of a block** is denoted as

$$p_{blk} \quad (58)$$

and defined as lengths of the block polygon exterior in m.

56. **Circular compactness of a block** is denoted as

$$CCo_{blk} = \frac{a_{blk}}{a_{blkC}} \quad (59)$$

where a_{blkC} is an area of minimal enclosing circle. It captures the relation of block footprint shape to its minimal enclosing circle, illustrating the similarity of shape and circle.

57. **Equivalent rectangular index of a block** is denoted as

$$ERI_{blk} = \sqrt{\frac{a_{blk}}{a_{blkB}}} * \frac{p_{blkB}}{p_{blk}} \quad (60)$$

where a_{blkB} is an area of the minimal rotated bounding rectangle of a block (MBR) footprint and p_{blkB} its perimeter of MBR.

58. **Compactness-weighted axis of a block** is denoted as

$$CWA_{blk} = d_{blkC} \times \left(\frac{4}{\pi} - \frac{16(a_{blk})}{p_{blk}^2}\right) \quad (61)$$

where d_{blkC} is a diameter of the minimal circumscribed circle around the block polygon, a_{blk} is an area of a block and p_{blk} is a perimeter of a block. It is a proxy of permeability of an area.(Feliciotti, 2018)

59. **Solar orientation of a block** is denoted as

$$Ori_{blk} = |o_{blkB} - 45| \quad (62)$$

where o_{blkB} is an orientation of the longest axis of bounding rectangle in a range 0 - 45. It captures the deviation of orientation from cardinal directions.

60. **Weighted neighbours of a block** is denoted as

$$wN_{blk} = \frac{\sum blk_n}{p_{blk}} \quad (63)$$

where $\sum blk_n$ is a number of block neighbours and p_{blk} is a perimeter of a block. It reflects granularity of a mesh of blocks.

61. **Weighted cells of a block** is denoted as

$$wC_{blk} = \frac{\sum cell}{a_{blk}} \quad (64)$$

where $\sum cell$ is a number of cells composing a block and a_{blk} is an area of a block. It captures the granularity of each block.

62. **Local meshedness of a street network** is denoted as

(65) $Mes_{node} = \frac{e-v+1}{2v-5}$

where e is a number of edges in a subgraph, and v is the number of nodes in a subgraph (Feliciotti, 2018). A subgraph is defined as a network within topological distance five around a node.

63. **Mean segment length of a street network** is denoted as

(66) $MSL_{edg} = \frac{1}{n} \sum_{i=1}^{n} l_{edg_i}$

where l_{edg_i} is a length of a street segment i within a topological distance 3 around a segment.

64. **Cul-de-sac length of a street network** is denoted as

(67) $CDL_{node} = \sum_{i=1}^{n} l_{edg_i},$ if edg_i is cul-de-sac

where l_{edg_i} is a length of a street segment i within a topological distance 3 around a node.

65. **Reached cells by street network segments** is denoted as

(68) $RC_{edg} = \sum_{i=1}^{n} cells_{edg_i}$

where $cells_{edg_i}$ is number of tessellation cells on segment i within topological distance 3. It captures accessible granularity.

66. **Node density of a street network** is denoted as

(69) $D_{node} = \frac{\sum node}{\sum_{i=1}^{n} l_{edg_i}}$

where $\sum node$ is a number of nodes within a subgraph and l_{edg_i} is a length of a segment i within a subgraph. A subgraph is defined as a network within topological distance five around a node.

67. **Reached cells by street network nodes** is denoted as

(70) $RC_{node_{net}} = \sum_{i=1}^{n} cells_{node_i}$

where $cells_{node_i}$ is number of tessellation cells on node i within topological distance 3. It captures accessible granularity.

68. **Reached area by street network nodes** is denoted as

(71) $a_{node_{net}} = \sum_{i=1}^{n} a_{node_i}$

where a_{node_i} is an area covered by a street node i within topological distance 3. It captures an accessible area.

69. **Proportion of cul-de-sacs within a street network** is denoted as

(72) $pCD_{node} = \frac{\sum_{i=1}^{n} node_i, \text{ if } deg_{node_i}=1}{\sum_{i=1}^{n} node_i}$

where $node_i$ is a node whiting topological distance five around a node. Adapted from (Boeing, 2017b).

70. **Proportion of 3-way intersections within a street network** is denoted as

(73) $p3W_{node} = \frac{\sum_{i=1}^{n} node_i, \text{ if } deg_{node_i}=3}{\sum_{i=1}^{n} node_i}$

where $node_i$ is a node whiting topological distance five around a node. Adapted from (Boeing, 2017b).

71. **Proportion of 4-way intersections within a street network** is denoted as

$$p4W_{node} = \frac{\sum_{i=1}^{n} node_i, \text{ if } deg_{node_i}=4}{\sum_{i=1}^{n} node_i} \quad (74)$$

where $node_i$ is a node whiting topological distance five around a node. Adapted from (Boeing, 2017b).

72. **Weighted node density of a street network** is denoted as

$$wD_{node} = \frac{\sum_{i=1}^{n} deg_{node_i}-1}{\sum_{i=1}^{n} l_{edg_i}} \quad (75)$$

where deg_{node_i} is a degree of a node i within a subgraph and l_{edg_i} is a length of a segment i within a subgraph. A subgraph is defined as a network within topological distance five around a node.

73. **Local closeness centrality of a street network** is denoted as

$$lCC_{node} = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)} \quad (76)$$

where $d(v,u)$ is the shortest-path distance between v and u, and n is the number of nodes within a subgraph. A subgraph is defined as a network within topological distance five around a node.

74. **Square clustering of a street network** is denoted as

$$sCl_{node} = \frac{\sum_{u=1}^{k_v} \sum_{w=u+1}^{k_v} q_v(u,w)}{\sum_{u=1}^{k_v} \sum_{w=u+1}^{k_v} [a_v(u,w)+q_v(u,w)]} \quad (77)$$

where $q_v(u, w)$ are the number of common neighbours of u and w other than v (ie squares), and $a_v(u, w) = (k_u - (1 + q_v(u, w) + \theta_{uv}))(k_w - (1 + q_v(u, w) + \theta_{uw}))$, where $\theta_{uw} = 1$ if u and w are connected and 0 otherwise (Lind *et al.*, 2005).

The final selection consists of 74 morphometric characters spanning across the subsets of the relational framework and covering all categories, even though not equally.[4] The set is a result of the identification process proposed above. As such, it should provide an unbiased and non-skewed description of each of the elements.

7.1.2.2 Contextual characters

Looking at the primary characters and their spatial distribution, they could be abrupt and do not necessarily capture urban patterns as we would like them to (even though all capture some patterns as per spatial autocorrelation).

The characters defined above have to be expressed using their contextual, spatially lagged versions to become useful for pattern detection within the recognition model, which does not employ direct spatial constraints. *Context* here is defined as vicinity of each tessellation cell within three topological steps on morphological tessellation. That covers approximately 40 nearest neighbours (median 40, standard deviation ~13.4 based on Prague) providing a balance between the spatial extent large enough to capture a pattern and at the same time small enough not to over-smooth boundaries between different patterns (see Appendix 7.3 for sectional diagram analysis).

Within this method, four types of contextual characters are proposed. One is capturing a local central tendency and three capturing the properties of the distribution of values within the context. For each of the primary characters, each of the contextual is then calculated and then used within the clustering algorithm itself. The resulting set of used characters is then composed of 4 times 74 characters, giving 296 individual contextual characters.

[4]The balance across categories within the specific set is not required as different categories offer different information relevant for different purposes.

7.1.2.2.1 Local central tendency Statistics knows central tendency as a measure of a typical value for a probabilistic distribution (Weisberg and Weisberg, 1992, p. 2). Based on a set of data of unknown distribution, central tendency aims to simplify the whole set into one representative number. In the case of morphometric characters, we can measure the central tendency of values of a single character across the whole case study, but that would not give us much information. As contextual characters are defined on three topological steps, it is proposed to measure *local central tendency*, thus a value unique for each building measured as a typical within its immediate context.

Commonly used measures of central tendency are mean, median or mode (Wilcox and Keselman, 2003). Each of them fits a different purpose. If one wants to use the arithmetic mean to determine central values, underlying distribution should not be skewed. Otherwise, outliers may significantly affect the resulting value. A mode is, by definition, not suitable for continuous variables like those obtained in primary characters. Median is the most robust of all, measuring the middle value. However, the robustness comes at a cost - the shape of a distribution is not reflected at all. Another option is to find a middle ground between easily distorted mean and robust median using truncated mean. Instead of computing arithmetic mean of the whole distribution, we can work with interquartile (smallest and largest 25% are omitted) or interdecile (smallest and largest 10% are omitted) range to minimise the outlier effect on the mean.

The distribution of values of individual characters vary and in some cases, tends to be skewed. As shown in Appendix 7.4 analysing the difference between mean, interdecile mean, interquartile mean and median (being equal to extremely truncated mean) on a selection of 8 characters, it is clear, that majority of data is somewhat asymmetric, causing volatility of mean, which should not be used in such cases. The question is then limited to the distinction between the median and truncated means (leaving aside midhinge and similar estimators). The data indicate that the difference between median and interquartile mean is minimal (but still present, e.g., in the case of *shared walls ratio*). As interquartile mean uses more information than the median, while being similarly robust to outliers, this research settles on implementation of the interquartile mean as a measure of local central tendency, denoted as

$$(78)\quad IQM_{ch} = \frac{2}{n}\sum_{i=\frac{n}{4}+1}^{\frac{3n}{4}} ch_i,$$

where *ch* is selected primary character. Formula assumes sorted values.

7.1.2.2.2 Properties of a distribution Apart from a local central tendency (in the geographical context sometimes present in literature also as spatial lag (Anselin, 2001)), which aims to capture representative value, it is fundamental to understand how the actual distribution of values within the *context* looks like.

That could be approached in multiple ways. Three notable are 1) capturing the *diversity* of values within the local context, 2) measuring the *statistical dispersion* of values, and 3) measuring *similarity of a target and an actual distribution* of values, like in the case of inequality.

7.1.2.2.2.1 Diversity While discussion on the importance of diversity has been central to urban discourse since the era of Jane Jacobs (Jacobs, 1961), there is not a very wide range of characters actually measuring diversity. The research focuses mostly on Simpson's diversity index (Bobkova *et al.*, 2017; Feliciotti, 2018), developed initially for categorical, not continuous variables and hence relying on pre-defined "bins" (classes of values). For example, Bobkova *et al.* (2017) use this index to measure the diversity of plot sizes, but their binning into intervals based on the actual case-specific values makes the comparability of outcomes limited: if we apply the same formula to another place, we will get different binning and different results. This appears to be a rather ubiquitous problem in applying Simpson's diversity index, i.e., it is necessary to set a finite set of pre-established bins prior to undertaking the analysis. However, despite the need for urban morphometric analysis to produce comparable outcomes, it is challenging to ensure specific descriptiveness to "universal" pre-defined bins. The use of the Simpson's diversity index in ecology is encouraged (Jost, 2006) because ecologists have a finite number of groups enabling them to pre-define all bins appropriately (moreover, bins are usually not defined on a continuous numerical scale). However, this is not often the case in urban morphology. Simpson's diversity index and similar characters based on binning provide values specific to individual cases where binning is set and have to be interpreted as such.

Recent literature shows that there might be alternative ways to measure the diversity of morphological characters. Caruso *et al.* (2017) and Araldi and Fusco (2019) applied the Local Index of Spatial Autocorrelation (LISA) in the form of local Moran's I, defined as

"the weighted product of the difference to the mean of the value of a variable at a certain observation and the same difference for all other observations, with more weight given to the observations in close spatial proximity." (Caruso *et al.*, 2017, p. 84) LISA aims to identify clusters of similar values in space, describing their similarity or dissimilarity, which could be seen as a proxy for diversity, but due to the limited number of significant categories (High-High, High-Low, Low-High, Low-Low), its application in this context is limited and somewhat reductionist.

7.1.2.2.2.2 Statistical dispersion The second approach is to measure statistical dispersion, i.e., the ratio to which the distribution is stretched (wide distribution) or squeezed (narrow distribution). Together with the central tendency, dispersion is often used to describe the basic properties of distributions.

There are multiple ways of measuring dispersion. The most used are probably standard deviation, range or interquartile range as examples of *dimensional* (resulting value have the same units as the original character) measures. Dimensional measures of dispersion are the most common as they are generally easy to understand and interpret. Similarly to the measure of central tendency, all can be measured on the full range of values or a limited one, usually again as interquartile (IQ) or interdecile (ID) range. Dimensionless measures are not expressed in the same units as original characters, so while a dimensional measure of dispersion for building area will be in meters, dimensionless will have no units (the values are relative). Among dimensionless measures are the coefficient of variation (CoV) or quartile coefficient of dispersion (QCoD).

7.1.2.2.2.3 Distribution matching The third approach focuses on the comparison of the actual distribution and the ideal distribution. One example is a test whether such a distribution follows the principle of the Power Law used by Salat (2017). However, that is not a straightforward measurement, especially if the distribution is of a different shape; it is hard to quantify the relationship. Specific distribution is also embedded in the Gini index customarily used to measure inequality or indirectly in entropy-based indices like Theil index of inequality (a special case of the generalised entropy index) (Novotnỳ, 2007).

7.1.2.2.2.4 Comparison of potential characters To understand the properties and behaviour of potential characters capturing properties of distributions on the real morphometric data, wide selection of the most relevant characters from each group is analysed as a way of selecting the most appropriate ones to be used as contextual characters.

In terms of diversity measures, the key question is not which one should be used, either Simpson's diversity index as in Bobkova *et al.* (2017) or Gini-Simpson diversity index as in Feliciotti (2018), but how to define binning as that can significantly affect the resulting diversity values. For that reason, Simpson's diversity is tested using *natural breaks* (Jenks, 1967) (number of classes is based on the Goodness of Absolute Deviation Fit (GADF) (Rey and Anselin, 2007)), *Head Tail breaks* (Jiang, 2013) and *quantiles* (5 and 10 bins). The reason for the inclusion of Simpson's diversity index, even though it may not be fully comparable across cases is the fact that the recognition of tissue types is always local, always case-specific.

Dimensional characters capturing dispersion included in comparison are *standard deviation (SD)*, *range*, and *absolute deviations (median - MAD, average - AAD)*. Both standard deviation and range are measured for IQ, ID and unrestricted range of values. Included dimensionless characters are *coefficient of variation (CoV)*, *quartile coefficient of dispersion (QCoD)*.

The last group is represented by both *Gini index*, and *Theil index*; both measured for IQ, ID and unrestricted range of values.

Using four morphometric characters as test data - *area of a building*, *height of a building*, *coverage area ratio of tessellation cell* and *floor area ratio of tessellation cell*, all potential contextual characters listed above are measured on three topological steps around each building. Resulting spatial distribution is visually assessed to eliminate those unfit for pattern recognition, either for relative randomness of result or significant outlier effect (typically present in measures based on unrestricted range of values). Finally, a correlation matrix is used to identify potential overlaps and uniqueness of values leading to the selection of optimal contextual characters.

The results section of this assessment is available in Appendix 7.5.

7.1.2.2.2.5 Resulting selection of contextual characters While the complete results of the analysis are available as Appendix 7.5, the main conclusions are as follows.

Because some of the values follow exponential (power-law or similar) distribution within the whole dataset, the binning method for Simpson's diversity index has to acknowledge that. For that reason, HeadTail Breaks are the ideal method as it is specifically tailored to exponential distributions (Jiang, 2013). Those characters which do not resemble exponential distribution should use natural breaks or similar classification method sensitive to the actual distribution, rather than quantiles, which may cause significant disruptions and very similar values may fall into multiple bins causing high diversity values in place where is not.

Within measures of statistical dispersion, IQ range and IQ standard deviation are better in capturing boundaries between types of development and are robust to outliers. Interquartile range was used by Dibble *et al.* (2017) and is easier to interpret. Due to its definition, CoV tends to infinity when the mean value tends to zero, being very sensitive to changes of the mean.

Theil index and Gini index are both used to assess inequality, but Theil index, unlike Gini, is decomposable to within-group inequality and between-group differences (Novotnỳ, 2007), making it more suitable for spatial analysis than Gini index would be. ID values used within the Theil index are better than other ranges as the resulting analysis is more sensitive, while outlier effect is still minimal. ID captures, for example, inner structures of blocks better than IQ, where such structures might be filtered out. It may help to distinguish between blocks with and without internal buildings.

The final selection of contextual characters is then composed of four distinct uncorrelated characters. Local central tendency is captured by *interquartile mean (IQM)* and describe the most representative value. The local measure of statistical dispersion is represented by *Interquartile range (IQR)* as dimensional character which expresses the range of values around IQM, indicating where the values mostly lie. IQR is denoted as

$$IQR_{ch} = Q3_{ch} - Q1_{ch}, \tag{79}$$

where $Q3_{ch}$ is third quartile of selected primary character and $Q1_{ch}$ first quartile. Formula assumes sorted values. *Interdecile Theil index (IDT)* is denoted as

$$(80)\ IDT_{ch} = \sum_{i=1}^{n} \left(\frac{ch_i}{\sum_{i=1}^{n} ch_i} \ln \left[N \frac{ch_i}{\sum_{i=1}^{n} ch_i} \right] \right),$$

where ch is selected primary character, and describes the (in)equality of distribution of values. Finally, *Simpson's diversity index (SDI)* is denoted as

$$(81)\ SDI_{ch} = \frac{\sum_{i=1}^{R} n_i \left(n_i - 1 \right)}{N(N-1)},$$

where R is richness expressed as number of bins, n_i is the number of features the ith type and N is the total number of features. It captures the presence of various classes of values. Together, these four characters have a potential to describe spatial distribution of morphometric values within a set context.

For the clarity in terms of classification of contextual characters, IQM inherits the category from the primary parental character, while IQR, IDT and SDI all fall into *diversity* category.

After linking together primary and contextual characters, each of the primary 74 characters is represented by all four contextual, based on the values measured within three topological steps on morphological tessellation around each building. That gives 296 contextual characters in total, the set which is spatially autocorrelated by definition and hence can be used within the clustering method to identify distinct homogenous clusters representing tissue types. The fact that all input data for clustering are measured using this spatially lagged method ensures that spatial clusters should be geographically coherent and mostly continuous. The nature of data allows the use of spatially unconstrained clustering methods.

Importance of the proper selection of morphometric characters and the effect it may have on the overall results in not debatable. A robust method described above is employed starting from the selection of primary morphometric characters from literature and their adaptation to fit *relational framework of the urban form* and to minimise the potential error in selection. The resulting set of 74 characters is established to cover a wide range of descriptive features capturing urban form configuration from dimensions of individual elements, through spatial distribution to diversity. Four contextual characters are introduced to describe a local central tendency and variation in the area capturing morphological patterns, rather than a description of individual elements. These, combined,

have the potential to capture the nature of each of the primary characters and their behaviour in the immediate spatial context.

7.1.3 Identification of clusters

The actual identification of urban tissue types is in principle statistical clustering of building/tessellation cell features with similar information about itself and its context. Moreover, clusters, in this case, need to be contiguous[5]. As mentioned above, the solution of the contiguity issue is built in the design of contextual characters. All characters are spatially autocorrelated by design[6]. There is a significant overlap between areas used for computation of contextual characters of two neighbouring cells that indirectly supports contiguity of clustering. However, this solution may result in less defined boundaries between two clusters, and every edge of the cluster needs to be interpreted as fuzzy rather than defined. Specific mitigation of over-smoothing of boundaries is embedded in the design of contextual characters as they are mostly based on truncated values, which not only eliminate outlier effect but also result in more defined boundaries.

The general principle of clustering is using the data to iteratively determine the optimal division of observed data into homogenous clusters. In case of probabilistic methods, this prediction can have associated probability that the chosen cluster is the correct one and have the probability of belonging to every other cluster.

Current progress in machine learning brings various methods to choose from. Every clustering method follows different principles and is able to identify different kinds of clusters. The most common is *k*-means clustering (MacQueen and others, 1967) and its derivatives

[5]Contiguity is not easy to accomplish as spatially constrained clustering methods, which are designed to be contiguous and take into account spatial relationship of clustered elements, like Skater (Assunção et al. 2006) or Max-p Region Problem (Duque et al. 2012) are computationally inefficient, which is multiplied by the size of the datasets used within this research. They would not be able to crunch the amount of data. The second option how to include spatial dimension in clustering is the actual inclusion of x and y coordinates of each object (in case of building likely x and y coordinates of building centroids). The geographical coordinates would then become another two dimensions in the dataset. This solution might work if the number of dimensions is low, and two additional characters could make a significant effect. As the dataset of contextual characters is in composed of 296 dimensions, the simple inclusion of two others might not make much of a difference and not ensure any spatial contiguity.

[6]Median of Moran's I is 0.77, St.Dev 0.12, with values ranging between 0.42 (Square Clustering of Street Network Theil Index) and 0.98 (Gross Density Interquartile Mean) all with *p < 0.001*. Complete Spatial Autocorrelation analysis is available as Appendix 7.6

(k-medoid (Park and Jun, 2009), k-median (Jain and Dubes, 1988) or Gaussian mixture models (Reynolds, 2009)). The algorithm divides observations into predefined k clusters based on the nearest mean value to minimise within-cluster variance based on squared Euclidean distances between observations (Reynolds, 2009). As a result, clusters tend to be of a similar size. In the case of urban form, it is unlikely that each urban typology is equally present, rendering the use of k-means as less fit for the purpose. It is expected that cluster will be on unequal size and also of unequal density - clusters capturing rigid patterns will be more densely packed than those capturing more diverse areas. The clustering algorithm needs to take into account all these requirements stemming from the specificity of urban morphometric data. Moreover, every building is by definition part of some urban tissue, which could be very heterogeneous, meaning that algorithms expecting and identifying noise (in this case buildings which do not belong to any cluster) in the data like DBSCAN (Ester *et al.*, 1996), HDBSCAN (McInnes *et al.*, 2017) or OPTICS (Ankerst *et al.*, 1999) are not ideal either.

7.1.3.1 Gaussian Mixture Model clustering

Clustering method which does reflect the nature of the problem is the Gaussian Mixture Model (GMM), a probabilistic derivative of k-means (Reynolds, 2009). Unlike the k-means itself, it does not rely on squared Euclidean distances only but is based on the assumption that a Gaussian distribution represents each dimension of each cluster. Hence the cluster itself is defined by a mixture of Gaussians. Where k-means looks for clusters of similar extent, GMMs embedded expectation-maximization (EM) algorithm, which allows identification of different shapes. EM is an iterative method which starts from random points (like k-means) but can find the maximum likelihood of parameters of expected underlying Gaussians.

GMM is probabilistic clustering, which means that it defines n components (equal to k in k-means) and their expected underlying Gaussian distributions and then predicts the probability that each observation belongs to each cluster. The exemplar observation A can then belong to cluster *1* with the probability *0.6*, to cluster *2* with the probability *0.35* and to clusters *3 - 9* with probability *<0.01*, considering 9-component-GMM.

The result of GMM applied to the illustrative artificial dataset, as shown on figure 7.1,

illustrates both resulting labelling, which correctly identifies known clusters, and underlying Gaussian distributions shown as ellipses, where the shade reflects the probability that the points in hyperspace belong to the selected cluster.

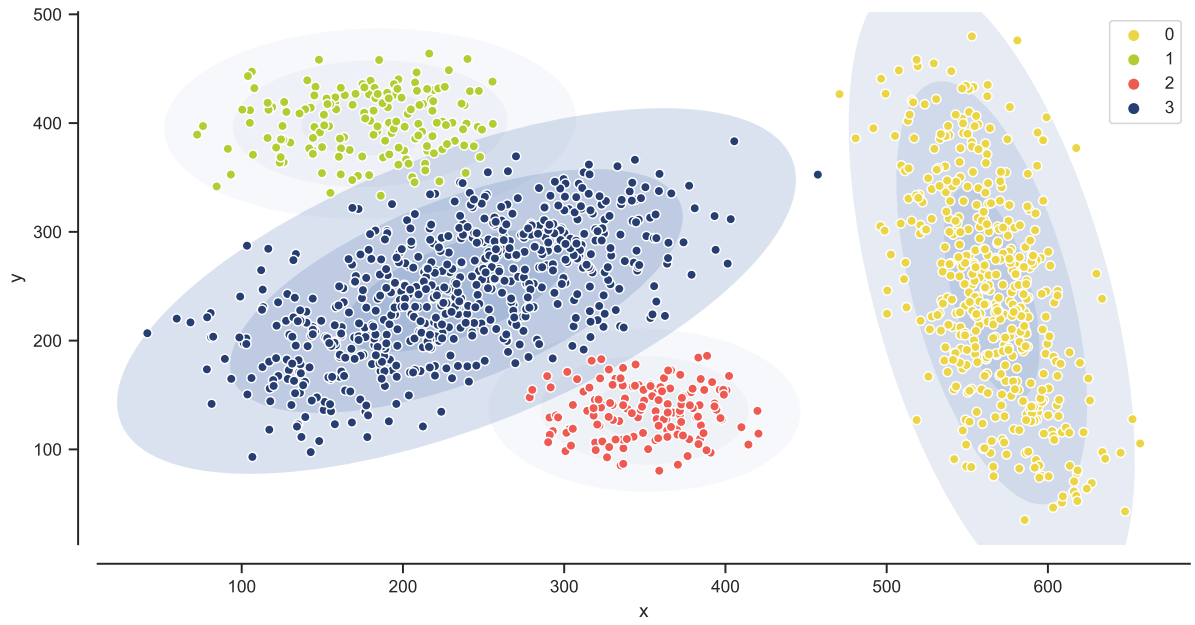


Figure 7.1: GMM clustering (4 components) of the illustrative artificial two dimensional (x, y) dataset containing four known clusters. All clusters are fairly successfully distinguished. The figure also shows underlying Gaussian distributions as ellipses reflecting the probability by the change of the shade.

Because in the first step of GMM, the seed points are placed randomly, this placement might affect the resulting model. This specificity makes GMM non-deterministic clustering, which means that each run will likely result in (slightly) different clusters. GMM has to be done repeatedly in several initializations, of which the best should be used to ensure the stability of the clustering.

Within the context of urban morphology, the method has been applied within a similar classification task by Jochem *et al.* (2020). Within this research, a `sklearn.mixture.GaussianMixture()` implementation of GMM within open-source python package scikit-learn v.0.22 (Pedregosa *et al.*, 2011) is used. Further details on the exact algorithm are available in scikit-learn documentation and code.

7.1.3.2 Levels of clustering resolution and its scalability

The ideal outcome of the tissue type recognition is each cluster as a distinct urban tissue type. However, the definition of urban tissue does not specify the threshold when two similar parts of the city are still the same tissue type and when they become a different one. This issue is mirrored in the clustering method. The ideal outcome of clustering is the optimal number of clusters based on the actual structure of the observed data. That might not be straightforward to determine as better-looking clustering (from the statistical, not visual perspective) might be just overfitted. Moreover, the relation between resulting clusters and urban tissues is always questionable as there is no ground truth for either of them. Detecting 5 large cluster in the whole Prague would likely be based on underfitted model and cluster would not represent urban tissues in the traditional sense, but their aggregations. On the other hand, detecting 100 would likely represent the overfitted model, and each cluster would be only a part of a tissue. It is expected that the statistically optimal number of clusters should be close to what we would typically call urban tissue. However, this link requires further interpretative work, which should happen based on the taxonomy of clusters to allow scale-dependent flexibility.

7.1.3.2.1 Number of components Gaussian Mixture Model clustering requires, similarly to k-means, specification of a number of components of the model (i.e., clusters) before clustering. However, that number is usually not known, especially in the case of urban form. Assumptions can be made based on the expert knowledge, but that would limit the application and unsupervised nature of the whole process and go against the prepositions set in chapters 1 and 5.

The way around is to estimate the ideal number of components based on the goodness of fit of the model for each of them. That means that the GMM is trained multiple times based on the range of feasible options of the number of components and each of the models is then assessed against the whole dataset (to determine how well clusters are distinguished). The assessment is of a quantitative statistical nature, keeping the method relatively unsupervised. The only input researcher needs to make at this stage is an interpretation of the resulting values and the curve of the goodness of fit to specify the number of components for the final clustering.

7.1.3.2.1.1 Goodness of fit The *goodness of fit* measures a fit of a trained model to a set of observations (e.g., the original dataset)(D'Agostino, 1986). It describes how consistent is the distribution of clustered model to the distribution of the whole dataset. With K-means clustering is often used silhouette method, which could, in theory, be used with GMM as well. Another option is measuring the average log-likelihood score. However, the optimal method for GMM is the Bayesian information criterion (BIC), a model based partly on the likelihood function (Schwarz and others, 1978). Unlike similar Akaike information criterion (Akaike, 1973), BIC implements penalisation for a high number of clusters trying to mitigate possible overfitting of the model.

In practice, BIC is measured for each n within the tested range. The lowest the BIC is, the better the model represent original data.

The interpretation of the goodness of fit score is not a question of comparing the numbers only, but understanding the resulting curve. In theory, the lower the BIC score is, the better the model fits the original data. However, it has to be kept in mind that there is a certain confidence interval and that BIC itself penalises a higher number of clusters. The optimal number is not always the one which reaches the lowest BIC score, especially if the score is within the confidence interval of other options. The clustering aims to simplify the whole dataset into the smallest number of meaningful clusters, but not too small. Hence in the situation with multiple options within the same confidence interval, we should select the first significant minimum, i.e., the smallest number of components which has its mean score within the confidence interval of the numerically best fit.

In the ideal case, the BIC curve would reach the minimum for an optimal number of components and then start growing again, making the interpretation relatively straightforward. However, due to the possibility of overfitting, the curve may not culminate but only change the gradient. In such cases, the gradient itself should be analysed and as optimum should be selected a number of components before the flattening of the gradient.

7.1.3.2.1.2 Stability of procedure Non-deterministic nature of GMM means that each of the trials should be repeated multiple times to understand what is the confidence interval of possible outcomes. Testing each number of components only once might lead to incorrect interpretation of results. The ideal situation is to compute multiple runs (the higher the number, the better the result) of each option and plot the confidence interval

to help with the interpretation later. To better understand the magnitude of the effect, the model should be trained multiple times and resulting BIC score should be reported for each of them. The same should happen during the final clustering based on the selected number of components - the model should be initialised repeatedly, and the best of the resulting models should be kept and used.

The result of clustering is never the same, especially with the amount of the data this research is using. There is an inevitable variability, but that is mostly represented by unstable boundaries between clusters rather that significant results in clusters themselves. The boundaries should never be interpreted as a fixed line. There is always a certain degree of fuzziness, which could be captured by an overlay of resulting clusters form multiple models of same parameters.

7.1.3.2.2 Sample-based clustering As the dataset grows, it may become increasingly impossible to perform clustering on the whole dataset, especially if we want our data with a meaningful confidence interval. The calculation of dimensions between components of the model in the hyperspace of 296 dimensions is a demanding task requiring time and computational power. While data for Prague (~140 000 features) could be processed on a desktop with modern multi-core processors within days (multiple options with a confidence interval, not a single run), that is not true for larger metropolitan areas where features count can reach millions. The data like this can be run in the same way on cloud-based services providing significantly more computational power and servers tailored to data analysis, but this solution can be costly.

For that reason, it might be worth training the method on sampled data before classifying the whole dataset. Instead of using all features to train the model, randomly samples subset could be used as a training set for GMM, which, once fitted, could be used to classify the whole dataset. This solution lowers computational demands as the number of features used in the learning process is smaller, but there are also issues with it. The random sample should reflect the structure of the whole dataset to provide results comparable with GMM trained on the whole dataset. However, that is never entirely true. The larger the sample is, the more similar to the complete data is, but at the same time, the effect of sampling on computation is becoming less significant. Even larger samples may, in some cases, miss smaller clusters present in the full-data clustering as features composing

these cluster would not be present in the sample (the smaller the cluster, the higher the probability than it will be missed in the sample).

The decision whether to train GMM on the full or sampled data should reflect the balance between what is ideal (full) and what is possible in certain conditions. The different options of sample-based clustering are tested and compared to the default clustering in the following section, to assess the behaviour of sample-based clustering in the case of Prague. The behaviour will be likely different at different places as the real structure and distribution of values affects the sampling-effect. Places with more diverse structure and several smaller tissues will be probably affected more than places with a homogenous structure where the likelihood of proper sampling of all clusters is higher.

7.1.3.2.3 Sub-clustering There are situations when resulting clustering is not refined enough for the specific analysis. The components are too big, and one may want a better resolution of clustering. One way to do it is to iteratively cluster individual already identified clusters, i.e. to do sub-clustering of existing clusters.

The morphometric dataset is rich in information, so if there is an assumption that a cluster should be divided, it is expected that the difference will be reflected in the data. The reason why it did not split the cluster in two initially is that such a difference is not significant from the perspective of the whole datasets, but it may be significant on a local scale. So when it is appropriate, the same data used for initial cluster recognition can be used again only on the sample belonging to one of the clusters.

The relation of sub-clusters to other than parental cluster is different from the relation between initial clusters themselves, and the difference has to be retained throughout the analysis and has to be correctly interpreted. Doing selective sub-clustering and then approaching initial clusters and sub-clusters as equal is not recommended even though there might be a particular situation when this approach might be viable. However, it has to be done consciously after an assessment of possible consequences.

The other way, aggregating clusters together based on their similarly will be discussed in the next chapter 8.

Either way, it is crucial to acknowledge that clustering is always based on the actual structure of the used data. That means that the result of clustering is always local. Clusters

identified in Prague using solely Prague-based data would not be equal to clusters identified in Amsterdam using Amsterdam-based data only. The structure of both datasets determines what the optimal division is and as both structures are different, the optimal division is done along different lines. It is expected that results will be comparable as the optimal cluster should reflect optimal urban tissues. Chapter 8 will test whether the misalignment is significant or not to further explore the link between two local clustering models.

Selection of the clustering model and its parameters affects the results of identification of urban tissues. The decision has to be made based on detail theoretical considerations of what the behaviour of morphometric datasets likely is. While many of its properties are still unknown, based on the assumptions outlined in this sections, it is believed that GMM, in combination with BIC for determination of the number of GMM components, can identify distinct homogenous clusters as a proxy of urban tissue types.

7.1.4 Data model

The data model representing the elements of the urban form consists of two input and three generated layers, all linked together through the proxy of a building based on the system of unique identifiers according to the structure presented in a table 7.1.

Table 7.1: Presence of different unique identifiers on different data layers. `buildings` contains all of them and are used as a connector.

layer	uID	nID	nodeID	bID
buildings	x	x	x	x
tessellation	x			
street edges		x		
street nodes			x	
blocks				x

Buildings are in the role of connecting elements and contain all identifiers. Morphological tessellation is based on the building layer, and cells hence inherit buildings' `uID`. Street edges are linked to buildings based on the proximity of building centroid to street segment

geometry (the nearest edge is linked using momepy). Street nodes are linked to buildings based on proximity either, but linked node has to be end node of linked nearest edge. Blocks are based on tessellation, and their id is linked to buildings using intersection-based spatial join during their creation.

Momepy uses unique identifiers to efficiently link elements together without the need for repeating costly spatial operations for every relevant character.

7.2 Tissue type recognition | Case study Prague

The first trial of the proposed tissue type recognition method outlined above is the case study of Prague, a dataset which after pre-processing contains 140 315 individual buildings, 22 503 street edges, 16 207 street nodes and 7 395 tessellation-based blocks. Following section reports on each step of the method in terms of both results and interpretation. The overall discussion on the method itself, its relevance and applicability is in chapter 9 and includes results of the taxonomical analysis presented in chapter 8. The validation of results is included in chapter 8.

7.2.1 Primary characters

The basis of the method lies with primary morphometric characters. These continuous variables describe individual aspects of fundamental elements and their combinations based on the relational framework. Following the method, all 74 of them are measured in Prague and then linked to the building-tessellation unit according to the data model. All morphometric characters are measured using momepy classes using reproducible Jupyter notebook presented in Appendix N.

The results of measured primary characters can be explored in two ways - 1) to assess a *spatial* distribution of values, and 2) to assess *statistical* distribution of values.

7.2.1.1 Spatial distribution

The spatial distribution of resulting values, i.e., spatial morphometric patterns, could be projected on maps and assessed visually, to determine the character of a pattern, or statistically. Since the aim of measuring is, eventually, to identify homogenous areas defined by distinct patterns of spatial configuration, each of the characters must capture local patterns. Statistically speaking, each of the characters needs to be spatially autocorrelated, which can be assessed using Moran's I (Moran, 1950)[7].

Based on the visual assessment, there are three types of characters within the measured set, represented by three examples below - 1) patterns with sudden changes, 2) smooth continuous patterns, 3) visually unclear patters.

[7]The same method has been used during the selection of primary characters to ensure that all capture spatial patterns. See Appendix 7.6 for details.

Figure 7.2: Spatial distribution of shared walls ratio of adjacent buildings in the area of Prague's city centre and its surroundings. The figure illustrates clear spatial patterns with the presence of sudden changes.

Figure 7.2 shows *shared walls ratio of adjacent buildings* in the part of Prague's city centre. There is a clear distinction between buildings having shared walls and those standing independently. The values show a relative homogeneity in the centre of the figure (Vinohrady), but high variability in some other places, especially in the Old Prague

neighbourhood (top left). There are sudden changes in values on neighbouring tessellation cells. This pattern is not unique, and it is somewhat expected for characters based on individual elements as these do not have a notion of contiguity.

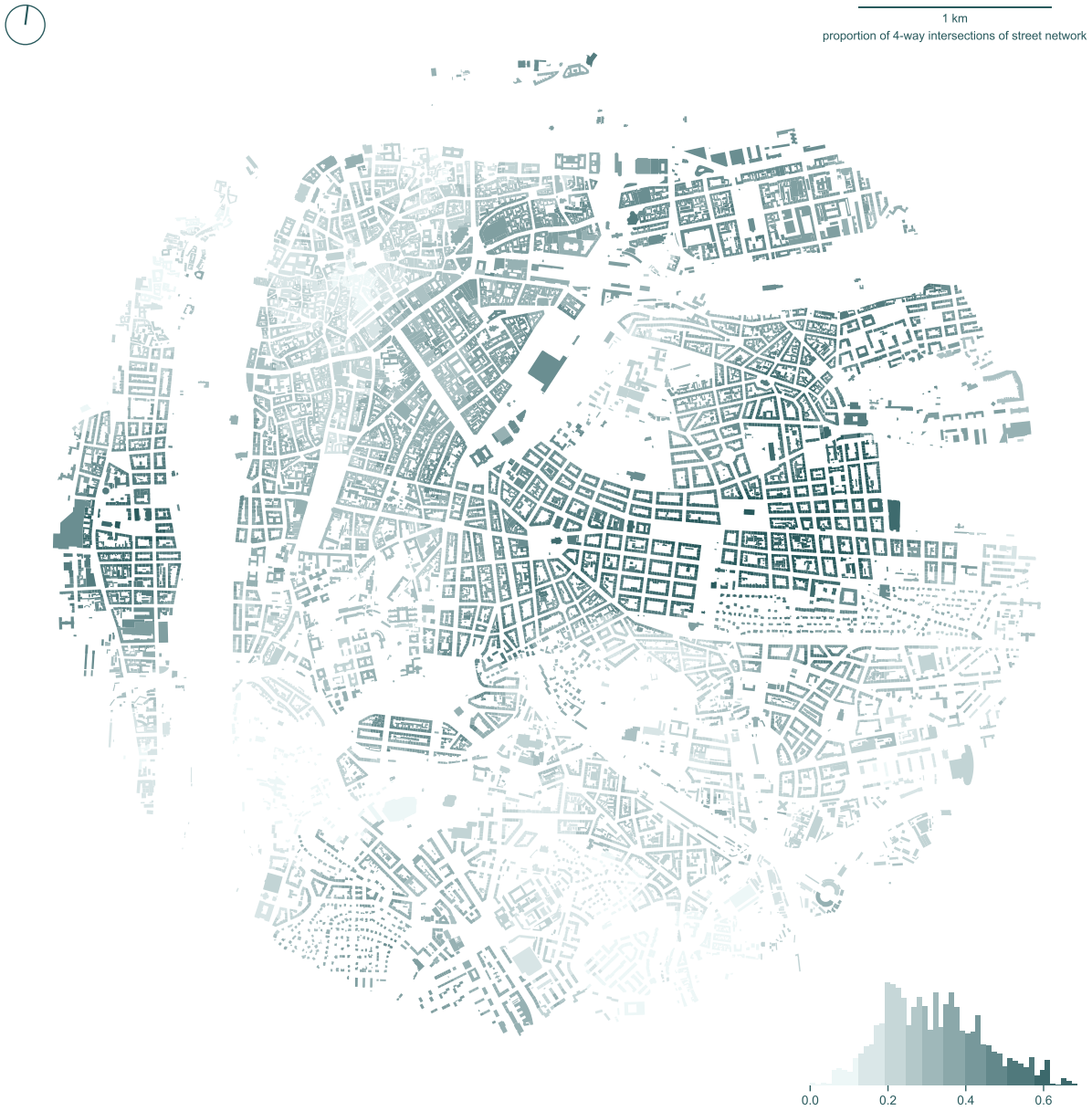


Figure 7.3: Spatial distribution of the proportion of 4-way intersections of the street network in the area of Prague's city centre and its surroundings. The figure illustrates clear continuous spatial patterns with unclear boundaries between low and high values.

The second example on the figure 7.3 shows *proportion of 4-way intersections of street*

network within the same area. This is purely network-based character measuring properties of subgraphs around each network node (i.e. a junction). Subgraphs, by definition, overlap causing the smooth transition of values across the study area. It is relatively simple to describe resulting patterns visually, with high values in more grid-like areas (Vinohrady - centre, Smíchov - left). However, the definition of boundaries between high and low values would be a relatively complicated procedure due to the inherent spatial smoothing. Characters based on a broader topological context tend all to have continuous patterns like this.

Figure 7.4: Spatial distribution of equivalent rectangular index of tessellation cell in the area of Prague's city centre and its surroundings. Figure illustrates visually unclear spatial patterns.

The last, not very frequent though, is the example on figure 7.4 showing visually unclear spatial distribution. The figure shows *equivalent rectangular index of tessellation cell* in the same area. To determine spatial patterns visually require much effort, and still, the results are questionable. This is one of the examples where one might want to exclude

such character for apparent randomness of resulting values. However, visual assessment should not be used for such a decision because it is naturally arbitrary and biased based on the ability of a researcher to detect patterns. For that reason, this work uses Moran's I index of spatial autocorrelation to determine whether a character captures meaningful spatial pattern or not.

Figure 7.5 below shows the value of Moran's I compared to reference distribution and a Moran scatterplot based on the contiguity of morphological tessellation. The I value for the whole of Prague is 0.07, showing significant autocorrelation. It is not a high value, for a reference two previous example have I 0.387 and 0.912 respectively. However, it is still significant (the value itself is likely not within a reference distribution), meaning that even not visually apparent, spatial pattern is still present. The whole set of characters contains a couple of other examples similar to this one, but overall this situation is not a frequent one.

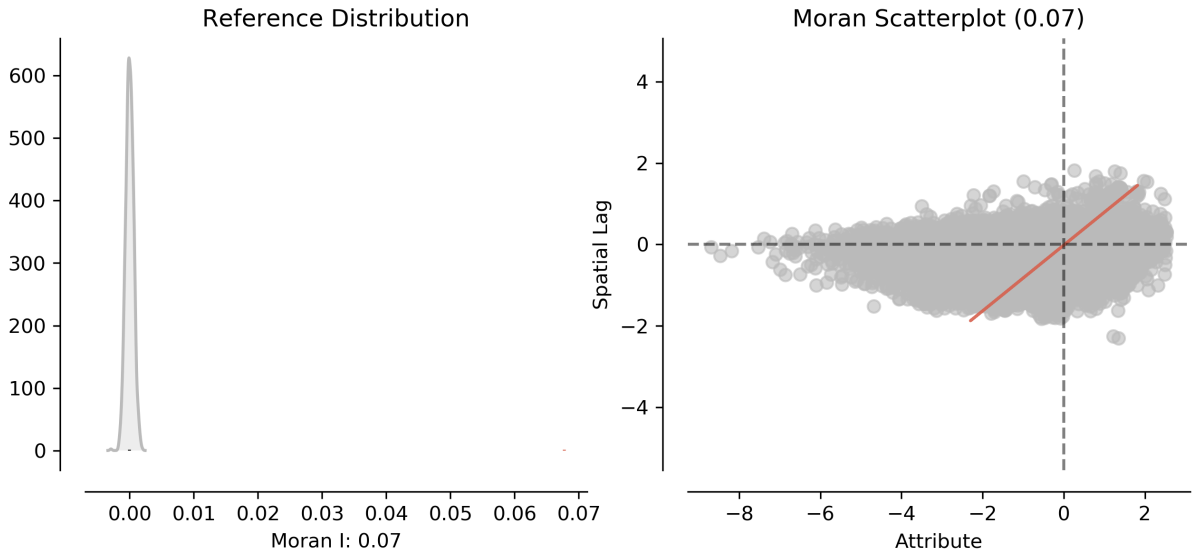


Figure 7.5: Reference distribution in relation to actual Moran's I value and Moran scatterplot of the equivalent rectangular index of tessellation cell based on the whole Prague. The results indicate significant, however weak spatial autocorrelation. (Rey and Anselin, 2007)

Due to the large variety of characters attempting to capture both structural complexity and cross-scale complexity within a single set, the spatial distribution of resulting values may vary. However, all show significant spatial patterns.

7.2.1.2 Statistical distribution

Statistical distributions of resulting values are also different, based on the nature of each character. From the literature is known, that urban context is often described by exponential distributions like a power law (Salat, 2017), but that is far from being a rule for a selected set of morphometric characters. Figure 7.6 shows four examples of distributions as captured in Prague.

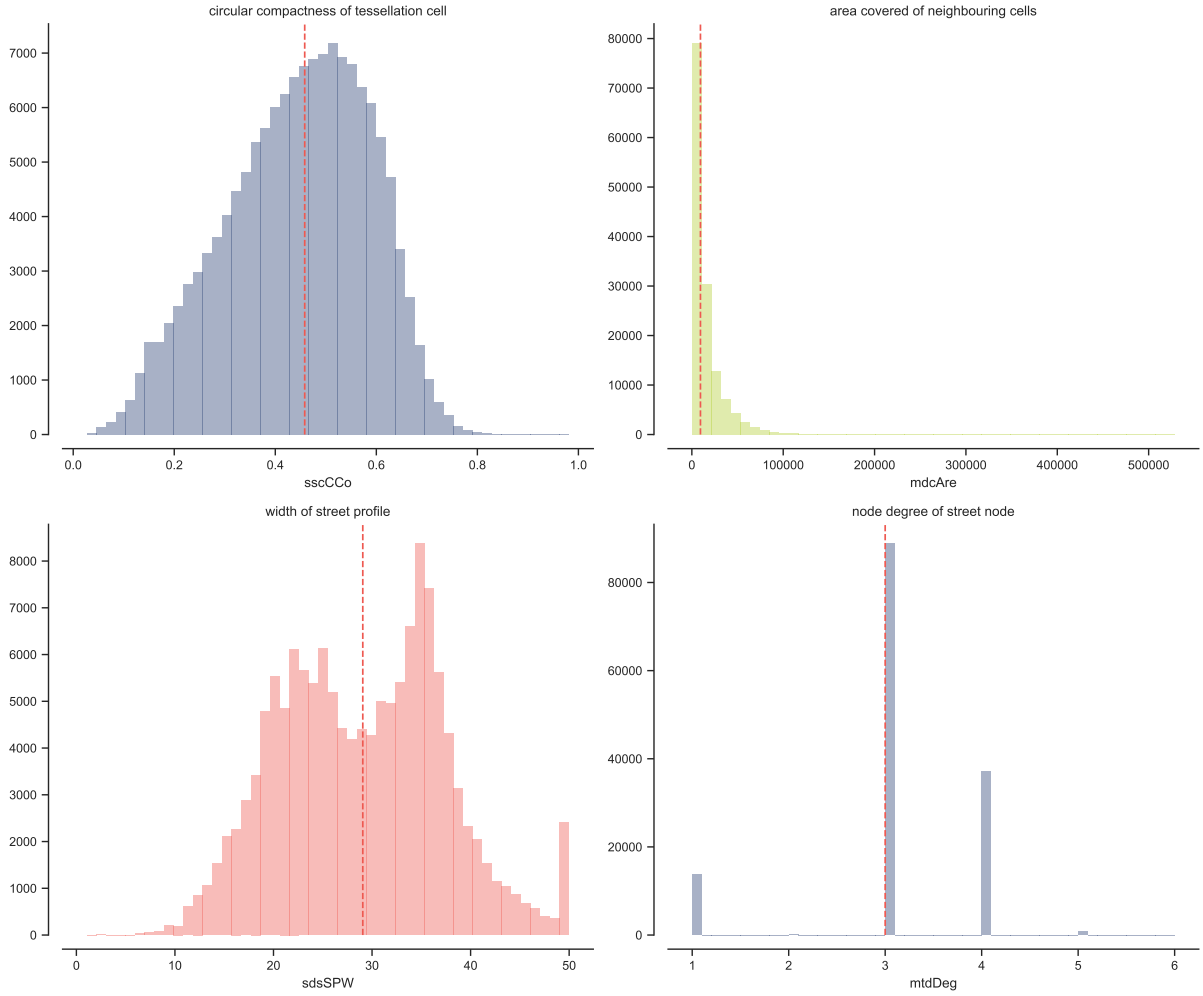


Figure 7.6: Histogram of four types of statistical distributions. Circular compactness of tessellation cell (top left), the area covered by neighbouring cells (top right), the width of a street profile (bottom left), and degree of a street node (bottom right).

The first case, *circular compactness of tessellation cell* (top left), is slightly skewed Gaus-

sian distribution with a minimum of values being in one of the extremes. It illustrates the range of characters with a more or less distorted normal distribution. Median value tends to be in the middle of the range.

The second case, *area covered by neighbouring cells* (top right), tends to follow already mentioned exponential distribution, with a majority of values being in the lowest extreme and only a few in the highest. Median value tends to be close to the overall minimum.

The third case, *width of a street profile* (bottom left), reflects the specific rule of the spatial organisation of cities. In this bimodal case, we can see peaks around 22 and 35 metres, which are likely predominant street widths in the context of Prague. The minor peak at 50 meters is caused by the maximum value of the defaulting to 50 metres in case of open spaces. Median value tends to be in the middle of the distribution, but that is not the overall rule for all characters of similar type of distribution.

The last case, *degree of a street node* (bottom right), is specific as the results are always integer values with a limited range. Cases like this are a minority. Apart from this example, only those measuring number of corners show similar behaviour.

These are not the only types of distributions in the set, but they illustrate the variability of morphometric characters.

Descriptive summary values of all character are presented in table 7.2.

Table 7.2: Overview of the primary morphometric values for the whole case study. The key to character IDs is available in table A7.3. Units, where applicable are in the section Identified set of primary characters

id	mean	std	min	25%	50%	75%	max
sdbAre	260	860	30	87	130	240	89000
sdbHei	9.9	6.7	3	5.5	7.4	12	110
sdbVol	3200	12000	90	550	960	3100	1.3e+06
sdbPer	64	56	20	40	51	67	3000
sdbCoA	2.1	64	0	0	0	0	11000
ssbFoF	1.4	0.57	0.23	1	1.3	1.6	11
ssbVFR	3	1.7	0.43	2.1	2.6	3.5	67
ssbCCo	0.53	0.11	0.026	0.47	0.56	0.61	1
ssbCor	8.8	7.4	0	4	8	10	390

id	mean	std	min	25%	50%	75%	max
ssbSqu	5.3	9.1	9.5e-09	0.48	1.1	5	85
ssbERI	0.94	0.086	0.25	0.91	0.96	1	1.1
ssbElo	0.71	0.2	0.026	0.56	0.74	0.87	1
ssbCCD	1.5	2.2	0	0.068	1	1.9	88
ssbCCM	9.4	6.6	3	6.3	7.6	10	210
stbOri	16	13	0	6.2	13	25	45
stbSAl	6.7	8.9	4.9e-10	0.61	2.5	9.5	45
stbCeA	6.9	9	8.9e-12	0.48	3	9.9	45
sdcLAL	67	42	7.9	40	52	79	970
sdcAre	2100	4100	31	540	940	1900	350000
sscCCo	0.45	0.14	0.027	0.35	0.46	0.55	0.98
sscERI	0.97	0.062	0.43	0.94	0.98	1	1.1
stcOri	18	13	0	7.1	16	29	45
stcSAl	9.2	9.7	1.9e-05	1.5	5.6	14	45
sicCAR	0.2	0.15	0.00092	0.092	0.16	0.26	1
sicFAR	0.67	0.92	0.00092	0.14	0.32	0.74	17
sdsLen	230	260	0.047	110	160	260	3300
sdsSPW	29	8.4	1	22	29	35	50
sdsSPH	10	6.1	0	6.4	8	13	57
sdsSPR	0.41	0.32	0	0.21	0.3	0.49	23
sdsSPO	0.58	0.21	0	0.44	0.58	0.71	1
sdsSWD	3.6	2.1	0	1.9	3.7	5.1	12
sdsSHD	2.3	2.3	0	0.94	1.5	2.7	24
sssLin	0.95	0.13	0	0.97	1	1	1
sdsAre	31000	56000	34	6900	13000	30000	740000
sisBpM	0.075	0.079	0.00056	0.046	0.068	0.095	21
sddAre	30000	46000	86	9400	16000	31000	660000
mtbSWR	0.18	0.2	0	0	0.15	0.32	1
mtbAli	4.8	5.1	1.4e-09	0.9	3	7	44
mtbNDi	25	18	0	13	20	30	200
mtcWNe	0.046	0.022	0.0012	0.03	0.045	0.059	0.26
mdcAre	16000	19000	390	5500	9400	19000	530000

id	mean	std	min	25%	50%	75%	max
misRea	44	25	1	27	40	55	290
mdsAre	86000	110000	770	33000	53000	94000	1.3e+06
mtdDeg	3.1	0.82	1	3	3	4	6
mtdMDi	170	150	0.047	99	130	190	3300
midRea	52	28	1	33	49	67	270
midAre	97000	110000	770	42000	65000	110000	1.3e+06
libNCo	0.6	3.3	0	0	0	0	58
ldbPWL	180	250	20	51	82	200	3400
ltbIBD	27	11	0	20	25	33	120
ltcBuA	0.65	0.24	0.043	0.49	0.7	0.84	1
licGDe	0.57	0.67	0.0022	0.18	0.35	0.66	5
ltcWRB	9e-05	6.7e-05	1.7e-06	3.9e-05	7.3e-05	0.00012	0.00072
ldkAre	120000	240000	710	15000	31000	110000	2e+06
ldkPer	1500	1800	100	550	830	1700	13000
lskCCo	0.43	0.13	0.11	0.33	0.44	0.53	0.98
lskERI	0.86	0.13	0.35	0.79	0.9	0.96	1.1
lskCWA	360	470	0.43	87	170	430	3100
ltkOri	18	13	0.00098	7	15	28	45
ltkWNB	0.0074	0.0043	0	0.004	0.0066	0.01	0.04
likWBB	0.00089	0.00066	8.3e-06	0.00037	0.00074	0.0013	0.006
lcdMes	0.15	0.06	-0.33	0.11	0.15	0.19	0.34
ldsMSL	150	76	45	110	130	170	1600
ldsCDL	280	390	0	13	160	380	4200
ldsRea	350000	310000	770	190000	260000	400000	4.2e+06
lddNDe	0.013	0.0055	0	0.0095	0.012	0.014	0.13
lddRea	190	86	1	130	190	240	680
lddARe	370000	310000	770	200000	280000	420000	4.2e+06
linPDE	0.13	0.087	0	0.067	0.11	0.17	1
linP3W	0.64	0.11	0	0.57	0.64	0.71	0.97
linP4W	0.23	0.12	0	0.15	0.22	0.3	0.73
linWID	0.025	0.01	0	0.019	0.024	0.029	0.18
lcnClo	5.3e-06	2.5e-06	0	3.4e-06	5.1e-06	6.9e-06	2e-05

id	mean	std	min	25%	50%	75%	max
xcnSCl	0.056	0.087	0	0	0	0.086	1

Without exploring the table 7.2 above in detail, it is worth pointing out two characters standing out - *courtyard area of a building (sdbCoA)* and *number of courtyards of adjacent buildings (libNCo)*. Both are capturing similar concepts of closed courtyards (either in a single building or in a composite of adjacent buildings), and both are relatively invariant (min, 25%, 50% and 75% are all 0). While these might not be critical for identification of clusters in Prague, there are urban tissues, especially in warmer environments, characterised by these properties. While the overall aim of this research is to be comparable, not tailored to a specific context, these characters are still included.

Figures 7.7 - 7.11 show histograms capturing the (truncated) distribution of all measured characters. Note the differences outlined above and overall variety of distributions.

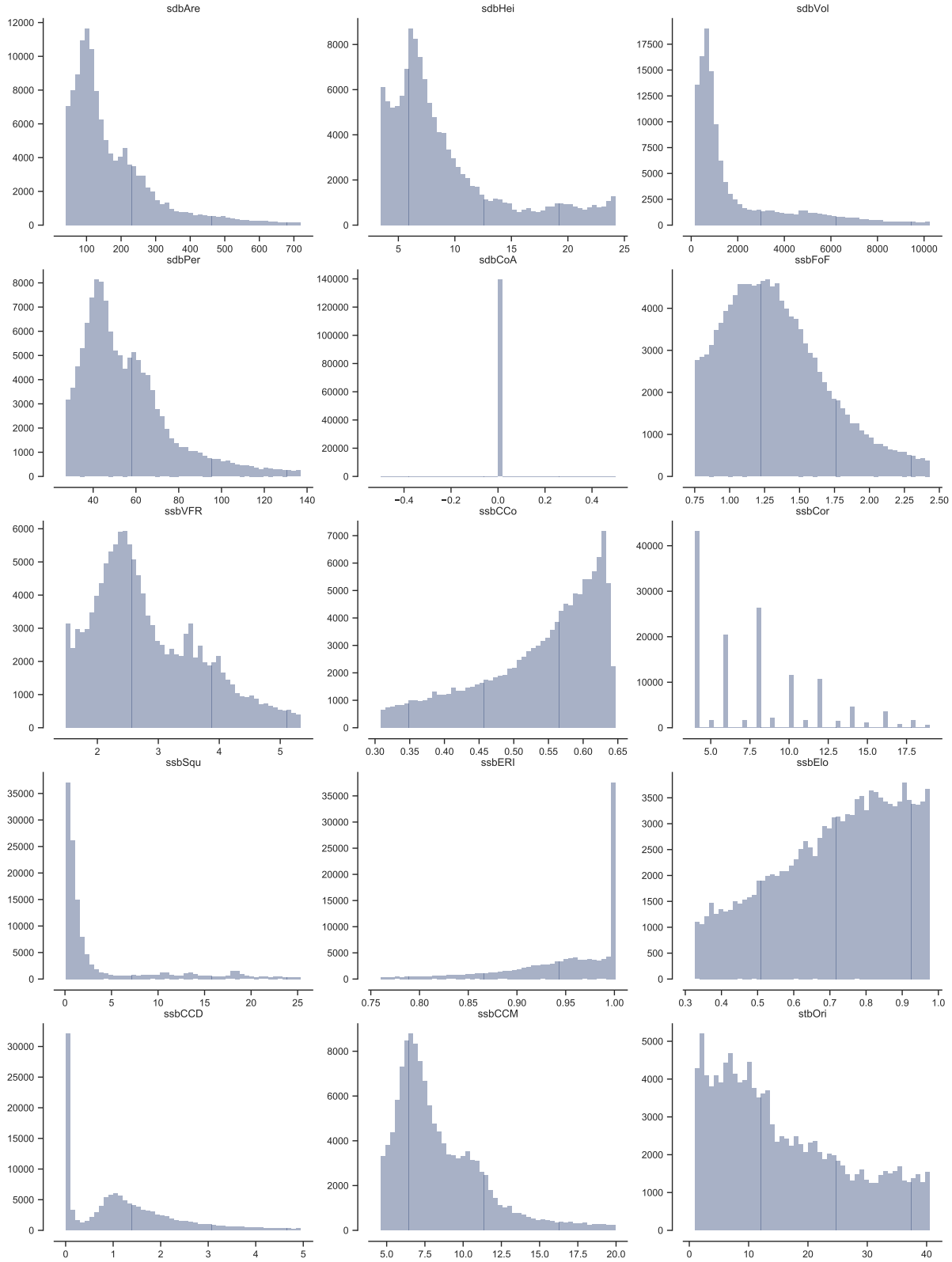


Figure 7.7: Histograms of characters 1-15 are showing the variety of distributions within the measured primary data. Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table are presented complete for reference.

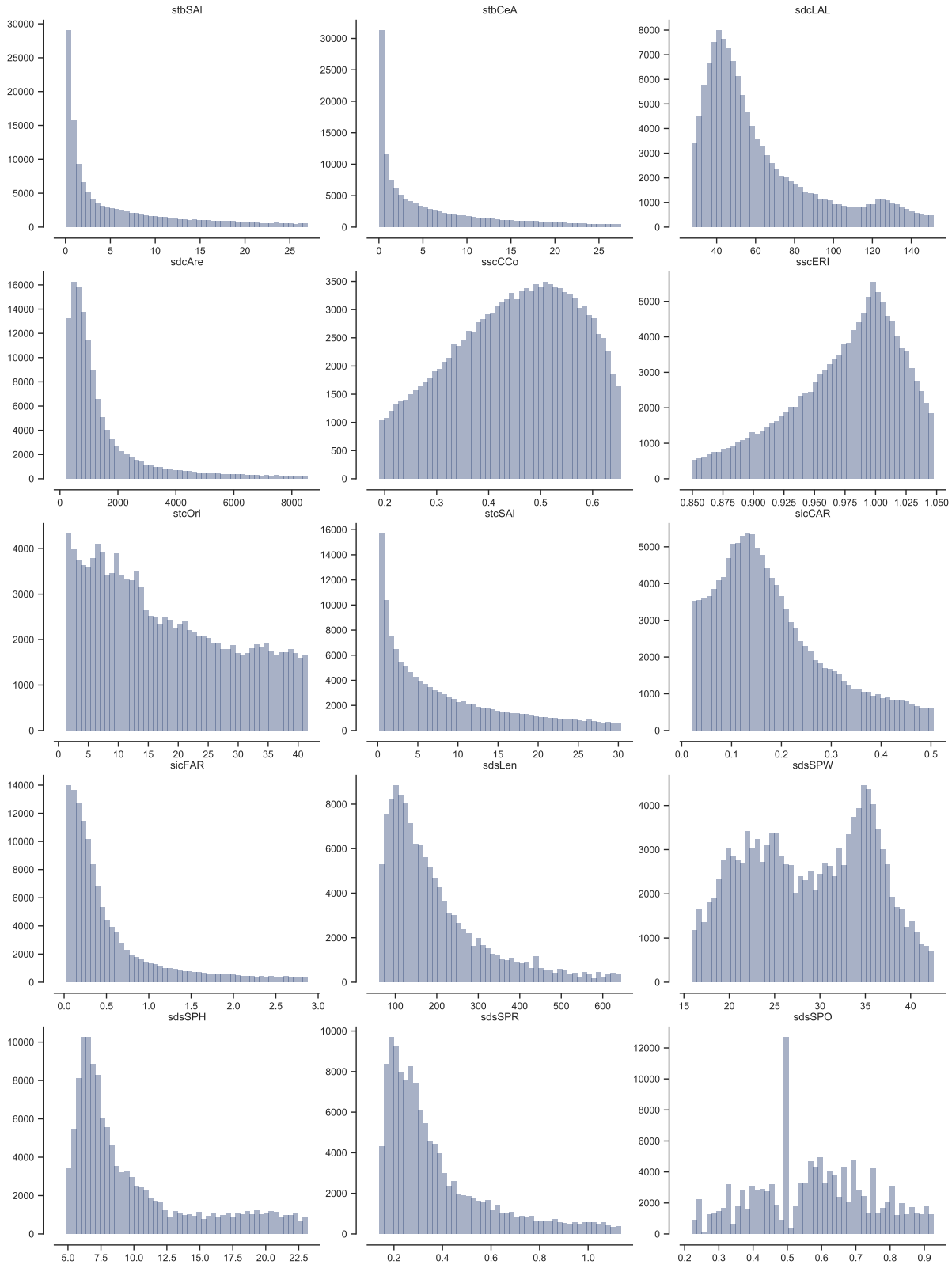


Figure 7.8: Histograms of characters 16-30 are showing the variety of distributions within the measured primary data. Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table are presented complete for reference.

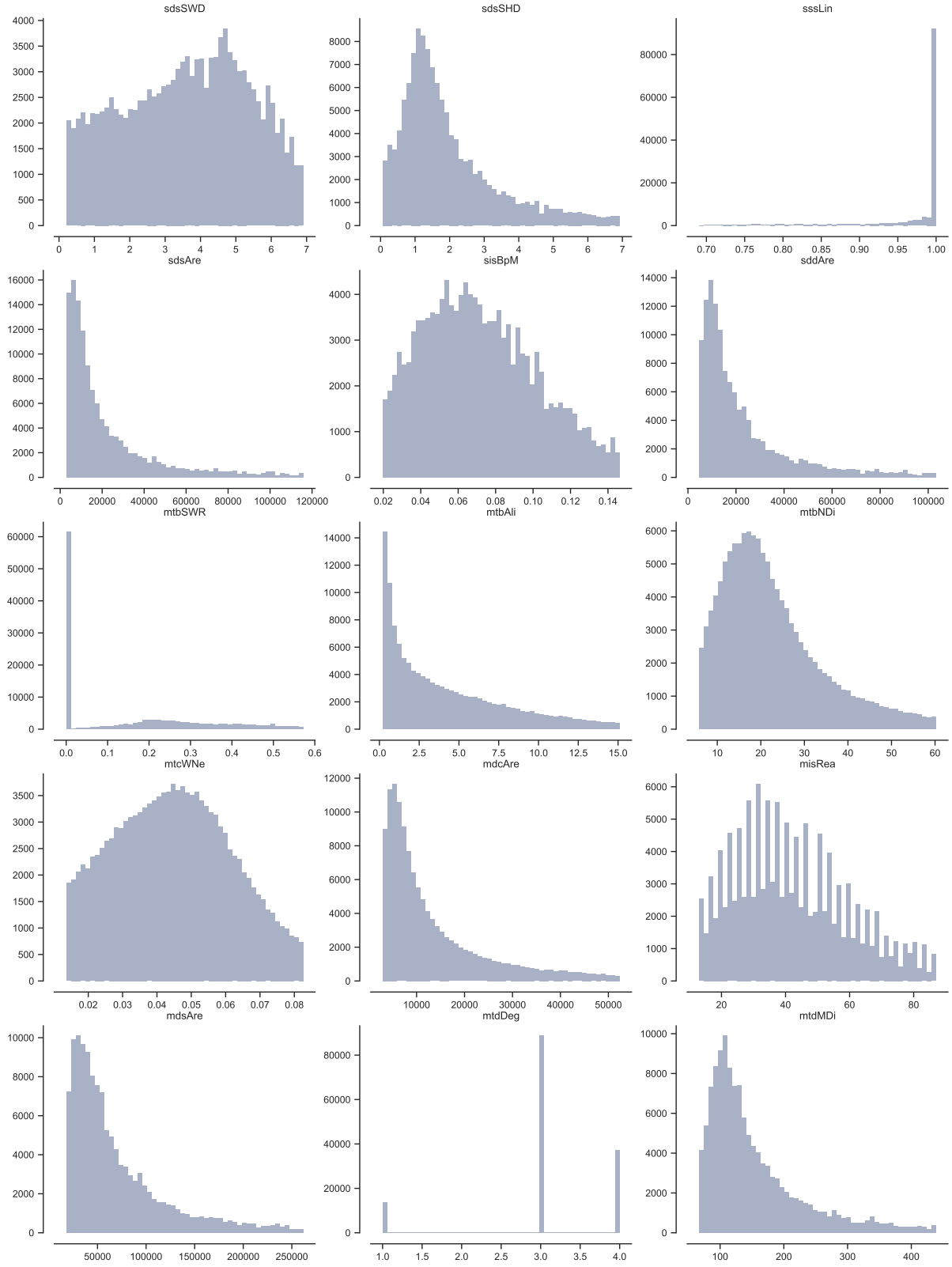


Figure 7.9: Histograms of characters 31-45 are showing the variety of distributions within the measured primary data. Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table are presented complete for reference.

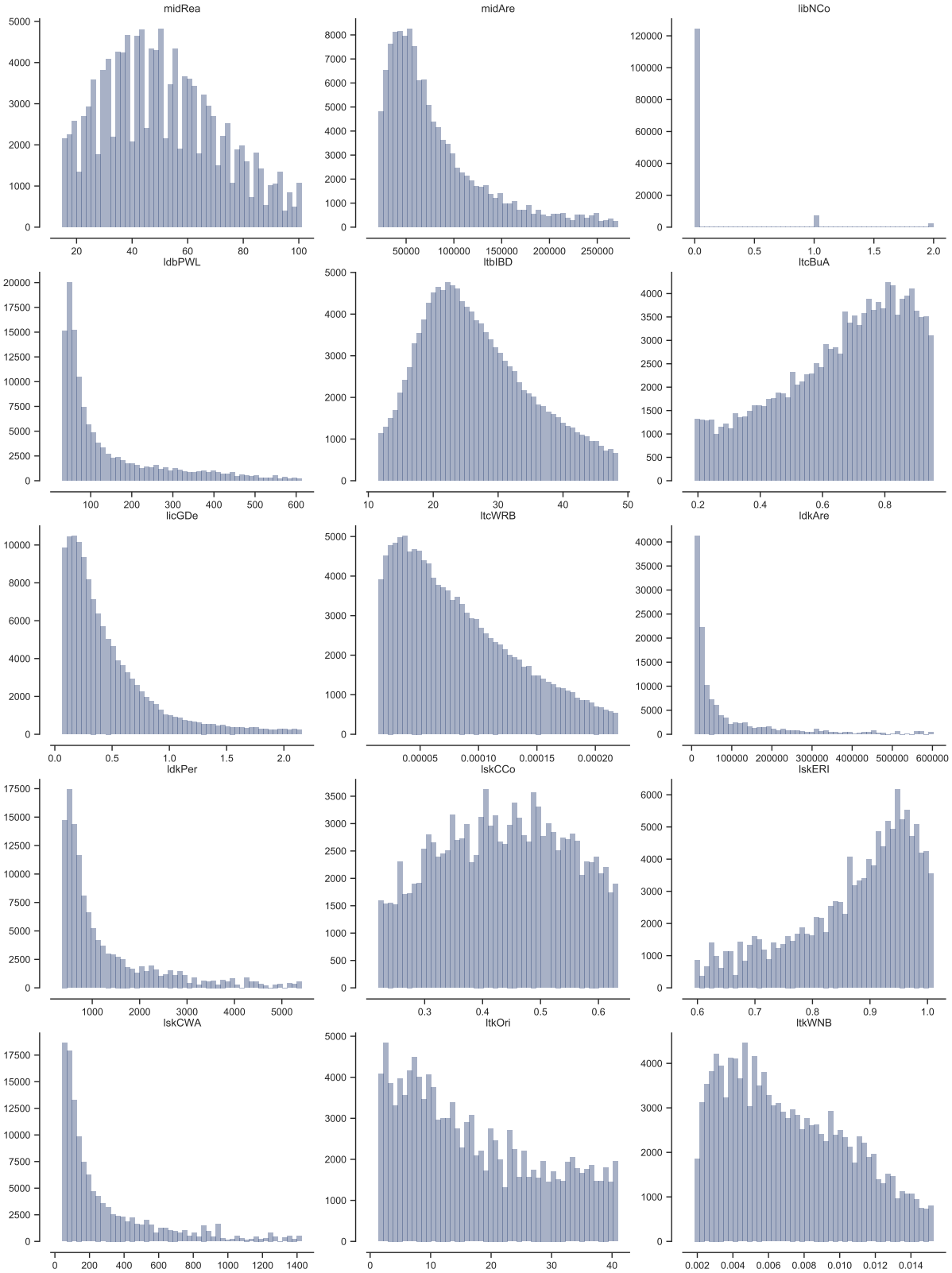


Figure 7.10: Histograms of characters 45-60 are showing the variety of distributions within the measured primary data. Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table are presented complete for reference.

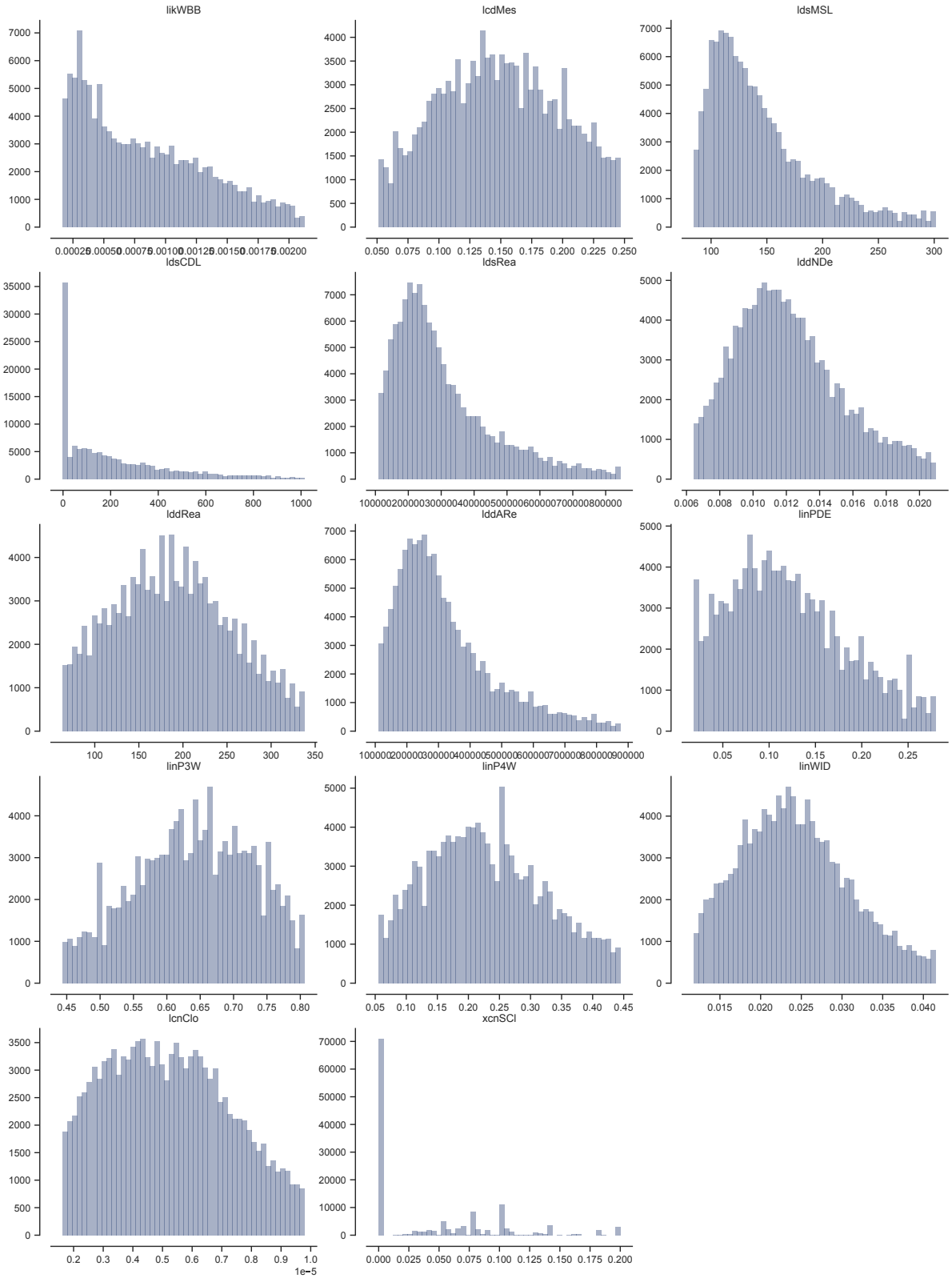


Figure 7.11: Histograms of characters 61-74 are showing the variety of distributions within the measured primary data. Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table are presented complete for reference.

7.2.1.3 Statistical relationship of characters

Understanding the relationship between measured characters is an essential aspect of the morphometric assessment. As specified in section 7.1.2.1.1, characters should not include many empirical correlations. Collinear characters (those being correlated and reflecting the same concept) should not be present in the resulting data as they might skew the hyperspace and adversely affect the result of clustering. As characters do not follow always follow a normal distribution, Spearman's rank correlation (Spearman, 1961) is used to assess the relationship between characters. The results are illustrated in the correlation matrix on figure 7.12 below.

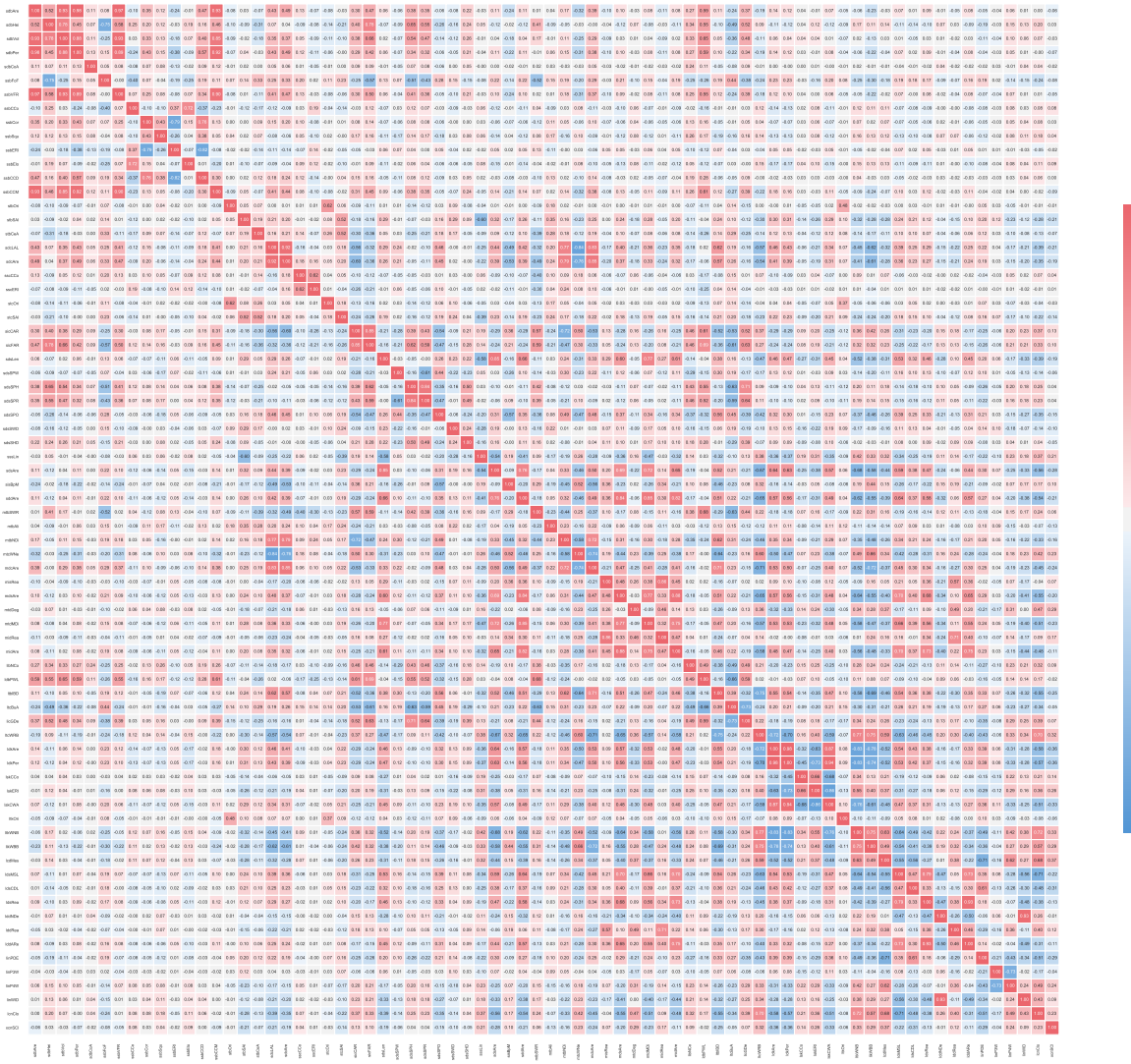


Figure 7.12: Correlation matrix of Spearman's rho values capturing the statistical relationship between resulting morphometric values of primary characters. With a few exceptions, the relationship is none or very weak.

As expected (the set is designed in such a way), characters generally show minimal correlations, with a few exceptions. These are reflecting different concepts and are capturing different phenomena, which makes them admissible.

Primary morphometric characters are the core of the method of identification of tissue types. Their selection, to capture different complexities aspects of urban form, results in a very heterogeneous set of measured data showing variable spatial patterns as well as statistical distributions. However, this data are the direct input of the clustering

procedure, but they are an input of calculation of contextual characters. The results of this following step are illustrated in the next section.

7.2.2 Contextual characters

While the importance of primary morphometric characters is that they bring the fundamental information about the spatial order of elements of urban form, the values which are used for the identification of clusters itself are based on the contextualisation. Resulting contextual characters are of four types (interquartile mean, interquartile range, interdecile Theil index, Simpson diversity index), where each describes the same primary character from a different perspective. Together, they reflect the context of each tessellation cell, defined as three topological steps, comprehensively and inclusively.

The actual values measured in Prague could, once again, be explored visually to assess the spatial distribution of resulting values and numerically to assess resulting statistical distributions.

7.2.2.1 Spatial distribution

Unlike in the case of primary characters, contextual characters are always capturing spatially consistent patterns. The reason is the inclusion of the topological context in each of them. However, the actual distribution of values differ. Following four figures show contextual characters based on *width of a street profile* to illustrate the differences and similarities between contextual characters. Note that this is an only illustrative example, and spatial distribution would differ for other characters.

Figure 7.13: Spatial distribution of the interquartile mean of a width of a street profile measured within three topological steps on morphological tessellation in the area of Prague's city centre and its surroundings.

Figure 7.13 shows *interquartile mean of a width of a street profile* measured within three topological steps on morphological tessellation. As a version of truncated mean, this character directly reflects the actual values of primary characters, and it is relatively simple to indicate areas with generally narrow streets (historical core) and those with a

wider profile (heterogenous areas on south and south-east). The overall distribution of values within the shown area is very symmetrical with a peak at 22 metres, which seems to be a common street width in Prague.

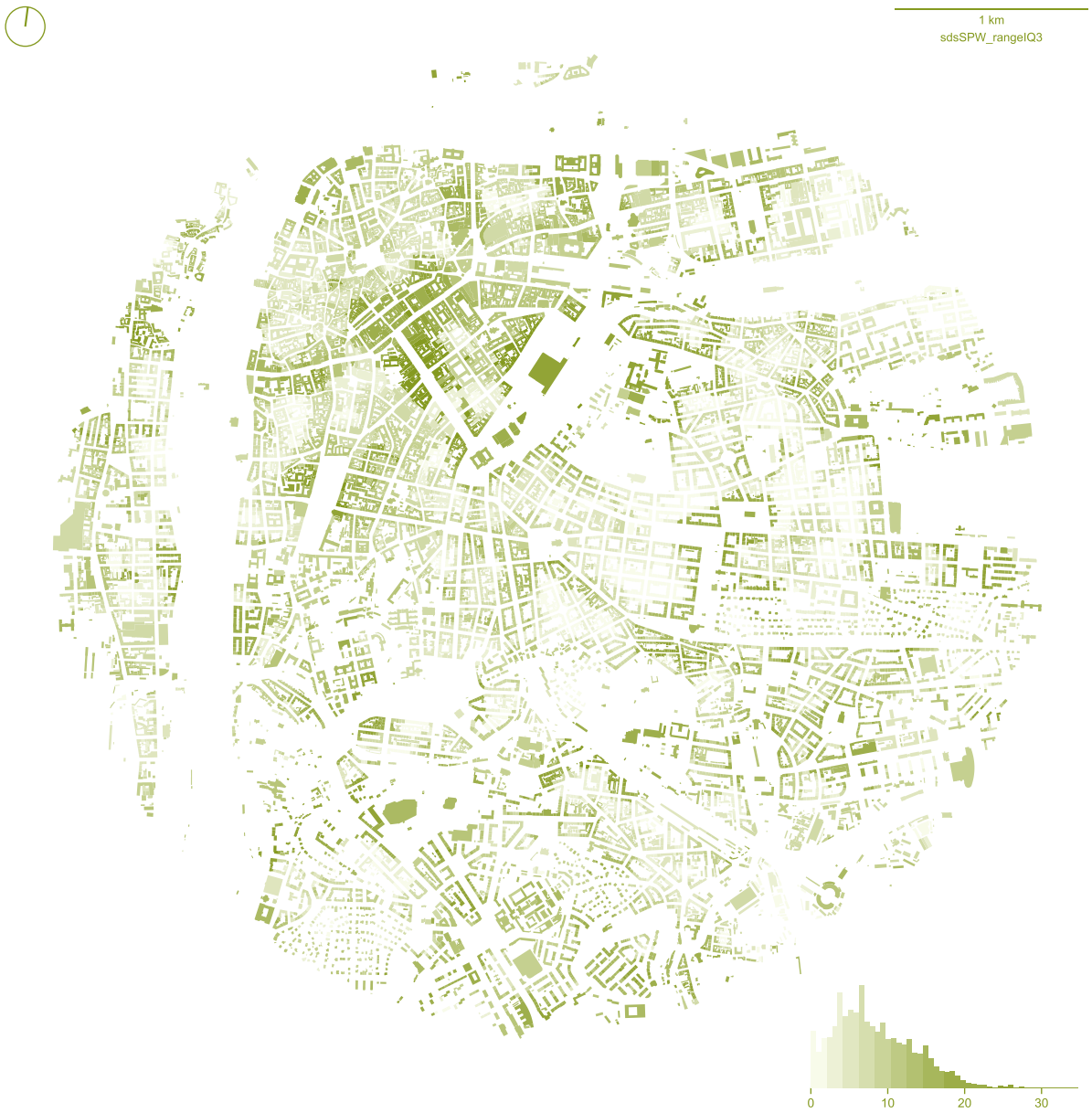


Figure 7.14: Spatial distribution of the interquartile range of a width of a street profile measured within three topological steps on morphological tessellation in the area of Prague's city centre and its surroundings.

Figure 7.14 illustrates *interquartile range of a width of a street profile* measured within

three topological steps on morphological tessellation. That reflects the range of values, so it is a proxy of statistical dispersion. In the example above it does divide places with either major street or generally wider streets from predominantly homogenous areas. The distribution of values is balanced but truncated at 0 (range could not be negative). We can identify certain similarity with the patterns on previous figure 7.13, because wider street (i.e., higher interquartile mean) causes a bigger range, but the patterns are not identical.

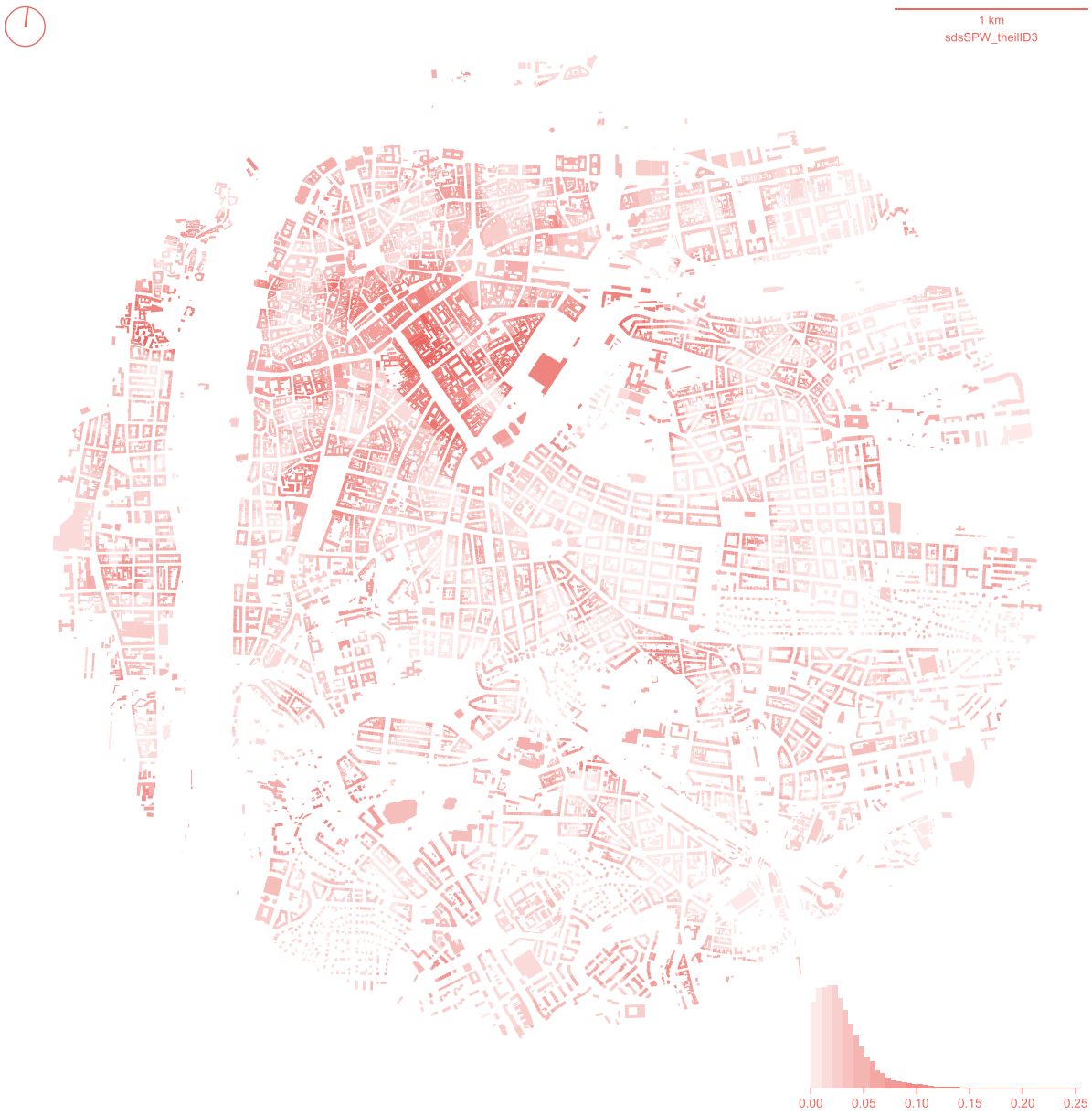


Figure 7.15: Spatial distribution of interdecile Theil index of a width of a street profile measured within three topological steps on morphological tessellation in the area of Prague's city centre and its surroundings.

Figure 7.15 illustrates *interdecile Theil index of a width of a street profile* measured within three topological steps on morphological tessellation. The resulting map shows as the most *unequal* area around Vaclavske sq. (darker red) where one street (in this case, elongated square) is significantly different from the other. Previously highlighted areas of wider

streets are not so from the perspective of Theil index. The distribution has a long tail, somewhat typical for Theil index applied to morphometric characters.

Figure 7.16: Spatial distribution of interdecile Simpson index of a width of a street profile measured within three topological steps on morphological tessellation in the area of Prague's city centre and its surroundings.

The last contextual character, shown on figure 7.16, is *Simpson diversity index of a width of a street profile* measured within three topological steps on morphological tessellation.

The values, in this case, are binned using Natural Breaks (k=7). It captures (inversely) similar information as Theil index, but that is not the rule. These two are somewhat related as both capture dispersion of values, but the relationship between them is not fixed as will be illustrated in the next section.

Depending on the spatial distribution of primary characters, the contextual pattern may be more similar to each other (like in the case above) or less similar. However, as illustrated in the chapter on primary characters, the visual assessment is not enough.

7.2.2.2 Statistical distribution

Figure 7.6 in previous section showed four types of distribution of primary characters. This section illustrates how each of them translates into the distribution of contextual characters.

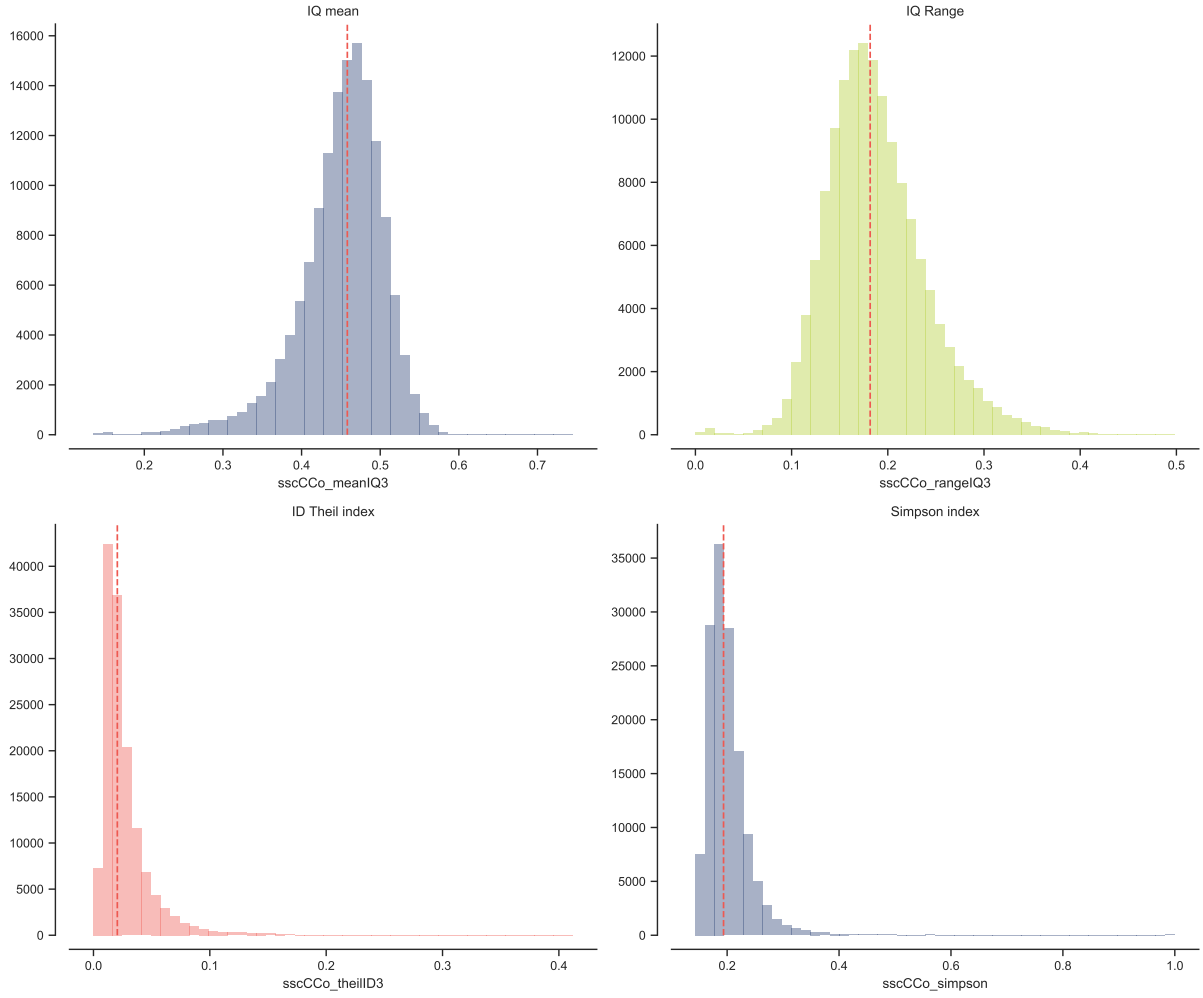


Figure 7.17: Histograms of statistical distribution of contextual versions of circular compactness of tessellation cell. Interquartile mean (top left), interquartile range (top right) interdecile Theil index (bottom left), Simpson index (bottom right).

The first example, *circular compactness of tessellation cell* on figure 7.17, was initially mildly skewed Gaussian-like distribution. In terms of IQ mean and IQ range, this property remains the same. Both are relatively symmetrical distributions with small tail on one or the other side. On the other hand, the distribution of the Theil index and Simpson diversity resembles exponential curve due to heavy tail in both.

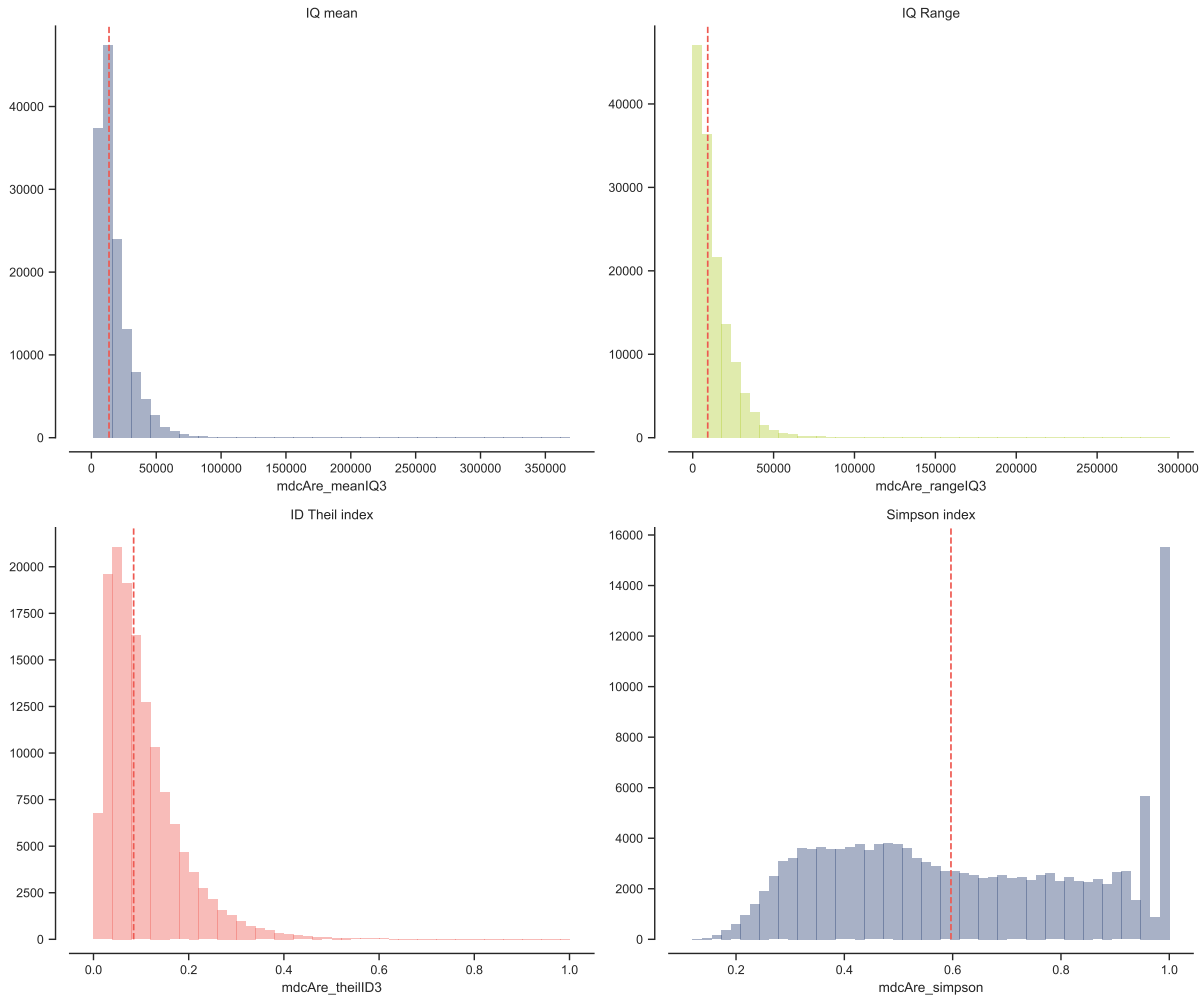


Figure 7.18: Histograms of statistical distribution of contextual versions of area covered by neighbouring cells. Interquartile mean (top left), interquartile range (top right) interdecile Theil index (bottom left), Simpson index (bottom right).

Initially, exponential distribution of *area covered by neighbouring cells* remains exponential in both IQ mean and IQ range cases (figure 7.18). Theil index is also exponential, although the curve is not so unequal. Simpson index is significantly different from all three. The HeadTail binning used within the calculation is tailored to exponential distributions and resulting Simpson diversity is then relatively balanced across the values. The values 1 showing a significant spike mean that the probability that any other value is within the same bin is 100%, hence no diversity in the area. It is typical for relatively homogenous compact urban tissues.

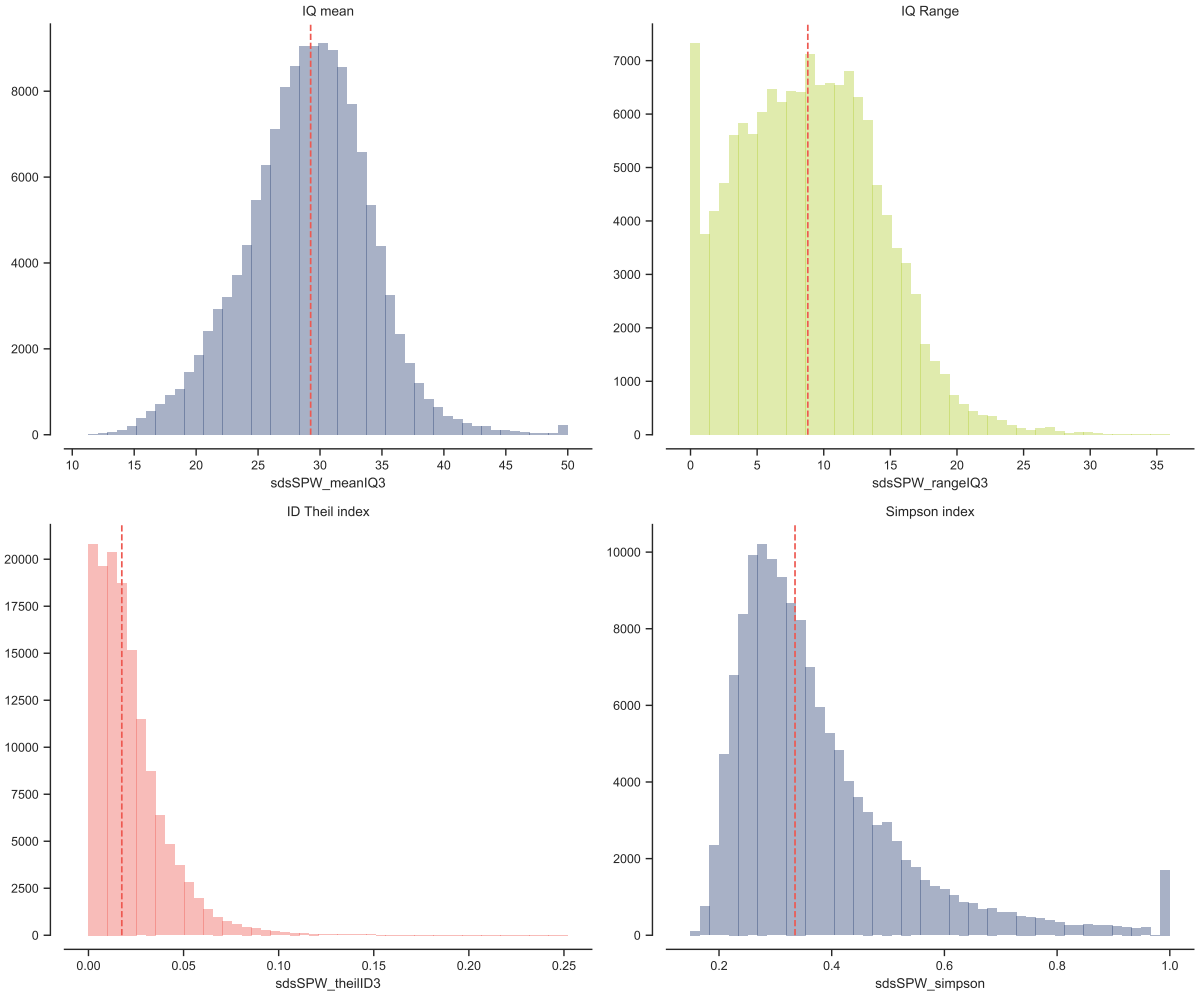


Figure 7.19: Histograms of statistical distribution of contextual versions of width of a street profile. Interquartile mean (top left), interquartile range (top right) interdecile Theil index (bottom left), Simpson index (bottom right).

The third example, *width of a street profile* which was illustrated on maps on previous pages, had initially specific distribution affected by rules on which streets are designed (there were spikes for narrower and wider streets). Figure 7.19 shows that none of the contextual character share this profile and, more importantly, all have different distributions. IQ mean is almost ideal Gaussian distribution, IQ range is right-skewed and truncated with a spike on 0, Theil index is again exponential, and Simpson diversity is right-skewed, but relatively symmetrical distribution. Even though figures 7.13 - 7.16 may seem similar, the difference in distributions on figure 7.19 indicates otherwise. What is important are

numerical values, not visual perception.

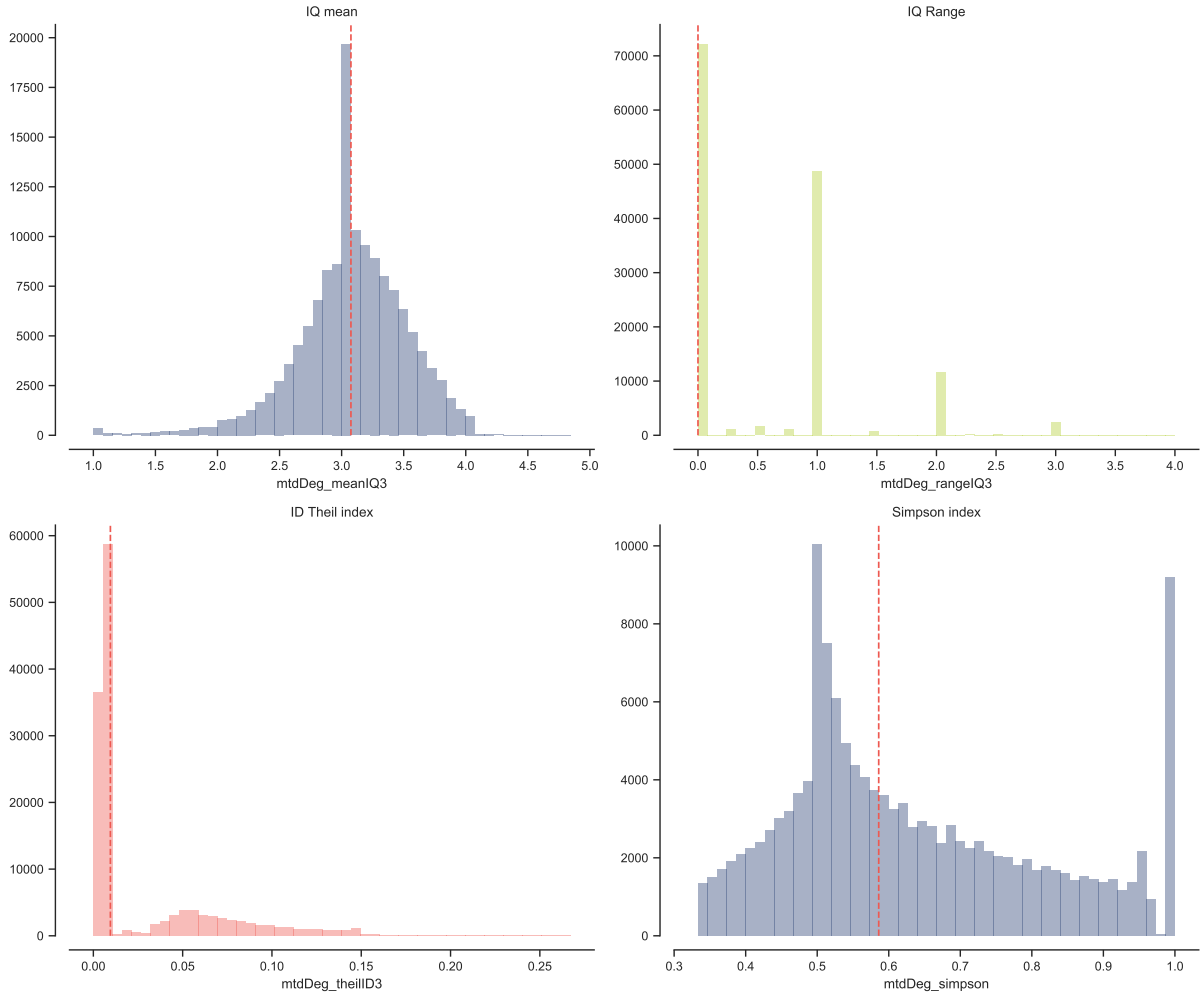


Figure 7.20: Histograms of statistical distribution of contextual versions of degree of a street node. Interquartile mean (top left), interquartile range (top right) interdecile Theil index (bottom left), Simpson index (bottom right).

Initially, restricted values of *degree of a street node* remained present in IQ range (figure 7.20 but not in the other contextual characters. IQ mean is symmetrical with a large spike on 3, which is almost its median value. Theil index is very different from previous examples and does not follow exponential distribution this time, while the Simpson diversity index has two spikes on 0.5 and 1.0 and relative balance otherwise.

As the examples above indicate, the variety present in primary characters remained present, in a different way, in contextual characters as well. Complete results for all

contextual characters are available as Appendix 7.7.

7.2.2.3 Statistical relationship of characters

The statistical relationship between contextual characters will directly influence the results of clustering in the next steps. For that reason, we should aim for minimisation of such relationship in terms of Spearman's correlation. As illustrated above, we may expect some relations, however only selective, affected by the nature of primary characters. Below (figure 7.21) is a correlation matrix of contextual characters for illustration of the measured relationship[8].

[8]Due to a large number of characters, the matrix is not optimal for presentation in this form. Its high-quality version is available in Appendix 7.8.

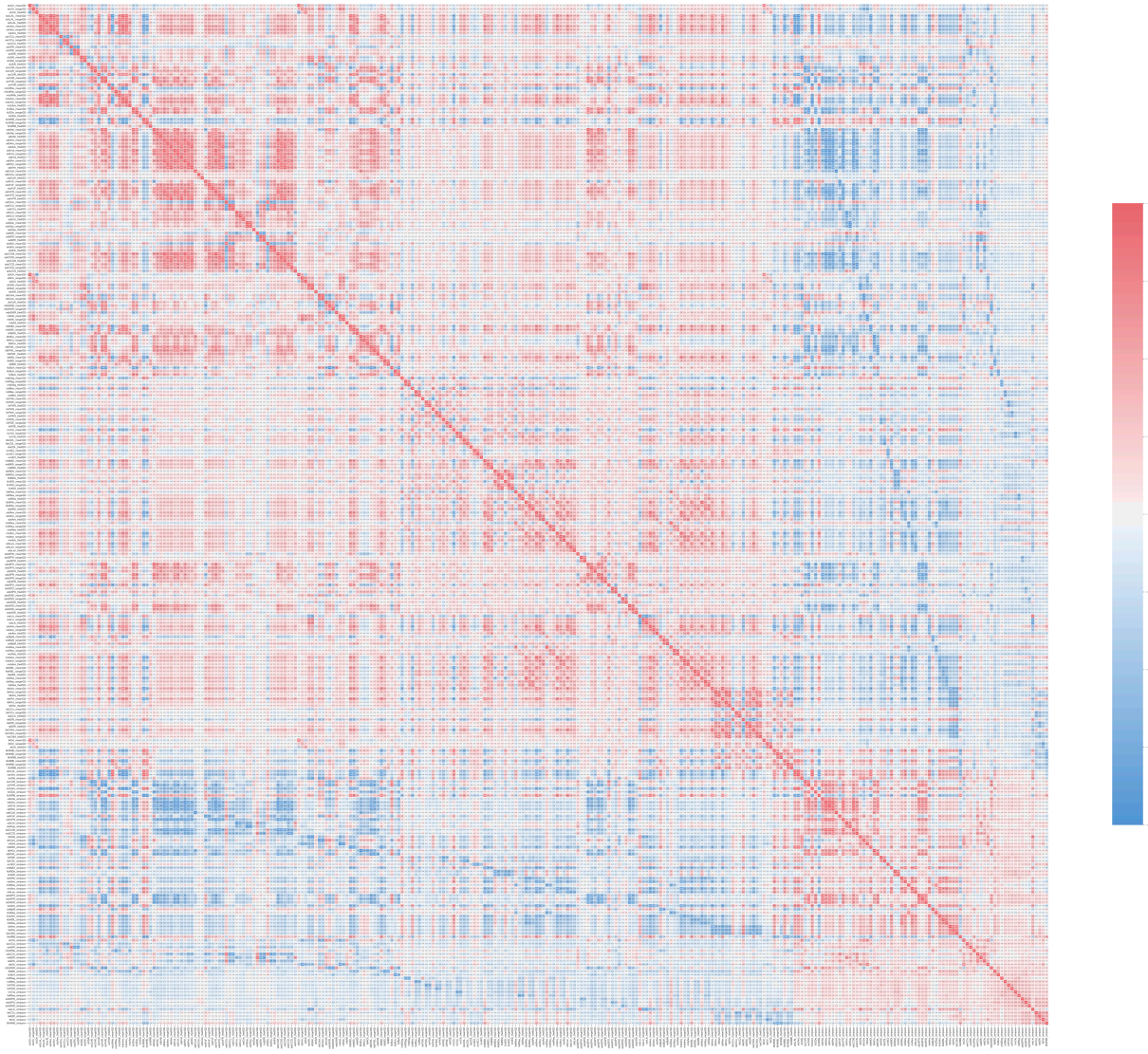


Figure 7.21: Correlation matrix of contextual capturing the statistical relationship between resulting morphometric values of contextual characters. With a few exceptions, the relationship is none or very weak. High-quality version of the matrix is in Appendix 7.8.

Even though there, once again, are some characters which tend to be correlated, the overall correlation is minimal. Such data then have the potential to provide the meaningful, unskewed result of clustering.

The exploration of measured primary and consequent contextual character shows that the data comply with requirements set by the method and show a high variety of information. It is assumed, that they describe urban form in its structural complexity as well as cross-

scale complexity.

7.2.3 Cluster analysis

The critical point in the whole process if the identification of distinct clusters is the clustering procedure itself. This section will explore the results of clustering on the complete set of data and then on sampled data, to understand the difference and possibility to lower computational demands. Both ways will start with an assessment of an optimal number of components based on the Bayesian Information Criterion (BIC) and follow with Gaussian Mixture Model (GMM) clustering itself. Furthermore, the final part of this section explores the potential of sub-clustering, i.e., generating even more detailed distinction of urban tissues.

7.2.3.1 Complete data

Clustering based on complete data is likely the key result of the whole research. It will either support the main hypothesis or reject it depending on the resulting clusters. GMM clustering of a complete dataset means that all features (n=140315) are used within a training set. It is expected that the algorithm will be able to detect clusters, although there might be present some adverse effects of the dimensionality curse[9].

Before analysis itself, data are standardised by mean removal and variance scaling:

$$(82)\quad z = \frac{x-\mu}{s}$$

[9]The morphometric description of each building/cell has 296 values (each for each contextual character). In the case of Prague, composed of approximately 140,000 buildings, it means that clustering has to deal with more than 40,000,000 values (140,000 buildings * 296 characters). That is a significant number, which is not only demanding in terms of computational power but also tricky in terms of statistics itself. The high dimensionality of the dataset (each character is a dimension in a hyperspace) may come with a *curse of dimensionality*. That means that even though there is the value in additional data (additional dimensions), it may negatively affect results. The high-dimensional hyperspace tends to become inflated (bigger), which in turn may render clusters very sparse. Individual data points are further away, and density-based or distance-based clusterings (GMM is distance-based) may struggle to correctly identify them, because Euclidean distances between pairs of points on sparse high-dimensional data would be of little difference, rendering clustering extremely unstable and insignificant. However, that is not always the case as it depends on the internal structure of the dataset and relations between dimensions.

where μ is the mean of the training values, and s is the standard deviation.

7.2.3.1.1 Bayesian Information Criterion To perform GMM clustering, one needs to specify the number of components to look for. While this information is not a priori known, one has to determine the optimal number using other methods before GMM. In this research, the Bayesian Information Criterion is used. BIC analysis repeatedly generates GMM clusters for a different number of components in a range (2, 40) and measures the goodness of fit of resulting clusters to the original dataset. The results in Prague are shown in figure 7.22. The lower the value, the better fit.

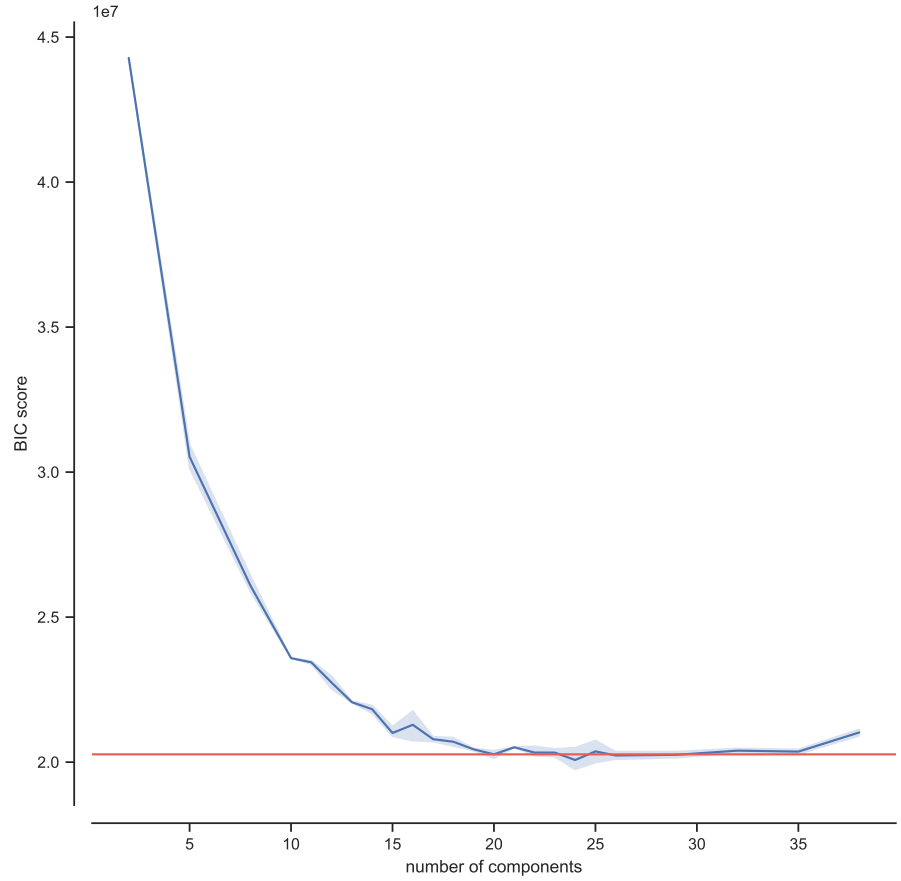


Figure 7.22: Bayesian Information Criterion score for the variable number of components. Shaded area reflects .95 confidence interval, red line marks the first significant minimum.

The pattern shows a steep decline from a small number of components to approximately 15 components where it starts flattening. The results between 15 and 35 components

are very similar, and then the BIC starts growing again. That suggests that the optimal number of components for the final clustering is between these two values. The optimum is 20 as the value, which is the first significant minimum. It is the smallest number after which no other is significantly (with the confidence interval) below the achieved score. The differentiation within the range in question is better recognisable in a zoomed figure 7.23 below.

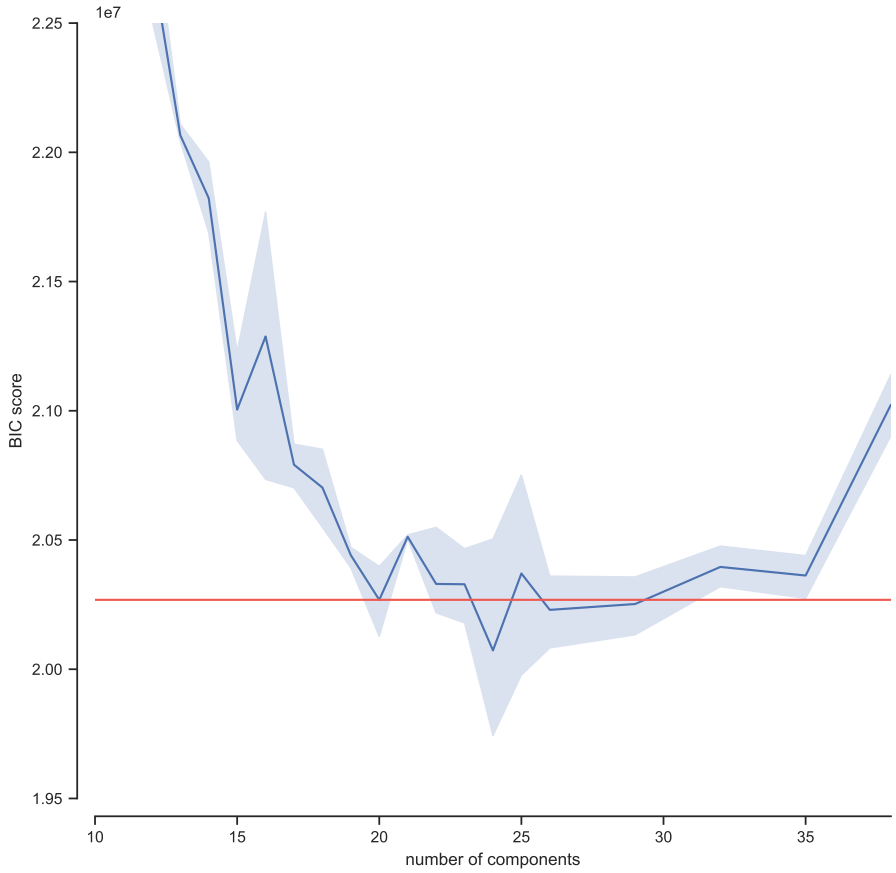


Figure 7.23: Trimmed plot of Bayesian Information Criterion score for the variable number of components, to see the differentiation within values. Shaded area reflects .95 confidence interval, red line marks the first significant minimum.

Within the trimmed figure, it is more evident the difference between the BIC score. The reason why 24 or 26 are not selected as optimum, while both being smaller than 20 is the significance. The aim is to detect the smallest optimal number of clusters as larger numbers may have better due to overfitting. Hence we want the first significant minimum, which is 20. 24 is below, but its confidence interval goes above the score of 20. The same

applies to 26.

Based on the BIC results, GMM clustering of complete data will focus on the identification of 20 components (clusters).

7.2.3.1.2 Distinct clusters as urban tissue types Gaussian Mixture Model clustering with 20 components is done using full covariance matrix and five initialisations, from which the best is selected based on the per-sample average log-likelihood. The complete Jupyter notebook is available in Appendix N.

The resulting prediction of cluster membership is shown visually on the figure 7.24. Each feature (building/tessellation cell) is coloured according to a cluster of the highest probability[10]. The map shows the delineation of distinct homogenous clusters and their spatial distribution across the whole case study area. At this moment, it is possible to say that the proposed method did identify the particular type of proxy of urban tissues using the purely quantitative method based on urban morphometrics. How well it did that will be assessed on the following pages and later validated by other data in Chapter 8.

[10]As mentioned above, Gaussian Mixture Model clustering is probabilistic, which means that each feature has predicted the probability that it belongs to any of the components. What is shown on all maps and data below is the cluster with the highest probability. In theory, it should be possible to work with secondary or tertiary labels for each feature, but the actual data on probability tell otherwise. The probability that features belong to any other than the primary cluster tends to be insignificant. Only 89 out of 140,315 features have the probability that they belong to any other than primary cluster bigger than 0.1. The reason behind it is likely related to the richness of the data and especially related dimensionality causing big differences in Euclidean distance between clusters. So while GMM is, in theory, probabilistic, in practice, it provides a single primary label only.

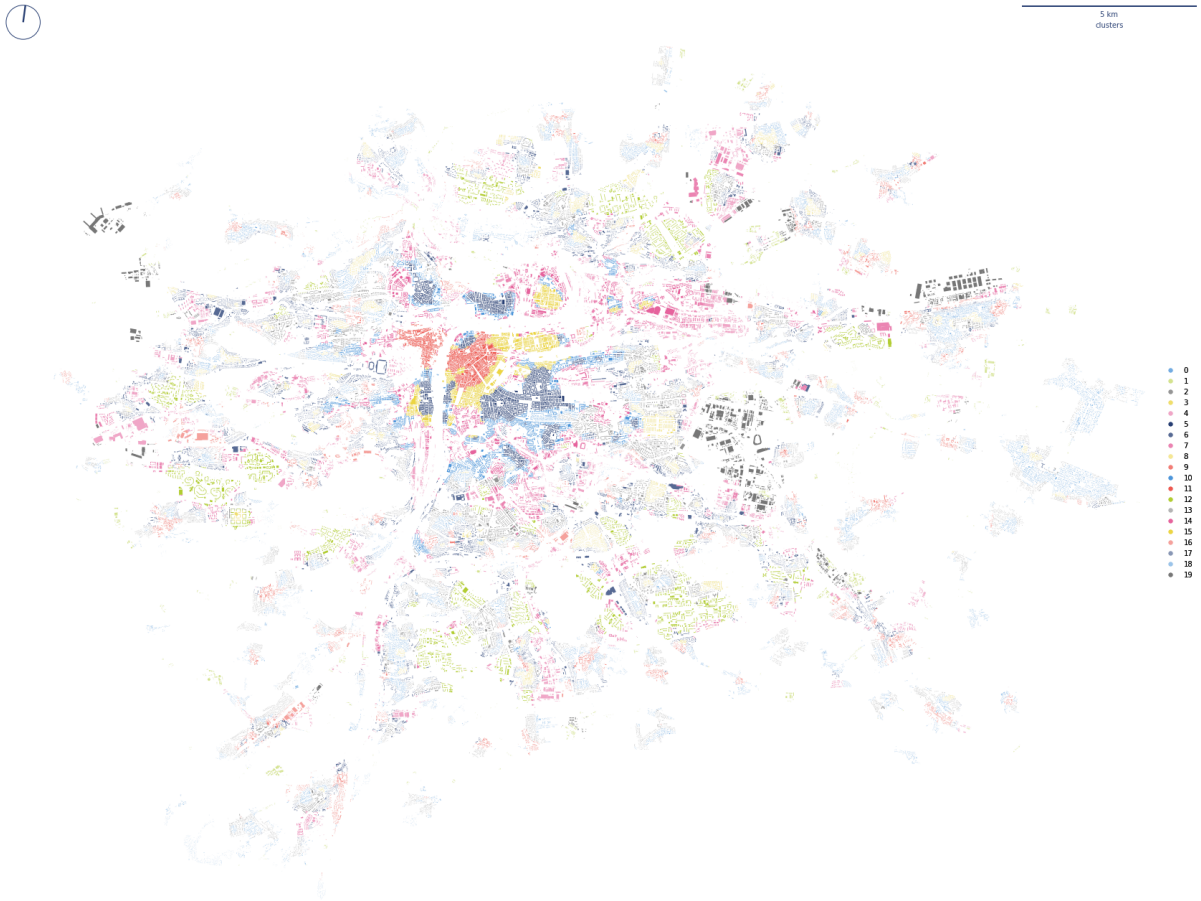


Figure 7.24: Spatial distribution of 20 clusters as identified by GMM based on complete data.

The 20 cluster seems to be relatively well defined and based on the first observation tend to reflect homogenous form. Even though there is no spatial constraint in the clustering itself, results show apparent contiguity caused by the design of contextual characters. The figure 7.25 shows detail of the section of Prague covering City Centre and the area towards the southern boundary for better understanding of results.

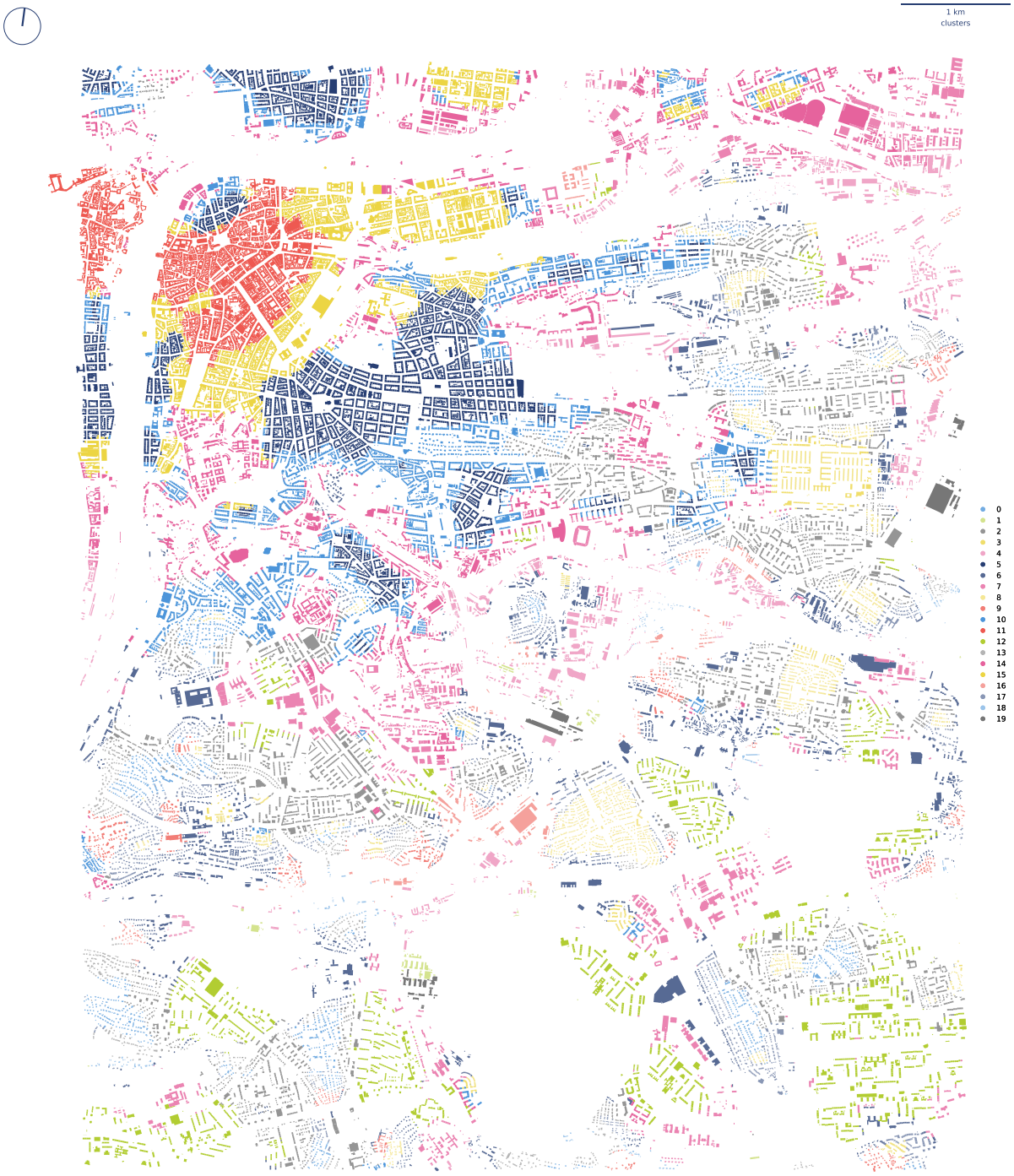


Figure 7.25: Detail of spatial distribution of 20 clusters as identified by GMM based on complete data.

Starting from the top left corner, where the historical core of Prague is, we can see (id 11 in red) delineation of what could be seen as medieval urban form, transitioning to compact perimeter blocks of Vinohrady neighbourhood (id 5 in dark blue). That is surrounded by less rigid heterogenous perimeter block-like tissues (id 10 in light blue) and then fringe areas (id 7 in pink). Towards south and east are present low-rise tissues (id 8 and 3 in lighter yellow) and modernist developments (ids 2 and 12 in grey and green). Drawing from the pure observation, clusters seems to be very precise and detailed and, most importantly, meaningful in terms of their link to the concept of the urban tissue.

The following section describes each of the identified clusters to give a detailed overview and understanding of what each cluster is composed.

7.2.3.1.2.1 Individual clusters Each of the individual clusters is presented by one example (usually the largest contiguous area) and its surroundings within 1,5km buffer. Colour schema is the same as in figures 7.24 and 7.25 and will be kept throughout the chapter. Clusters are sorted according to their ID, which is randomly assigned.

The first cluster (figure 7.26), noted as 0, is composed of predominantly low-rise, single-family housing. It has mostly residential character and tends to be located in the outer parts of the city, further away from the city centre. It is the largest of all clusters, with 15337 features, which is approximately 11% of all buildings in the study area.

Figure 7.26: Example of cluster 0 and its surroundings within 1,5km buffer located at the eastern boundary of study area. Aerial image courtesy of mapy.cz

The cluster on figure 7.27, noted as 1, contains mostly small-scale industry areas with small coverage, relatively small buildings. Often is adjacent to other clusters. It tends to be located in outer rings of the city but is overall very sparsely distributed. With only 2038 (less than 1.5%) features is one of the smallest clusters overall.

Figure 7.27: Example of cluster 1 and its surroundings within 1,5km buffer located at the north-east of study area. Aerial image courtesy of mapy.cz

Cluster 2, shown in figure 7.28 is one of the urban tissue types following modernist principles of spatial configuration, with linear buildings, but still relatively connected street network. These areas are mostly infills of the existing structure located within the city (except its central part) rather than on the periphery. It is relatively abundant with 12016 features (approximately 8.5%).

Figure 7.28: Example of cluster 2 and its surroundings within 1,5km buffer located at the north-west of study area. Aerial image courtesy of mapy.cz

Cluster 3 (figure 7.29) is one of the smaller ones. Its structure is defined by row-houses, a typology which is not very common in Prague. There are only 4133 features, less than 3% of all buildings scattered mostly in peripheral locations.

Figure 7.29: Example of cluster 3 and its surroundings within 1,5km buffer located at the east of study area. Aerial image courtesy of mapy.cz

Cluster 4 (figure 7.30) is one of the types with an industrial character, in this case being distributed as sort of infill development in the fringe areas. It is mostly adjacent to other urban tissues, relatively evenly distributed across the study area. It is composed of 5281, which is 3.8% of the total number.

Figure 7.30: Example of cluster 4 and its surroundings within 1,5km buffer located at the south of study area. Aerial image courtesy of mapy.cz

Cluster 5 (figure 7.31) can be best described as compact perimeter block-based residential area. This dense, grid-like development is located in the central areas of the city around the historical core and is one of the best defined urban tissues in Prague. There are 5930 of features belonging to this cluster, which is a bit more than 4.2% of the total count.

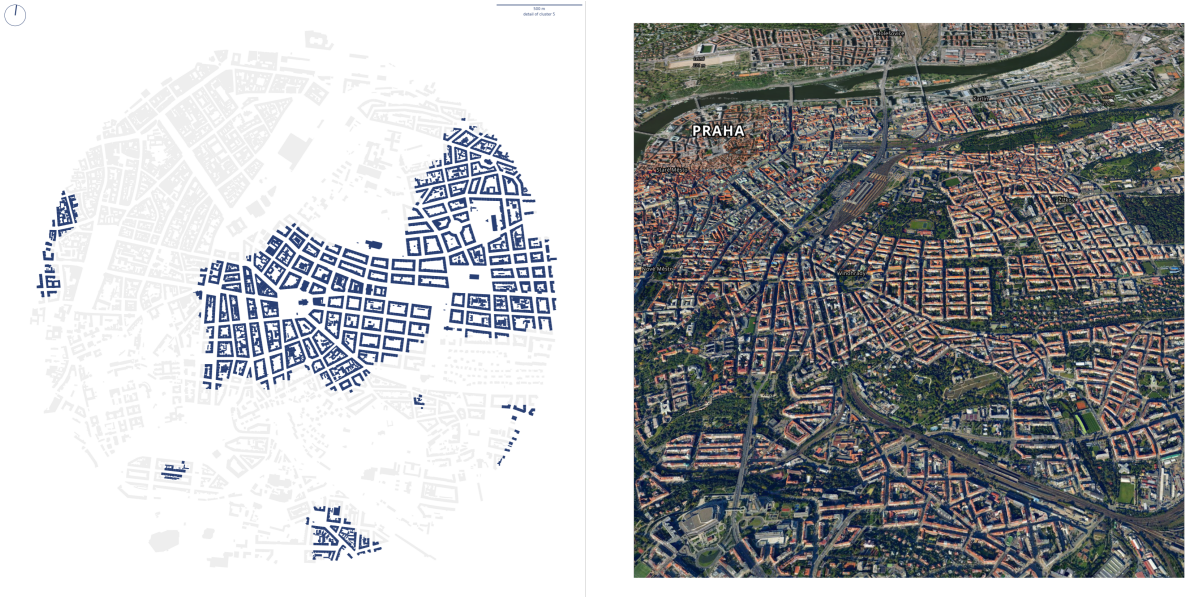


Figure 7.31: Example of cluster 5 and its surroundings within 1,5km buffer located at the centre of study area. Aerial image courtesy of mapy.cz

Cluster 6 (figure 7.32) is very different from the previous one as it contains fringe low-rise, not very well defined urban tissues. These are small-scale tissues scattered evenly around the study area, adjacent to other types of tissues, often filling topographically inconvenient areas. There are 10329 of these features, which is about 7.4%, so it is one of the more abundant clusters.

Figure 7.32: Example of cluster 6 and its surroundings within 1,5km buffer located the south-east direction from the city centre. Aerial image courtesy of mapy.cz

Cluster 7 (figure 7.33) is an example of more heterogeneous area. It has a similar character as cluster 4, but unlike that, it often contains other types of development with a less defined structure, like contemporary housing or office parks which do not reflect traditional rules of spatial configuration. That leads to higher heterogeneity in the area, making these tissues complicated to define. It consists of 4140 features, which is nearly 3% of the total amount.

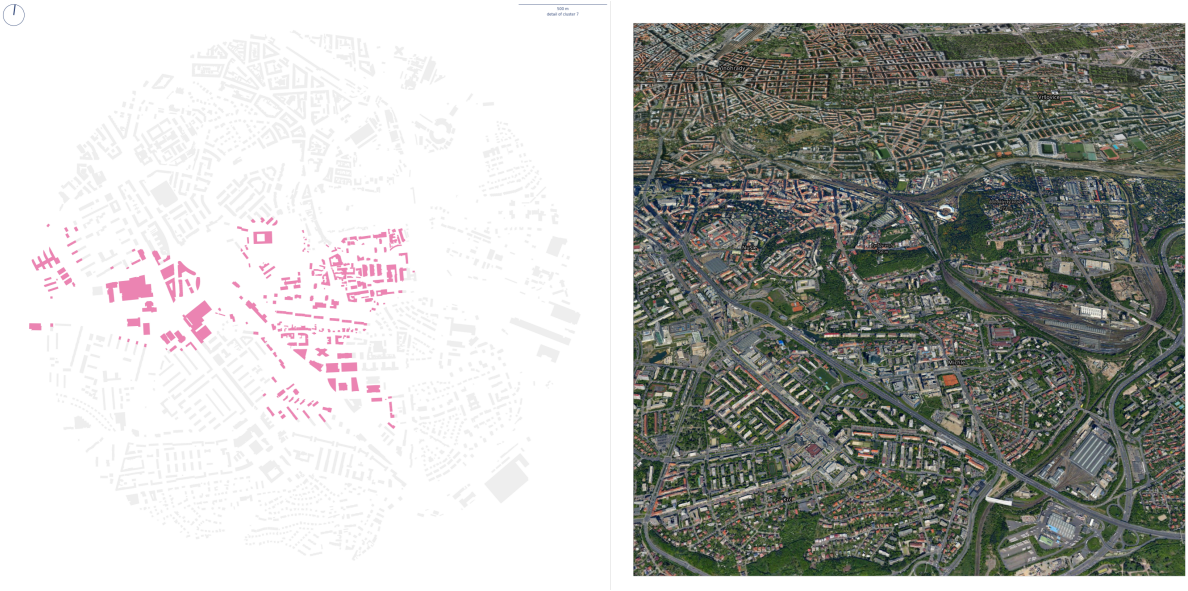


Figure 7.33: Example of cluster 7 and its surroundings within 1,5km buffer located the souther direction from the city centre. Aerial image courtesy of mapy.cz

Cluster 8 (figure 7.34) predominantly contains single-family housing in a relatively dense setting resembling garden city movement development. These places have interconnected network of relatively grid-like character, with buildings adjacent to each other either as row-house typology or similar. There are 7845 features within this cluster (5.6%).

Figure 7.34: Example of cluster 8 and its surroundings within 1,5km buffer located the souther direction from the city centre. Aerial image courtesy of mapy.cz

Cluster 9 on figure 7.35 seems to identify low-rise areas of organic development, which seems to be cores of the historical villages around Prague. These are small-scale tissues evenly distributed in the outer ring of development, now mostly embedded in the other development. They compose 5.6% of total features (7862).

Figure 7.35: Example of cluster 9 and its surroundings within 1,5km buffer located at the south-west of the study area. Aerial image courtesy of mapy.cz

Cluster 10 (figure 7.36) is very often adjacent to cluster 5 (compact blocks) or composes its own areas of block-based development. However, unlike in cluster 5, these blocks tend to be skewed or distorted in some other way. In some cases, this cluster could be seen as a transitional area between homogenous compact blocks and other types of urban tissue. These are 7203 features within this group, making 5.1% of the total amount.

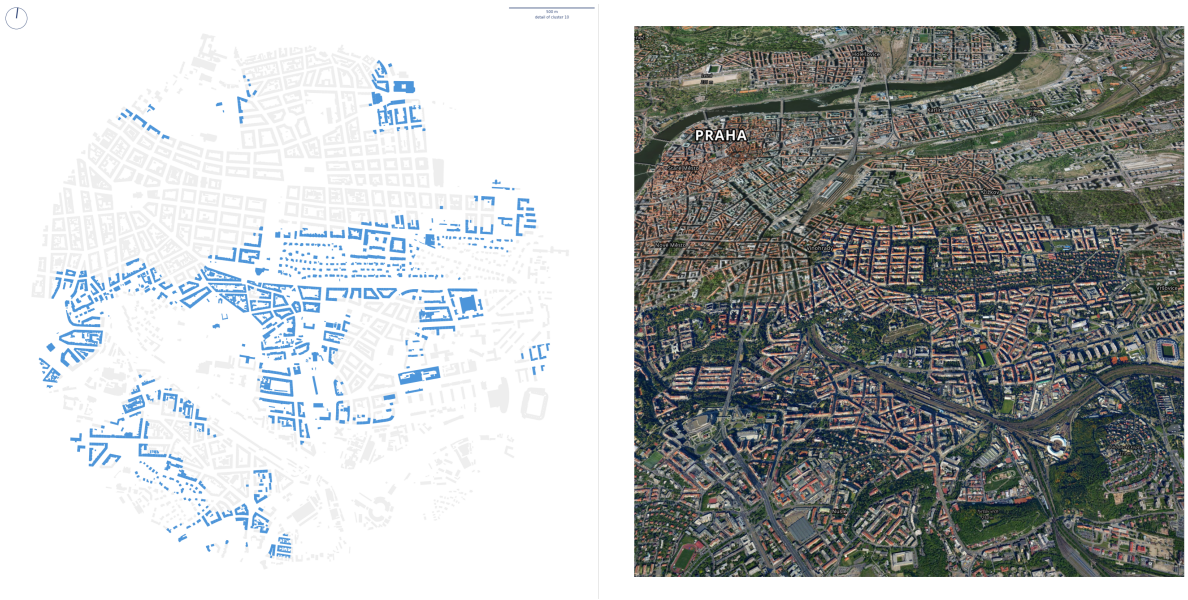


Figure 7.36: Example of cluster 10 and its surroundings within 1,5km buffer located next to the city centre. Aerial image courtesy of mapy.cz

Cluster 11 (figure 7.37) is a very straightforward one to describe as it composes solely of historic medieval core of Prague. It includes areas on both sides of the river and correctly excludes the area cut-out of the Old Town, which has been demolished in 19th century and rebuilt after that (Hrůza, 2003)f. There are 2167 features, making historical core one of the smallest clusters of all, composing only 1.5% of the total amount.

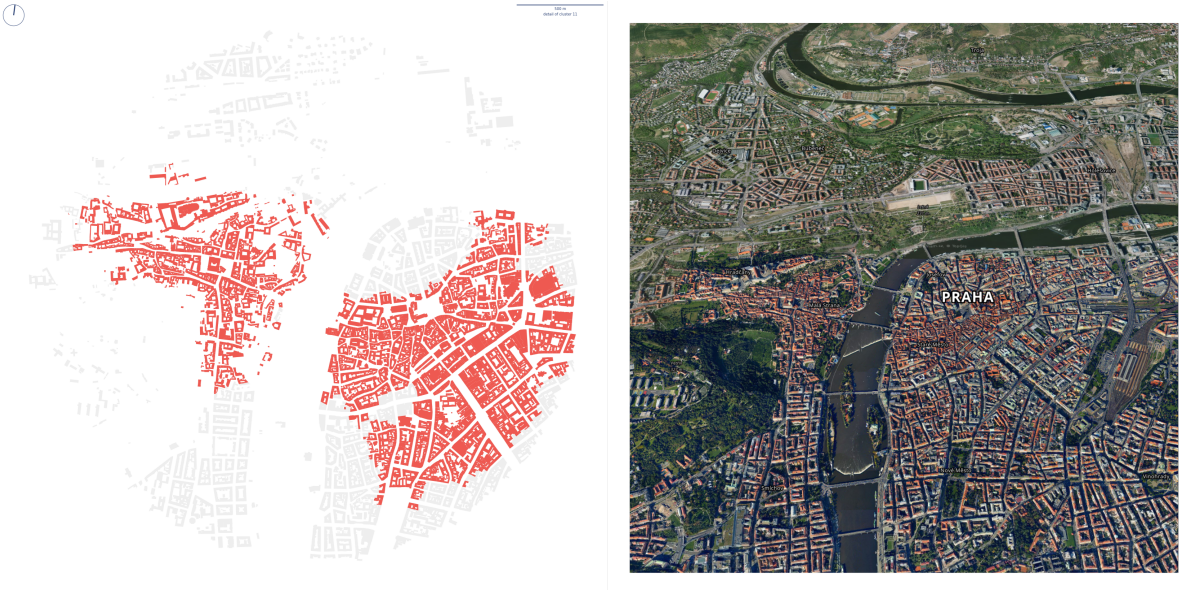


Figure 7.37: Example of cluster 11 and its surroundings within 1,5km buffer located in the city centre. Aerial image courtesy of mapy.cz

Cluster 12 is another very distinct one. As illustrated in figure 7.38, the origin of the development is modernist, covering large-scale modernist housing estates. These are typical with slab buildings, the incoherent relationship between buildings, plots and streets and large amounts of open spaces, among other characteristics. In Prague they are almost exclusively on the peripheral ring of the city, forming a so-called modernist belt of Prague. They consist of 6885 features, which is 4.9% of the total amount.

Figure 7.38: Example of cluster 12 and its surroundings within 1,5km buffer located on the southern edge of the city. Aerial image courtesy of mapy.cz

Cluster 13 (figure 7.39) is another example consisting of single-family housing. This time it is low-density development with predominantly detached buildings. It is typical with elongated blocks which in part is a reaction to the underlying topography. It is a very abundant cluster with 14992 features, making more than 10.6% of the total amount, distributed along the periphery of the city.

Figure 7.39: Example of cluster 13 and its surroundings within 1,5km buffer located on the south-western edge of the city. Aerial image courtesy of mapy.cz

Cluster 14 is distributed almost exclusively within the wider centre of Prague, often adjacent to the homogenous compact city as is illustrated on the figure 7.40. The cluster could be defined as an inner fringe composed of heterogeneous developments on the edge of existing homogenous one. There are 4984 features within it, making 3.6% of the data.

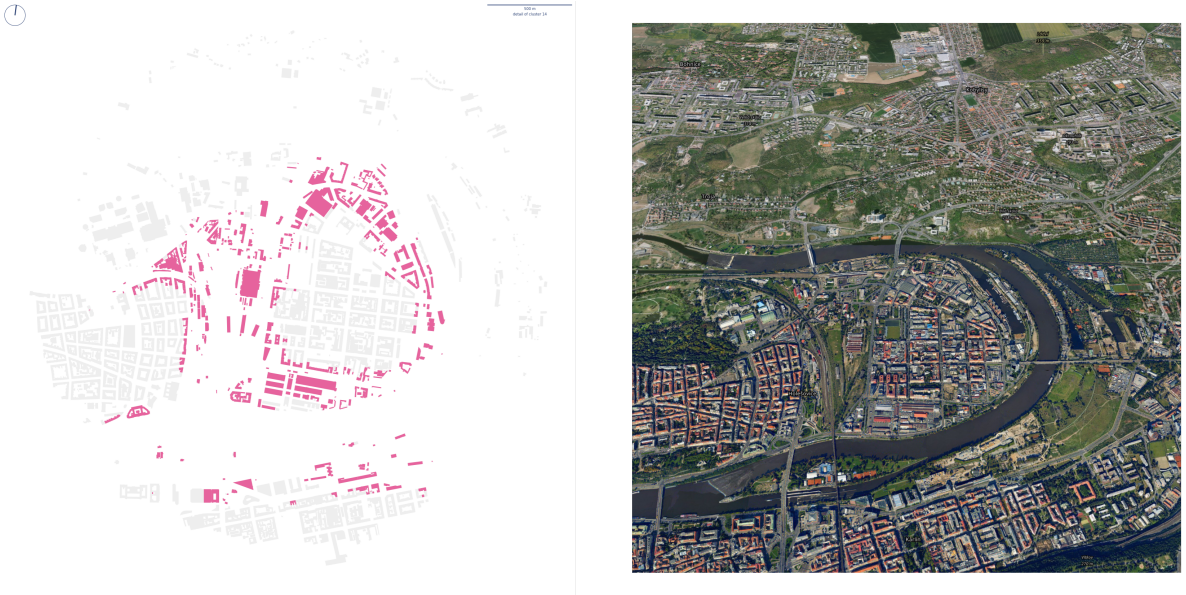


Figure 7.40: Example of cluster 14 and its surroundings within 1,5km buffer located north of the city centre. Aerial image courtesy of mapy.cz

Cluster 15 (7.41) is perimeter-block based tissue type with very heterogeneous development in the block interiors. It has a very high coverage area ratio located in in the city centre either as a transitional area between medieval core and compact city or as industrial development. There are only 3060 features within the cluster (2.2%).

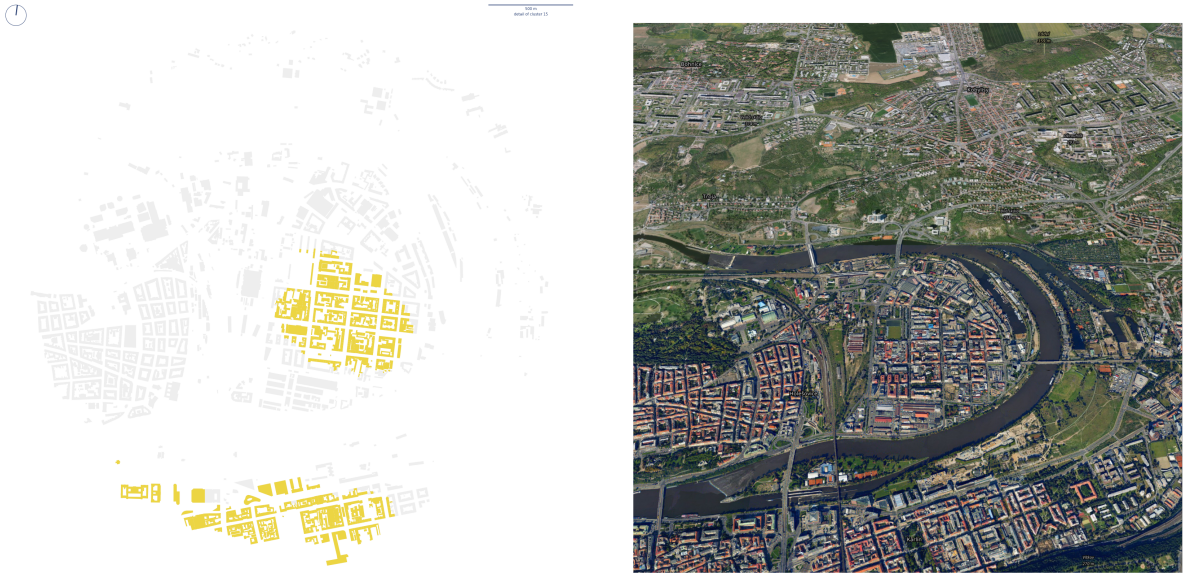


Figure 7.41: Example of cluster 15 and its surroundings within 1,5km buffer located north of the city centre. Aerial image courtesy of mapy.cz

Cluster 16 (7.42) is not very straightforward to define as it is a heterogeneous one. It mostly consists of small patches of not very well defined tissues with the predominant role of small-scale buildings but not exclusively. It may be seen as *other*, combining parts of the dataset which do not fit elsewhere, but at the same time, all places have the similar character of being *out of sight*. It is evenly distributed, but not very abundant one with 3548 making approximately 2.5% of the dataset.

Figure 7.42: Example of cluster 16 and its surroundings within 1,5km buffer located north of the city centre. Aerial image courtesy of mapy.cz

Cluster 17 (7.43) is another of the low-density single-family tissue types. It has a less defined and rigid structure, and it is often adjacent to open space. It does have a certain inner heterogeneity expressed as various kinds of buildings from detached to row houses. Like the other similar clusters, this is also relatively abundant with 12145 features (8.7%).

Figure 7.43: Example of cluster 17 and its surroundings within 1,5km buffer located west of the city centre. Aerial image courtesy of mapy.cz

Cluster 18 (7.44) consists of relatively independent detached areas of low-density village-like development. Clusters can be only a strip along the road or other open-space facing tissues. It is located mostly on the periphery of the city and entails 8764 features (6.2%)

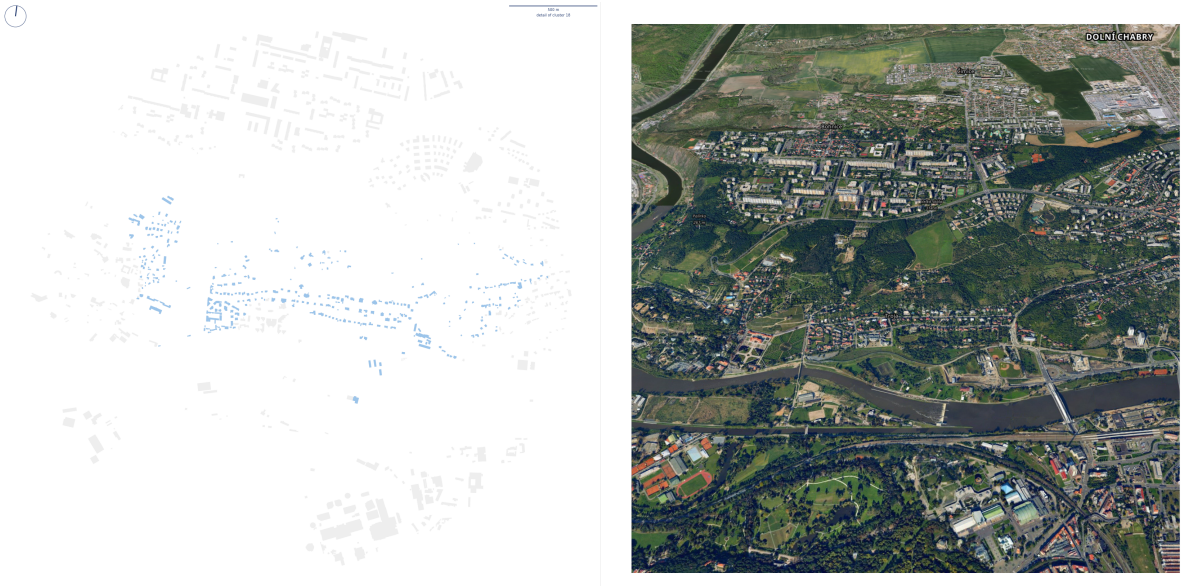


Figure 7.44: Example of cluster 18 and its surroundings within 1,5km buffer located on the north of the city. Aerial image courtesy of mapy.cz

The last cluster, 19 illustrated on figure 7.45 is industrial urban tissue type, consisting of a mixture of large-scale and small-scale buildings, convoluted street network and a minimum of residential use. It is only at a few places, but of a large-scale, mostly towards the edge of the city. There are only 1656 features within the cluster, making 1.2% of the total amount.

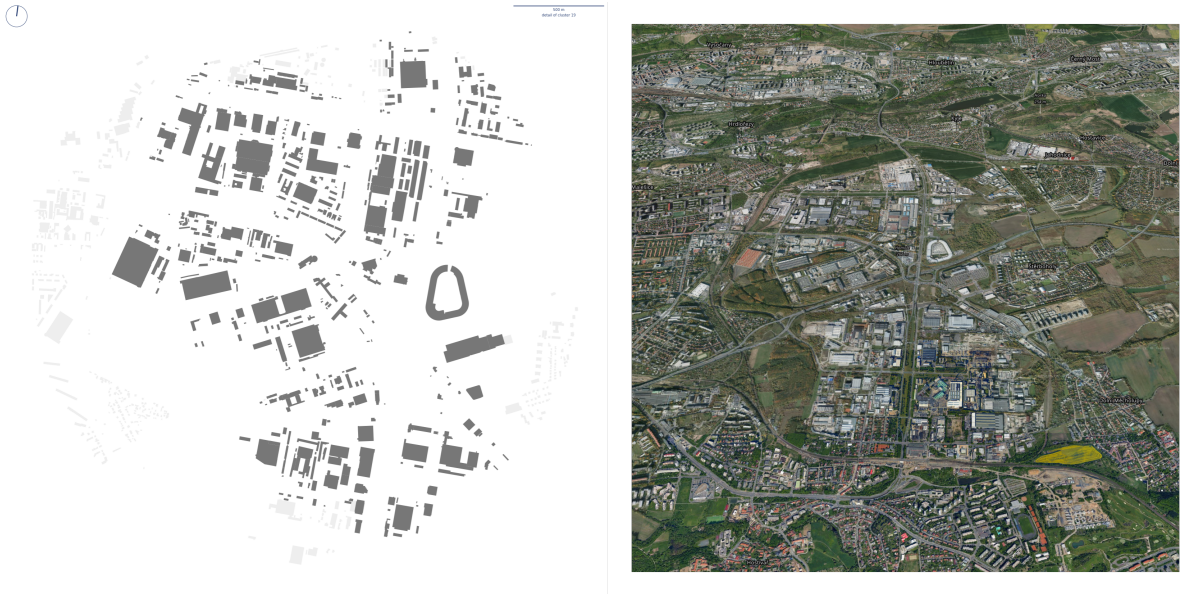


Figure 7.45: Example of cluster 19 and its surroundings within 1,5km buffer located on the eastern edge of the city. Aerial image courtesy of mapy.cz

From the overview is clear that some clusters are very distinct like the historical core (11) or modernist estates (12), while others resemble each other as is the case of low-density single-family clusters (0, 8, 13, 17). However, even between these seemingly similar clusters are recognisable differences. Numerical assessment of differences between clusters is part of Chapter 8, to determine which characters are causing the distinction and understand the clusters based on their morphometric profiles.

7.2.3.2 Sampled data

Gaussian Mixture Model uses training data on the input to estimate the optimal clustering and then predicts the probability that each feature belongs to any of the components. That means that training data do not have to equal the data we want to classify. The GMM and

especially the estimation of the number of components, which does GMM repeatedly, could have relatively high computational demands as the size of the dataset grows. The Prague example, with 140,000 features took approximately 60 hours to measure all BIC values and do the final clustering on a desktop computer with a 12-core Intel Xeon processor. Running larger areas at once may get unfeasible, it is hence critical to understand if the method can work with sampled data.

Sampled clustering would use a randomly selected fraction of the data as a training set and then use it for prediction on the complete data. That might significantly reduce computational demands because they rise exponentially with the growing dataset, but at the same time might not provide useful results. Sampling procedure might miss some clusters entirely (none or very few features are included in the sample) or affect the results in another way. The following section tries to answer some of the questions comparing the clustering based on the complete dataset with sampled one.

7.2.3.2.1 Sampled Bayesian Information Criterion Three versions of sampling are assessed - 10%, 25%, and 50% of the dataset. Because random sampling results in different samples each time, which could affect BIC, each option is sampled three times and GMM is run three times on each (in total nine runs of GMM per option). To assess the number of components, values from range 2 - 40 were tested. BIC is measured using the complete dataset. The resulting values are shown in figure 7.46 below.

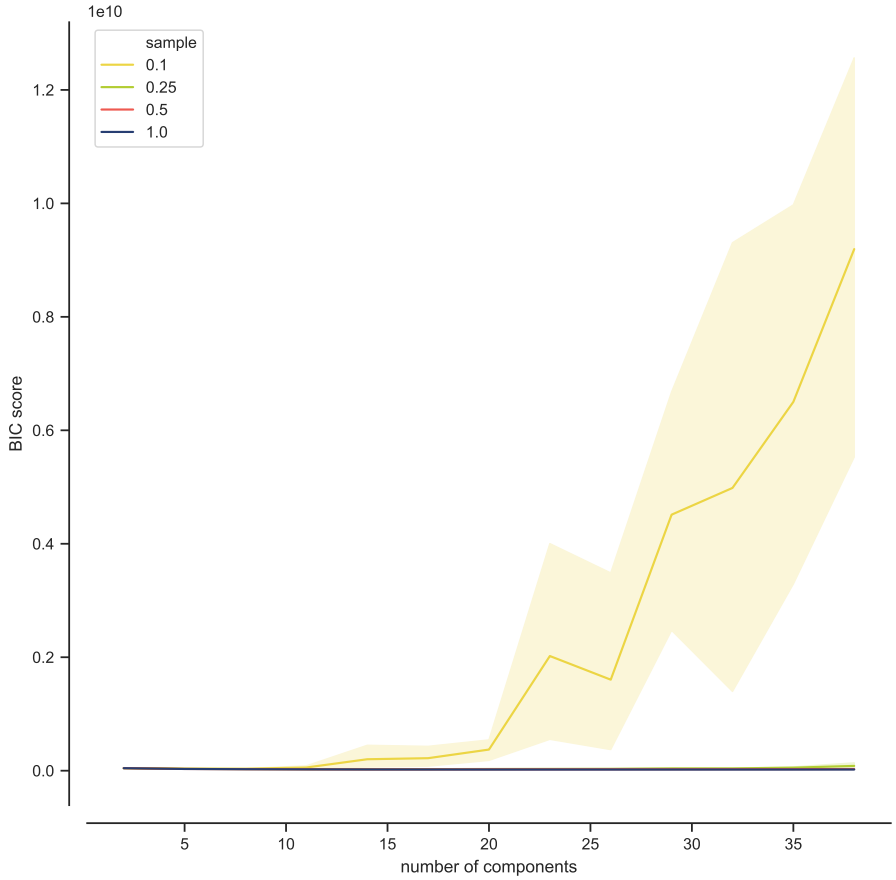


Figure 7.46: Bayesian Information Criterion score for sampled clustering. Shaded area reflects .95 confidence interval.

Figure 7.46 shows a striking difference between the results of 0.1 (10%) sampling and the rest. The BIC score for this option is significantly higher than for the rest, indicating that the sample is way too restricted to capture the structure of the dataset and generate meaningful clustering. Due to this difference, any differences between the other options are not recognisable. For that reason, figure 7.47 shows the same data without 0.1 option.

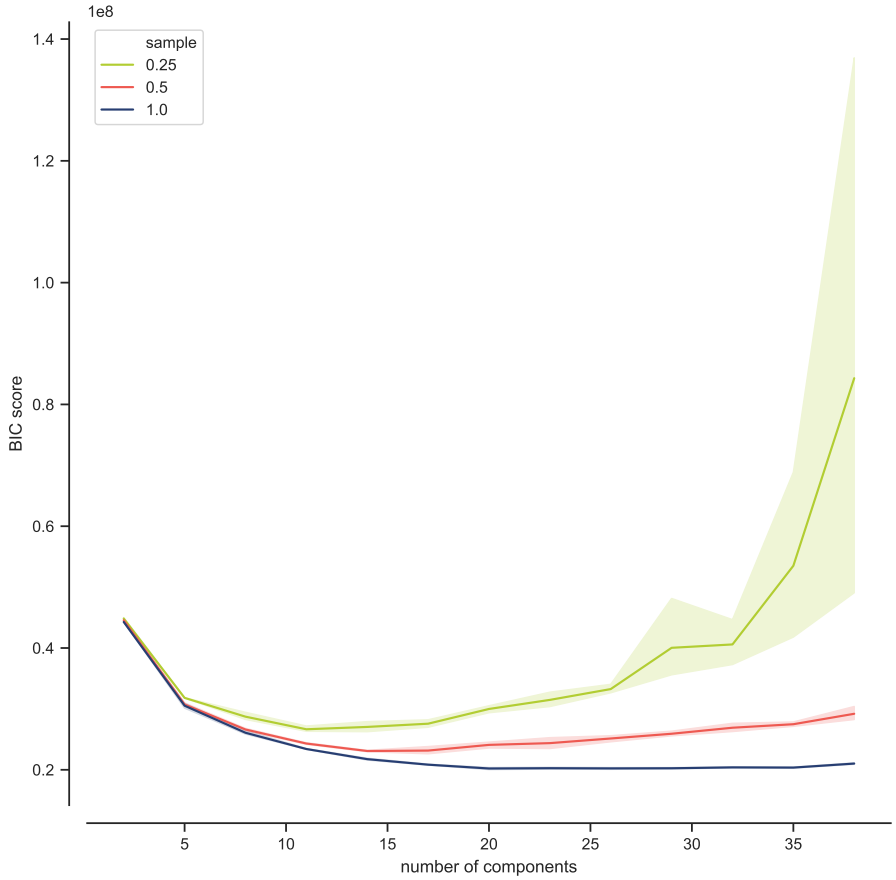


Figure 7.47: Bayesian Information Criterion score for sampled clustering excluding 0.1 sampled results. Shaded area reflects .95 confidence interval.

All two remaining options show similar curves as was already seen in the complete clustering. The bigger the sample is, the better results can GMM provide. One key difference between the samples is the resulting optimal number of components. It seems that the smaller the sample is, the sooner BIC curve culminates, which results is a smaller number of optimal components. 0.25 sampling culminates at 11 components, 0.5 at 15 components and 1.0 (complete data) at already mentioned 20 components. The difference between 0.25 and 1.0 both in terms of BIC and an optimal number of components is big, so it is questionable if such a small sample can provide any similar results.

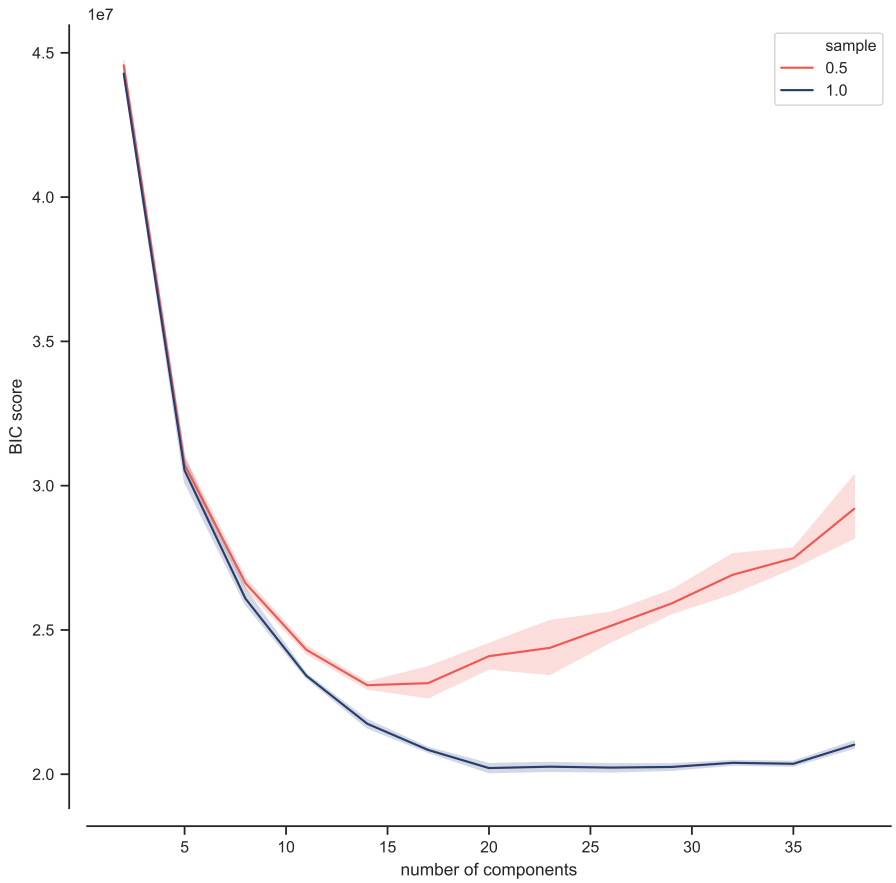


Figure 7.48: Bayesian Information Criterion score for sampled clustering excluding 0.1 and 0.25 sampled results. Shaded area reflects .95 confidence interval.

Figure 7.48 compares 0.5 and 1.0 sampling only as the difference is not so dramatic. Unlike 1.0 sampling, where the point of culmination is not as clear, 0.5 culminates sooner, and the curve starts ascending quicker. That makes the decision of optimum easier. Following the same principle as in the previous case, the first significant minimum is 15 components. The BIC score overall is worse than in non-sampled case, but it is worth testing the similarity of the actual cluster recognition.

7.2.3.2.2 Sampled distinct homogenous clusters Identification of distinct clusters based on sampled data (random sample 50%) using 15 components results in the spatial distribution of clusters illustrated on figure 7.49.

Figure 7.49: Spatial distribution of clusters based on sampled data (0.5) and 15 components within the whole study area.

Visual assessment of clustering indicates that the clusters seem to be meaningful and could be seen as a proxy of urban tissues. Due to the smaller number of components, some areas are showing less differentiation than in the complete clustering, but the difference does not seem to be in terms of correctness or wrongness of one or the other clustering, but only in terms of the change of the resolution of results.

7.2.3.2.2.1 Comparison of sampled and complete clustering Three easy-to-interpret clusters are compared 1:1 and their composition and shape are assessed to understand what are the actual on-ground differences between two versions of clustering apart from the different number of components.

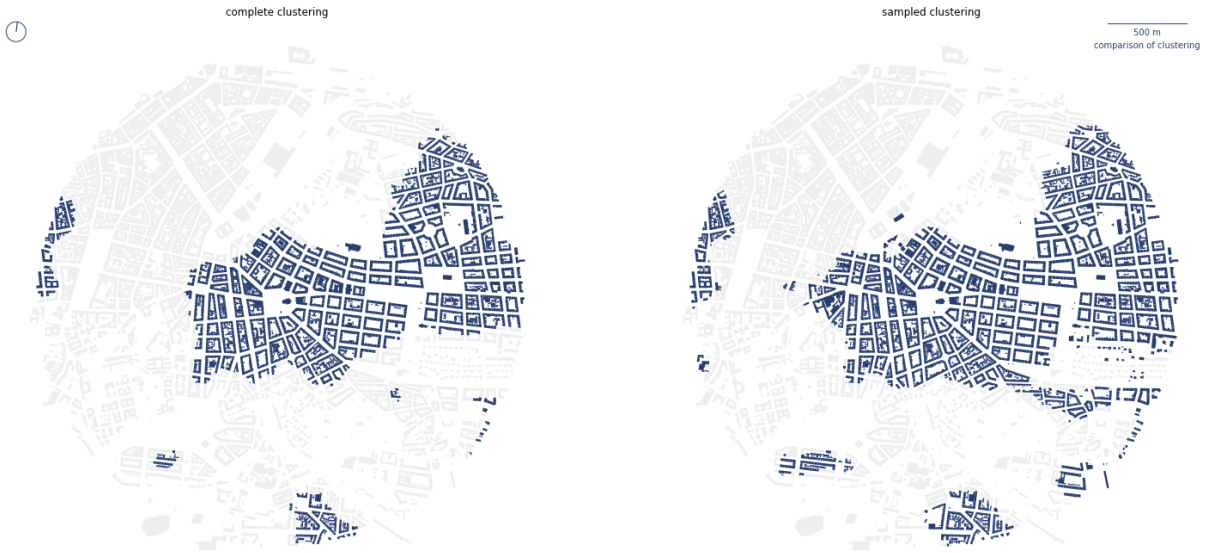


Figure 7.50: Comparison of spatial distribution of cluster 5 and sampled cluster 4 in the city centre.

Original cluster 5, capturing compact perimeter blocks, has its counterpart in sampled cluster 4. The example of their spatial distribution is illustrated in figure 7.50. Both versions capture mostly the same type of urban tissue with a very similar footprint. The only apparent difference is that sampled cluster is more inclusive (covering larger area) than a complete cluster, likely due to the smaller overall number of clusters (clusters needs to be naturally more abundant).

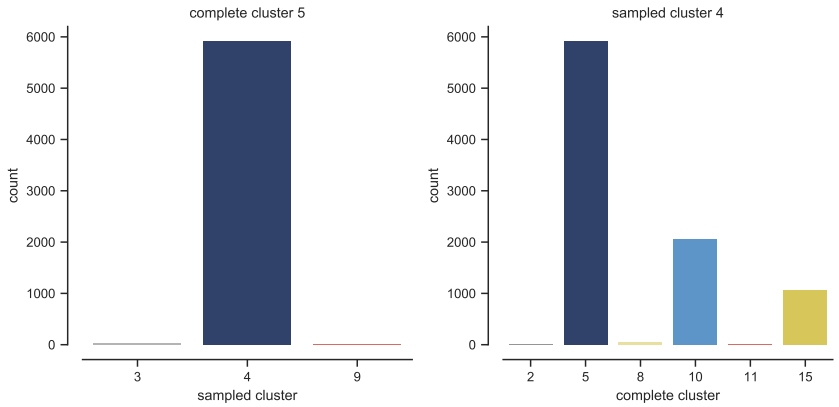


Figure 7.51: Composition of cluster 5 and sampled cluster 4 in relation to each other. Shows the number of features labelled as studied cluster and their labels in the other clustering variant.

The comparison of the composition of both clustering versions in relation to each other on figure 7.51 shows that features initially marked as being in the complete cluster 5 are almost entirely within sampled cluster 4. On the other hand, features labelled as sampled cluster 4 are predominantly located in complete cluster 5, but due to higher inclusiveness also in clusters 10 and 15.

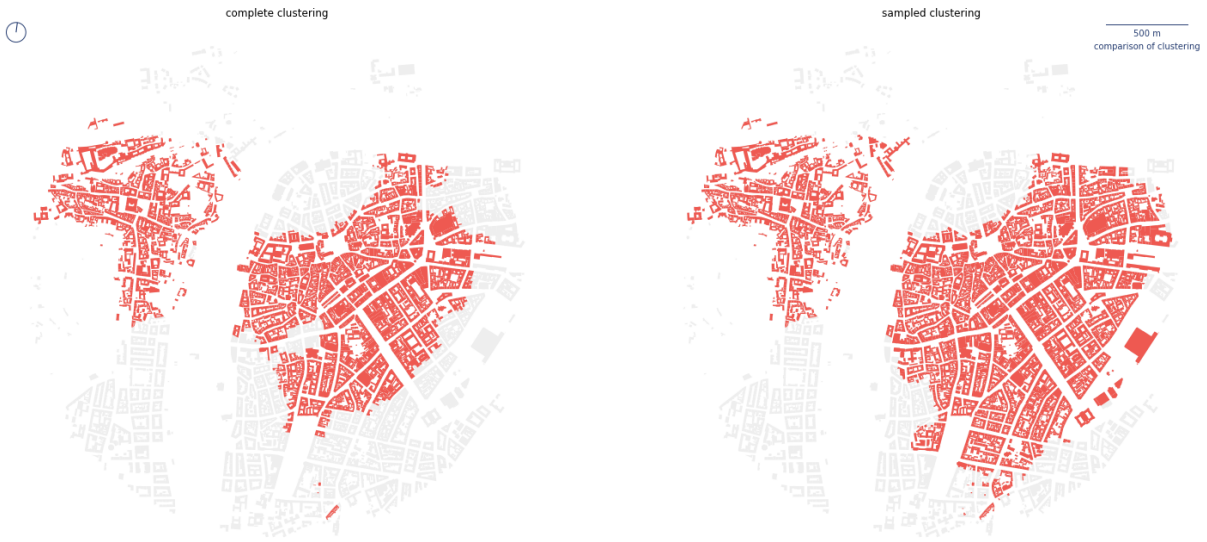


Figure 7.52: Comparison of spatial distribution of cluster 11 and sampled cluster 9 in the historical core.

The second example based on complete cluster 11 (figure 7.52) representing the historical core of Prague shows a similar story as the first one. Both versions correctly delineate medieval urban tissue and avoid newer redevelopment of the former Jewish quarter in the North. The difference is in inclusiveness, where the sampled cluster covers more extensive areas, which are in the complete clustering seen as cluster 15, the transitional one.

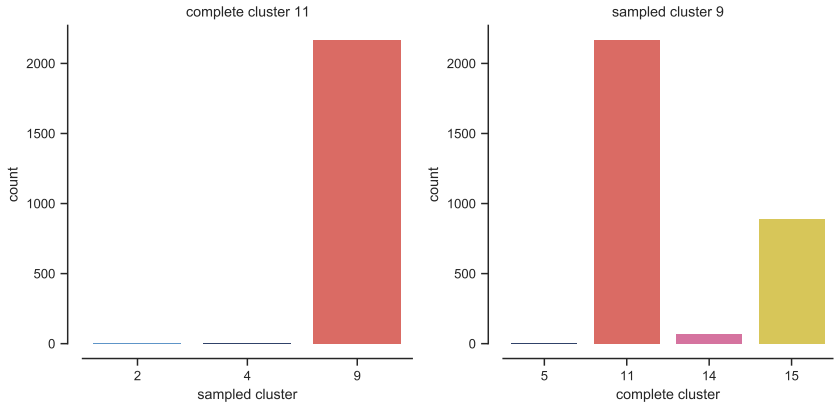


Figure 7.53: Composition of cluster 11 and sampled cluster 9 in relation to each other. Shows the number of features labelled as studied cluster and their labels in the other clustering variant.

The assumption derived from the visual assessment is correct based on the numerical data on the actual composition (figure 7.53). Features classified as cluster 11 in the complete clustering, are classified as cluster 9 in sampled clustering. Features classified as cluster 9 in sampled clustering, however, include parts of complete cluster 15 (less than 1/3 of cluster 9 comes from 15).

Figure 7.54: Comparison of spatial distribution of cluster 12 and sampled cluster 5 in the west of the city.

Even more, similarity shows a comparison of large-scale modernist urban tissues (figure 7.54) with very few differences which could be derived from visual observation. Even

the complicated case of modernist area mimicking perimeter blocks in the South of the example shows the same pattern, with middle part being excluded from the rest (likely an effect of a configuration of the street network).

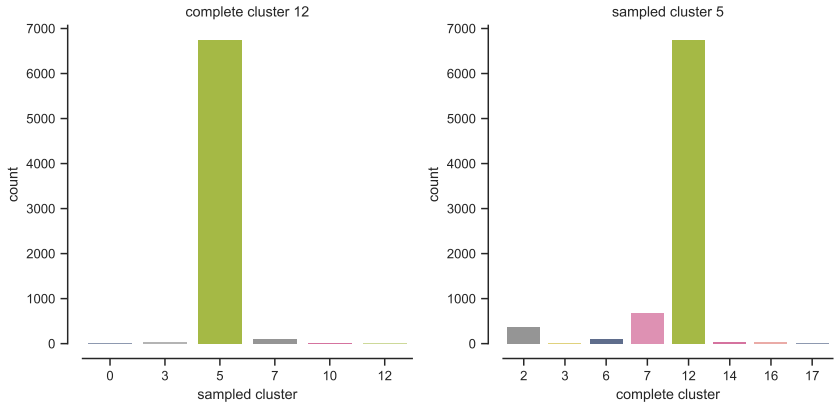


Figure 7.55: Composition of cluster 12 and sampled cluster 5 in relation to each other. Shows the number of features labelled as studied cluster and their labels in the other clustering variant.

The composition of clusters is, with a few exceptions equal (figure 7.55). The same features which belong to cluster 12 in the complete clustering are labelled as cluster 5 in the sampled clustering and vice versa.

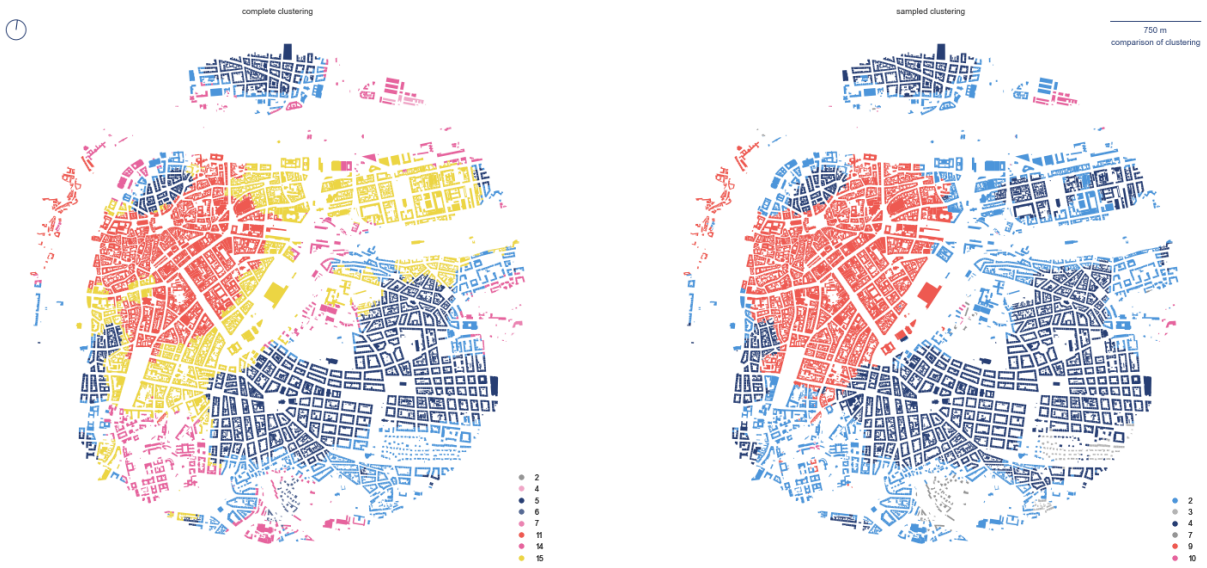


Figure 7.56: Comparison of the city centre focusing on cluster 15, which is not present in the sampled clustering.

Three examples above show that there is a striking similarity between both results. However, there is a different number of clusters, so where is the difference? The example on figure 7.56 shows cluster 15 based on complete data, which does not have its counterpart in sampled clustering. Instead, it is split into three almost equal parts (figure 7.57) each linked to another cluster. What was the so-called transitional area between medieval core and historical compact city is no longer present. That by itself is likely not a big issue, but it illustrates the behaviour of sampled clustering with a smaller number of components. It does not necessarily merge two similar clusters into one, but at some places splits clusters into multiple pieces. GMM, in this case, sees different data and hence might exclude some smaller clusters. Because these might have been *in between* other, parts are now closer to one and other parts closer to another cluster. The resulting clustering should then be seen as a different perspective using different resolution, rather than a coarser version of complete clustering.

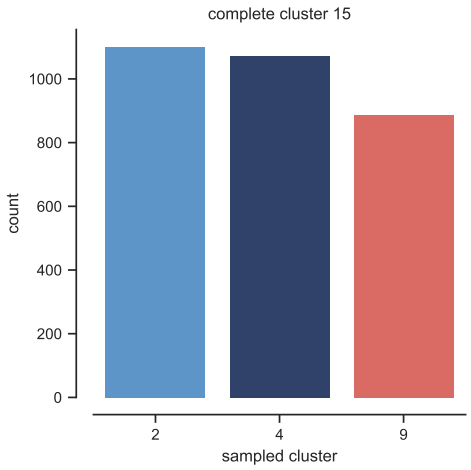


Figure 7.57: composition of cluster 15 in relation to sampled clustering. Shows a number of features labelled as studied cluster and their labels in the other clustering variant.

Looking onto other clusters which do not have a counterpart in sampled clustering (apart from 15, 2, 6, 9, and 10), none of them is *swallowed* by one larger cluster. All are split into two sampled clusters. Sometimes more equally (e.g., cluster 2 is equally split between 10 and 14), sometimes less equally (e.g., cluster 6 is more present in sampled cluster 0 than sampled cluster 8). This illustrates the probabilistic rather than hierarchical nature of GMM. The full comparison is available in Appendix 7.9.

Depending on the aim of the study, sampled clustering could likely be used instead of complete clustering, considering the fact that results based on samples smaller than 50% are not precise enough. However, for the ideal, detailed identification of urban tissues, sampled clustering might provide sub-optimal results.

7.2.3.3 Sub-clustering

The trial of sub-clustering, i.e., division of existing clusters, obtained using the complete dataset will be done on two of the original clusters, which are very different. The first example will focus on cluster 5, compact perimeter blocks, and the second on the modernist belt of Prague labelled as cluster 12. The assumption behind sub-clustering is that the richness of the data may allow us to determine differences within the cluster. These are not significant from the perspective of the whole dataset, that is why they were not picked initially as independent clusters, but they might be significant internally.

7.2.3.3.1 Compact Prague The first case is the cluster 5, which could be interpreted as the urban type of compact, rigid perimeter blocks. The reason for its selection is that due to the varied topography. These blocks have to react to the steeper surface at some places, and the perceptional character of such areas is different from those laying on the flat grounds.

Sub-clustering uses contextual data of individual features within the cluster and performs the identification of types once again, starting from determination of the optimal number of components using BIC and consequent training of the model and prediction of labels.

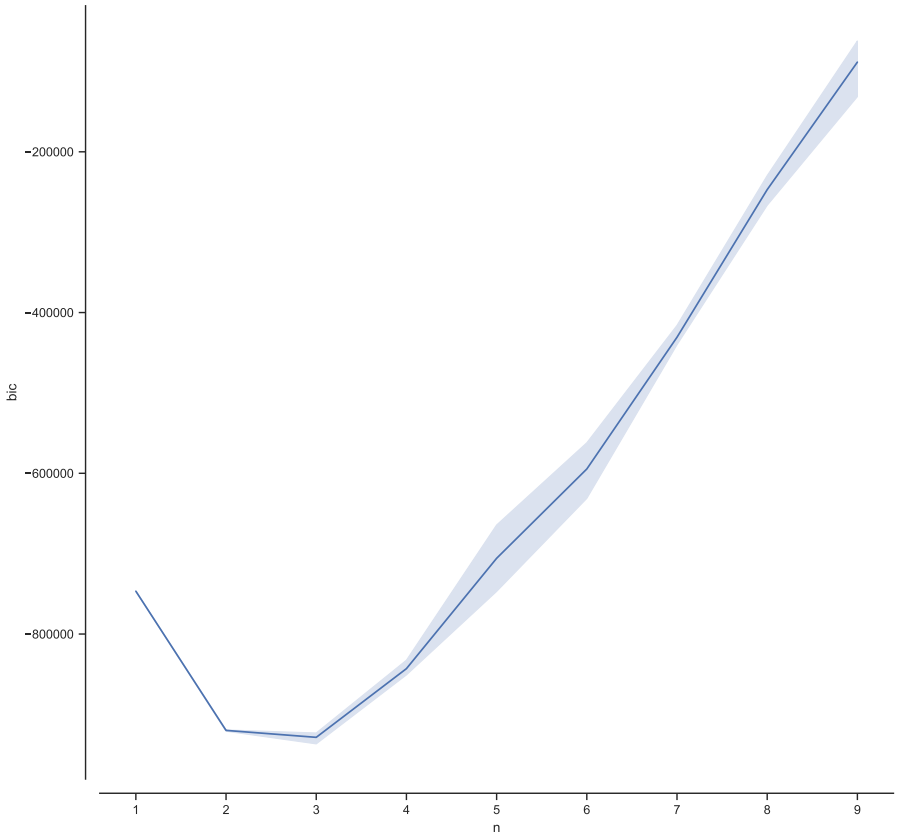


Figure 7.58: Bayesian information criterion score for cluster 5 sub-clustering to determine optimal number of components.

Bayesian Information Criterion illustrated in figure 7.58 indicates that there is a scope for sub-clustering as both 2 and 3 components have a better score than a single component. If the situation would be otherwise, and a single component would have the lowest BIC score, there would be no significant sub-clusters in the data and results of forced clustering would likely suffer from discontinuity. Following the rule of the first significant minimum, this trial works with two components.

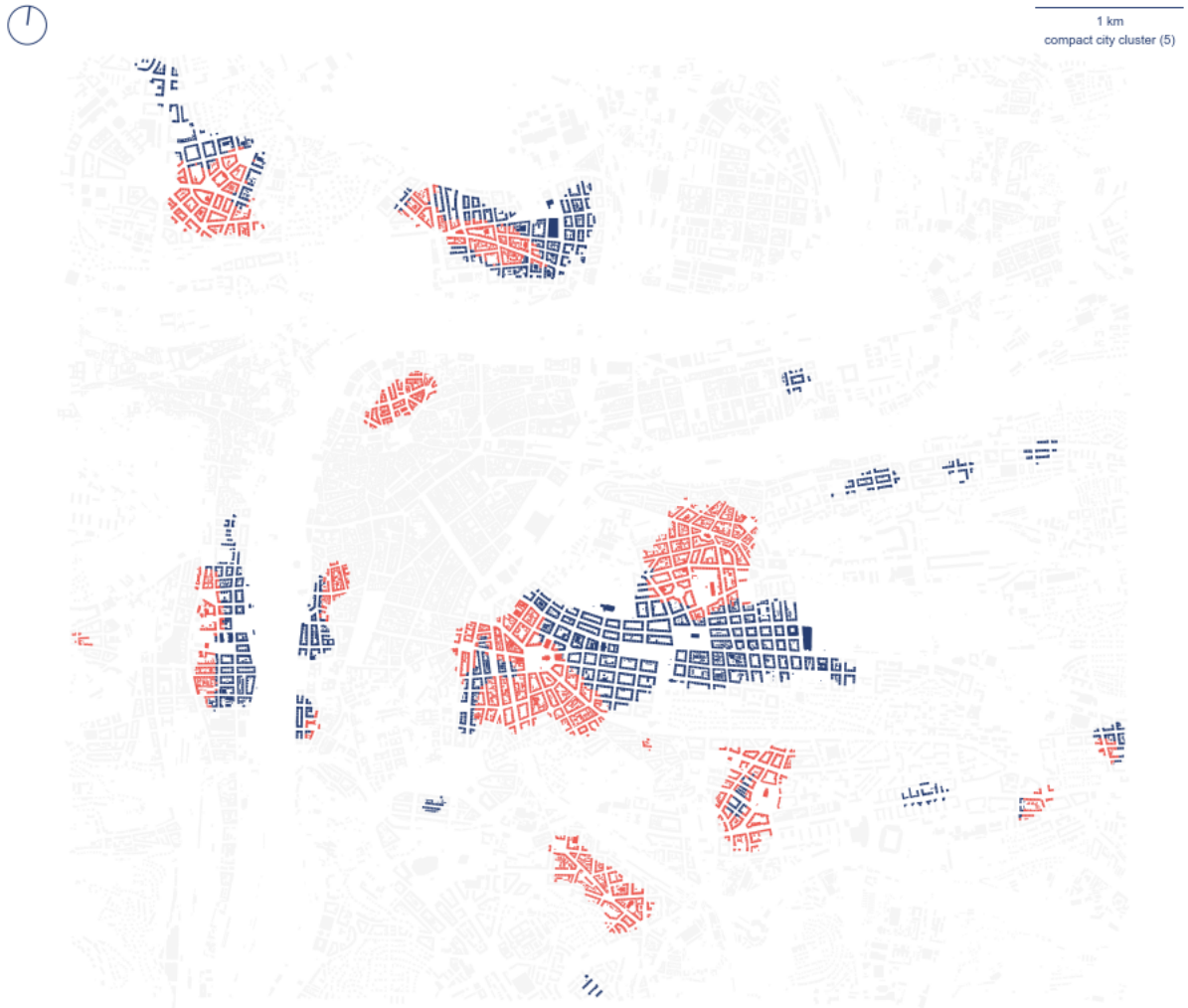


Figure 7.59: Spatial distribution of sub-clusters of cluster 5 marks the distinction between rigidly and distorted grids.

The result of sub-clustering of cluster 5 is shown in figure 7.59. It feels fair to conclude, that newly identified sub-clusters have a meaning and distinguish between areas which are more rigidly gridded and those which tend to have grid distorted.

7.2.3.3.2 Modernist Prague The second sub-clustering trial focuses on large-scale modernist housing estates on the periphery of Prague. There is an assumption of the inner differentiation of the relevant cluster 12 because each of these neighbourhoods has

been designed, and there were different authors and approaches in different places and periods of development (Hrůza, 2003). It is assumed that morphometric data should be able to reflect this difference.

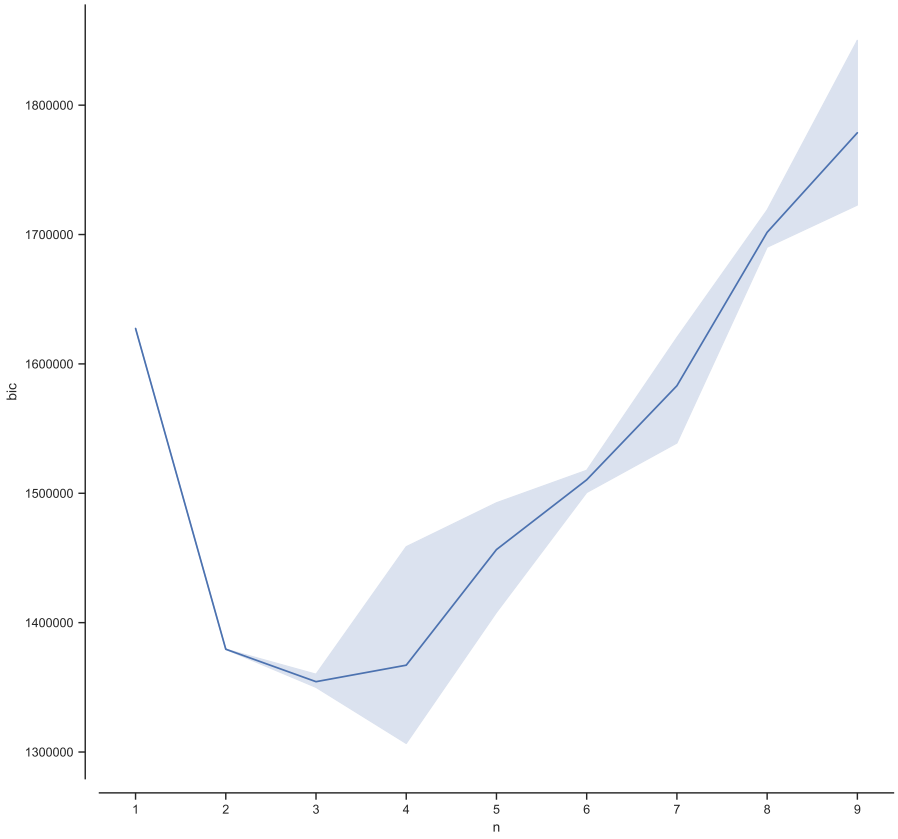


Figure 7.60: Bayesian information criterion score for cluster 12 sub-clustering to determine optimal number of components.

BIC results on figure 7.60 indicates that subdivision of the cluster is significantly better than a single group with all the options between 2 and 7 having a lower score that one component. The first significant minimum, in this case, are three components.

Figure 7.61: Spatial distribution of sub-clusters of cluster 12 showing the different location of all three groups.

The map on the figure 7.61 shows the whole cluster 12 divided into three sub-clusters. The interesting case is the green group, located exclusively on the western edge of the study area. The fact that it is not present anywhere else indicates that sub-clustering indicates that results are not affected by randomness. A closer look at the differences as illustrated on the figure 7.62 shows why these tissues are split in such a way. The green sub-cluster has large blocks and a circular character. The red one tends to be a large-scale orthogonal configuration, while blue is smaller-scale more compact urban tissue.

Figure 7.62: Comparison of examples of sub-clusters of cluster 12 illustrating the structural differences.

Both examples above indicate that there is a scope for sub-clustering if the research using this method needs a more refined level of detail. As noted above, sub-clustering ability depends on the internal homogeneity of each cluster, and it may not be possible in some cases. However, in cases where this possibility is available, results show meaningful patterns, enabled by the richness of the morphometric dataset.

7.3 Summary

This chapter took morphometric elements organised within the relational framework of urban form and defined the rich set of characters to be used within the rest of the study. Morphometric characters, divided into 74 primary and 296 contextual characters were then tested on the case of Prague. As the set followed specific rules driving its definition, it proved to provide a complex multi-scale characterisation of the local context of each individual building. That served as an input of cluster analysis using Gaussian Mixture Model method, which delineated 20 potential types of urban tissues within the fabric of the city.

While the validation is left for Chapter 8, results of clustering illustrated on previous pages indicate that the morphometric method of identification of urban tissues and their types has a potential. The outcome of the Gaussian Mixture Model learning procedure does match the expectations of what a tissue type should be. The question remains what the relation of these clusters to the actual concept of urban tissues is.

One should be aware that cluster is a numerical, morphometric statistical **proxy** of urban tissue, not its definition and replacement. GMM clustering is non-deterministic, so boundaries are not fixed, but rather indicative. It is not a ground truth (there is no ground truth at all in fact), and the meaning of clusters and relation between them has to be determined and interpreted before any further steps. The one approach on how to do so is proposed in the next chapter.

Chapter 8

Taxonomic relationships of urban tissues

The previous chapters introduced the framework implementing comprehensive urban morphometrics on a metropolitan scale. Based on the selection of measurable characters, it established a method of a complex description of urban form on the granularity of individual buildings. As reported, derived information then enabled the identification of urban tissue types within urban form. Consequent results of the cluster analysis indicate the validity of the morphometric method in recognition of urban form patterns and the potential for both additional subdivision and upscaling of the model.

In this chapter, therefore, the observed clusters are employed in the role of input data and investigated from two perspectives - 1) conceptualisation of clusters as *OTUs*, leading to the taxonomic classification, 2) validity of clusters as a proxy of urban tissues via assessment of relation to additional data, the transferability of the method to the different geographical context, and expandability of the taxonomy.

This chapter relates to chapter 3 and introduces numerical taxonomy of urban tissues. The resulting classification is subsequently validated together with initial clusters based on the study of its relationship to the supplementary, non-morphological data known to be descriptive of urban form. Finally, the case study of Amsterdam is included to examine the ability of the proposed morphometric framework to identify urban tissues in the context of another heterogeneous historical city, although of different patterns of

development.

Structurally, this chapter is divided into two major parts. It first outlines the need for a classification of urban tissues (section 8.1) and bridges the concept of numerical taxonomy from chapter 3 to the context of morphometric tissue types (section 8.1.1). Furthermore, it explains the importance of validation and two validation models used within this study (section 8.1.2).

The two major parts are Methodological proposition (section 8.2) and Case studies (section 8.3). The structure of the methodological section is reflected in the structure of case studies, having each part of the proposed method tested in the relevant part of the associated case studies section.

Methodological propositions first introduce the method of hierarchical clustering, leading to the taxonomy of urban tissues (section 8.2.1). Moreover, the results are utilised in the next section 8.2.2 establishing the method of validation of identified urban tissues and their taxonomy, using data on historical origin, land-use patterns and qualitative classification of urban areas. Finally, the second part of the section 8.2.1 brings in another case study (Amsterdam) and presents the method of evaluation of transferability of the morphometric assessment. The evaluation focuses on the identification of urban tissue types, their hierarchical classification and combination of two geographically distinct datasets to a single taxonomy, examining the expandability of the method. Following section 8.3 presents the results of the proposed method applied to case studies in the same order and hierarchy of sections.

The final section (8.4) of the chapter summarises the findings and prepares the foundation for chapter 9, synthesising and discussing the whole research.

8.1 Classification and validation

The urban tissue types presented in the previous chapter are all seen as equal, meaning that there is no specified relationship between them at this stage. However, the degree of similarity between individual clusters varies and can be quantified. Therefore, a numerical representation of similarity, based on already measured morphometric characters, becomes

a basis for the further classification of urban tissues.

As illustrated in chapter 2, there have been attempts to deliver classification frameworks, some even taking hierarchical structure but still deviating from the *optimal classification model*. This section proposes the application of the biological model of classification in the form of a numerical taxonomy leading to the establishing hierarchical classification as a conceptual basis for the *atlas of urban form.*

However, the model of identification of urban tissues and their classification needs to be validated. There are two critical questions regarding the validity which needs to be answered:

1. Are morphometric clusters a valid proxy of urban tissue types?
2. Is the method transferable outside of the context of the initial case study?

To answer the first question, this chapter introduces a validation as an assessment of the relation of the tissue types and their taxonomy to additional data, which are known to be linked to the form of cities. The significant relationship between them and proposed clustering and therefore, the classification would indicate their validity and the validity of the proposed method. The second question requires the inclusion of another case study from a different geographical and historical context. Clusters identified in that case study should be internally valid and comparable with those identified in Prague.

8.1.1 Taxonomy of tissue types

Resulting clusters can be theoretically interpreted, in a conceptual sense, as populations and studied as such. Based on the propositions outlined in chapter 3, the biological analogy is taken further in this chapter, to propose a numerical taxonomy of urban form. The operational taxonomic unit (OTU) in this case is a morphometric urban tissue type. Each cluster as a whole is considered as a unit for the classification, conceptually mirroring classification on the level of populations, following the principles of numerical taxonomy (Sneath and Sokal, 1973) with only minor adaptations related to the specificity of urban form.

Urban form is traditionally classified not in a taxonomic, but in a typologic way (Kropf, 2017). Where taxonomy is based on quantitative relations between its elements, typology follows conceptual division (Bailey, 1994). Both approaches are valid as they assess the same entity (urban form) from a different perspective. Moreover, if the classification results coincide, the method can validate each other. This study aims to explore the potential of quantitative description of urban form. Hence it is natural that it chooses the path of taxonomy.

The taxonomy results in the hierarchical tree capturing the relationship of clusters, or *taxa*. That allows agglomeration of the lowest-level taxa into higher-level ones, adaptively changing the resolution of classification. That is especially helpful for studies assessing the effect of urban form on socio-economic aspects as it can adapt to coarser data.

8.1.2 Validation and applicability

Validation is an assessment of the correctness of a result. In the context of this work, validation should mainly focus on understanding whether the identification of urban tissue types works as intended and whether the method is transferable to other contexts and eventually extensible.

The validation of a clustering method which does not have a ground truth data is always indirect. That means that any validation procedure can give only an indication of the method's performance, not a precision estimate. To make the validation more robust, it is better to compare results to more than one validation layer. Furthermore, it is critical to ensure that the literature identifies that there is an expected relation between the concept validation layers capture and urban form patterns.

Transferability of the proposed method should be tested by its application on other case studies and validating the results independently of the first one. If the method produces reasonable results and validation via proxy layers indicates statistical relation, we can assume that the method is transferable to contexts in which it was tested.

Finally, expandability of hierarchical classification (i.e. taxonomy) is critical for the future development of a taxonomy of urban form. The method is extensible if the results of tissue type identification from one case can be successfully linked to tissue types from the other

case to build a common taxonomy.

The following section first proposes a method of hierarchical clustering to derive the basis for a taxonomy of urban form. Then it outlines methods of validation using the test of relation to proxy layers, transferability to a different context and finally, expandability as an attempt to merge two cases into a single classification.

8.2 Methodological proposition

Methodological propositions of this chapter focus on two distinct questions. The first one has been defined in chapter 5 as one of the supplementary research questions:

1. How to determine the taxonomic relationship between OTUs to derive taxa of urban form?

To answer that, this section proposes the application of hierarchical clustering method outlined below.

The second question is related to both results of the previous chapter 7 and hierarchical clustering proposed in this chapter:

2. Is the overall method valid for classification of urban form?

That is a question which is critical but at the same time hard to answer. Therefore, this section proposes a series of tests to get a reliable indication of what the answer could be.

8.2.1 Hierarchical clustering

Hierarchical clustering, in case of urban tissues, aims to develop a hierarchy of similarities between observations based on their morphometric profiles. We can generally distinguish two main principles, agglomerative and divisive. The former starts with the pool of observations, each with its own cluster and identifies pairs of clusters while moving up

the hierarchy. The latter does the opposite as it starts with a single cluster and iteratively divides it while moving down the hierarchy. Statistics offer a wide range of procedures for both principles, which description is out of the scope of this research. This research employs Ward's minimum variance hierarchical clustering, as the method with a long lineage in academic use (Singleton and Longley, 2009) and recent application in urban morphology (Dibble *et al.*, 2017; Serra *et al.*, 2018). Each tissue type is represented by its centroid (mean of each character) within the hyperspace and Ward's algorithm agglomeratively links observations together in a way, which minimises an increase in total within-cluster variance (Ward Jr, 1963). The classification has a form of a dendrogram capturing a cophenetic relationship between observations (i.e., morphometric similarity).

Resulting dendrogram can be further interpreted and initial OTUs flexibly clustered together based on the branching of the diagram. That can, in turn, be mapped, and the spatial distribution of branches can be visually assessed. Furthermore, branching enables focused analysis of individual macro clusters, if that is of interested in one's particular study.

8.2.2 Validation

Validation of identification of tissue types and consequent hierarchical clustering is done in two ways. The first one studies the relationship of resulting classification to additional non-morphometric data to verify whether the expected link between morphology and other aspects is present in the data. The second approach tests the applicability of the method outside of the initial case study. As the design and decision making behind it were based on the Prague dataset, it may have incurred context-specific features which limits the applicability of the method elsewhere. To ensure that this did not happen, the method should be applied to unrelated data, and resulting classification should be examined to ensure that the results are comparable.

8.2.2.1 Relation to non-morpohological data

Capturing the relation of proposed classification to additional data is an indirect validation method. There are theoretical grounds on which we can expect that the relation between

urban form and other data exist and hence should be present in the classification. It is well known that urban patterns change based on the era in which they are built, meaning that there is a significant relation between urban types and their historical origin (Panerai *et al.*, 2004; Dibble *et al.*, 2017). Similar relation could be found with land-use patterns (Castro *et al.*, 2019) and some other data.

However, such additional data should not be seen as ground truth for classification as it does not reflect the same concepts. The relation should be seen as indicative.

In this research, the proposed classification will be compared to three datasets - 1) the period of the historical origin of a place, 2) predominant land-use patterns, and 3) qualitative typology of urban form. All three will use the method of validation, based on cross-tabulation, using a) statistical analysis using chi-square statistic and related Cramér's V, further interpreted based on b) compositional analysis focusing on the composition of each cluster in relation to the tested data, and c) visual assessment of spatial distribution to illustrate the behaviour of both compared data in space.

8.2.2.1.1 Analytical tools Validation is using cross-tabulation (contingency table) as an input for all Chi-square test, Cramér's V and compositional analysis. Cross-tabulation measures the number of observations within each cluster-category pairs, where categories are reflecting the different classes of used proxy data (e.g. land use types).

The detailed method of application of selected analytical tools is proposed below.

8.2.2.1.1.1 Chi-square test of independence The proposed classification, as well as proposed additional data, are categorical variables, including the historical origin which is presented as unevenly distributed eras rather than age. The chi-square test of independence of variables determines whether there is a significant relationship between two categorical variables based on a contingency table (Agresti, 2018). The null hypothesis (H_0) and the alternative hypothesis (H_1) of the analysis states that

> H_0: Morphometric classification of urban tissues is independent of *variable*.
> H_1: Morphometric classification of urban tissues is not independent of *variable*.

The statistic itself is denoted as

$$(83)\ \sum \chi_{ij}^2 = \frac{(O-E)^2}{E},$$

where O is the actual observed count, E is the expected value, chi^2 is the cell Chi-square value. The expected value E is calculated as

$$(84)\ E = \frac{M_R \times M_C}{n},$$

where M_R is the row marginal (sum of the row) for the cell, M_C is the column marginal (sum of the column) for the cell and n is the total sample size (Agresti, 2018).

The chosen alpha level of significance is $\alpha = 0.01$.

The actual implementation is based on `scipy.stats.chi2_contingency` function from open-source toolkit SciPy (Jones *et al.*, 2001).

8.2.2.1.1.2 Cramér's V The chi-squared statistic does indicate whether there is an association or not but does not tell how strong it is. Cramér's V coefficient is based on chi-squared statistic but extends it to provide a value between 0 and 1, reflecting the level of association similarly as Pearson's correlation does. Value 0 corresponds to no association while 1 to perfect association (Crewson, 2006).

Cramér's V coefficient is denoted as

$$(85)\ V = \sqrt{\frac{\chi^2}{n(q-1)}}$$

where q is the smaller number of either rows or columns. The strength of association described by V is illustrated in table 8.1.

Table 8.1: Strength of association of two categorical variables based on Cramér's V coefficient. Reproduced from (Crewson, 2006).

V	association
>.5	high
.3 to .5	moderate
.1 to .3	low
0 to .1	little if any

8.2.2.1.1.3 Cross-tabulation compositional analysis Each of the clusters and each of the branches is then studied independently to understand what is its composition in relation to the validation data, focusing on individual rows of a contingency table. The perfect relation would show all observations of a single class within a single cluster and none within any other. An equal count would reflect no relation among the classes. The compositional analysis is aimed to provide more in-depth interpretative values that chi-square and Cramer's-V, but it does not state any significance level.

8.2.2.1.1.4 Visual assessment of spatial distribution The visual assessment of spatial distribution overlays the boundaries defined by morphometric classification over the validation classes to determine spatial relationship visually. Alongside the compositional analysis, the visual assessment is meant to provide interpretative information, allowing a better understanding of the relation between tested data. It does not provide any numerical results as it only links compositional analysis with the geographical context.

8.2.2.1.2 Validation data Three datasets are used within the validation framework - historical origin, land use patterns and qualitative typology of urban form linked to the predefined boundaries.

8.2.2.1.2.1 Historical origin The link between the historical origin and the patterns of urban form is well established in the literature (Panerai *et al.*, 2004) and has been recently studied using morphometric tools, from the composition and configuration of street

networks (Porta *et al.*, 2014; Boeing, 2020a) to the complex morphometric assessment of sanctuary areas (Dibble *et al.*, 2017). The relation, which is described in the literature, should be present in the data studied in this research. The existence of the significant relationship between the historical origin and results of morphometric classification would indicate the validity of the proposed classification.

The data on historical origin provided by the Institute for Planning and Development Prague denotes the time frame in which was each part of the city first built-up. However, the data do not provide a single year, but a specific range, presenting what would be a continuous variable of age as ordered categorical one. Moreover, the categories are not equally distributed in time, with maximums within each category being 1840, 1880, 1920, 1950, 1970, 1990, and 2012. That may lead to the situation where three adjacent buildings, built in years 1878, 1879 and 1881 are not seen as of the (almost) same age. The first two are in the second category, being treated as equal while the last is in the third category being treated as different. Moreover, its difference from 1841 is the same as the difference from 1949.

These data ignore the redevelopment of parts of the city which happened later. Newly built areas, which are built in the area of previously demolished urban form are not reflected and should be interpreted accordingly.

However, even with these limitations, the dataset does represent the different periods of Prague's development, and there should be a significant relation.

Before doing the analysis itself, data on historic origin were spatially linked to the building layer. Each building got assigned a single category of origin denoting not its own period of origin, but the first moment of the development of the area it sits in.

8.2.2.1.2.2 Land use patterns Land use is determining building typology, which is partially reflected in the patterns of urban form. Single-family housing is always different from industrial or commercial areas, while multi-family housing can be developed in a plethora of ways, but still different from other uses. That gives us theoretical grounds for validation of proposed morphometric taxonomy using land-use patterns. However, it has to be noted that such a relation will likely not be perfect, as there are mixed uses in many places.

The data on land use in Prague capture land use to the level of individual building and plot and divides it into 123 categories. However, only 15 of them contain more than 1,000 buildings. Categories are providing detailed classification, but that does not reflect the predominant tendency of land use within the area but the individual buildings. For that reason, the initial data are used to compute predominant land-use patterns within three topological steps on morphological tessellation. As predominant land-use is seen the one with the highest frequency within the *context.*

Initial land-use data are spatially linked to buildings layers to have a single value representing category per building. Then the predominant land-use is calculated based on the context. Out of resulting categories, only 5 (*Multi-family housing, Single-family housing, Villas, Industry small, Industry large*) contain more than 1% of the dataset. For that reason, these five are used, and the rest is denoted as *Other*.

8.2.2.1.2.3 Municipal typology Planning system of Prague is based on the concept of *localities*, small neighbourhoods (Institut plánování a rozvoje hlavního města Prahy, 2018). Each neighbourhood has specified boundaries partially based on its morphology and partially on other aspects, from historical origin to social perception of the area. Furthermore, these neighbourhoods were qualitatively classified into one of the 10 *structural types*. This municipal typology tends to capture morphology and as such, could be used as a validation method. In the ideal world, this layer would become a ground truth for the morphometric classification. However, that is not possible due to methodological flaws embedded in the typology.

The typology consists of the following 10 types (loosely translated into English):

- organic structure
- perimeter block structure
- hybrid structure
- heterogenous structure
- village structure
- garden city structure
- modernist structure
- production area

- services area
- linear structure

While these types might work for planning purposes, they are conceptually incoherent mixing types based on morphology (organic, perimeter block) with those based on planning ideology (garden city, modernist), or those based on land use (production, services). Moreover, the fixation of the typology to *localities* comes with Modifiable Aerial Unit Problem (Openshaw, 1984), leading to the inclusion of ambiguous loosely defined hybrid and heterogeneous types.

However, the typology itself, considering above mentioned limitations, reflects what the planning authority thinks Prague is composed of, and it is worth studying the relation of this qualitative typology to the proposed quantitative classification. The only adaptation which needs to be done is the exclusion of hybrid and heterogenous types from the analysis due to their MAUP-based origin and of linear structure capturing railway structures only.

The data provided as polygons represented localities are spatially joined to buildings layer, and features containing excluded types are removed from the data.

8.2.2.2 Transferability of the method

The proposed method of identification of urban tissues is validated in the context of Prague using the methods above. However, that by itself does not ensure that the method is transferable and applicable elsewhere. Different geographical contexts, bringing various types of urban tissues, and their underlying spatial logic may be challenging for a method tested in a single, no matter how heterogenous, case. The transferability of the method is a critical feature for its robustness and applicability. The method should show similar performance, in terms of identification of tissue types and a consequent taxonomy, in additional case studies.

Therefore, the method as it stands is tested on the case of Amsterdam, NL. Both Amsterdam and Prague are heterogeneous cities with several historical layers, but of different planning context during their respective developments.

8.2.2.2.1 Urban tissues of Amsterdam The first part of the analysis of Amsterdam is the identification of urban tissue types using the method proposed in chapter 7. Precisely the same set of primary and contextual characters is used within Gaussian Mixture Model clustering and related BIC analysis of the number of components. For the details of the method, refer to the previous chapter. Similarly, the method of hierarchical clustering proposed in section 8.2.1 is applied to the resulting tissue types.

Both results will be assessed visually on a map to understand whether the clusters alone and within their branches are interpretable and contiguous.

8.2.2.2.2 Validation of clustering in Amsterdam The resulting clusters in Amsterdam are validated using the method proposed in section 8.2.2.1 above, using data on the historical origin of each building. In the case of the datasets obtained from Dukai (2020), each building has assigned a year of its construction. Unlike in Prague, the year does not represent the data when the area/plot was first built-up, but the latest construction. Even though the data are not initially binned, only buildings constructed after the year 1800 have a specific year. To ensure the compatibility of the data with those used in Prague and to avoid issues with pre-1800 periods, the origin dates are therefore binned into 11 groups following the classification of Spaan and Waag Society (2015). The rest of the validation follows the method outlined in 8.2.2.1.

8.2.2.3 Expandability of the classification

The study of expandability of hierarchical classification is the last methodological step in the whole thesis, and its role is to understand whether results of the morphometric study from one case study can be related to another case study. Expandability of the classification is crucial for further expansion of the database of urban tissue types. Even though clusters and hierarchical classification may work in individual cases, the question is whether we can combine the results to a single taxonomy. It tests the compatibility of results and a potential issue of clustering being tied to a single context. The optimal situation would mirror the biological world, where a newly discovered species can be usually embedded into an existing taxonomy. However, to get to the stabilised situation where taxonomy is not substantially changed by the discovery of a new species, we first

need a critical mass of species to be included. That is certainly not the case in newly built taxonomy of urban form, and it is expected that it could be relatively unstable in the beginning and stabilise by the inclusion of more cases.

The method can be considered extensible if the taxonomy of tissue types from Prague and Amsterdam combined does not substantially change the interpretative value of dendrograms.

The generation of the combined taxonomy is a straightforward process. All clusters from both cases are combined into a single pool and used as an input of Ward's hierarchical clustering. The resulting dendrogram is then compared to the initial individual dendrograms, and their structures are compared. That reflects whether resulting branches capture similar tissues in both contexts. The results are then visually assessed using branches mapped on to the urban form of both cities alongside.

The final step is an interpretative analysis of the reshuffle of clusters between individual dendrograms and a combined one. In the ideal case, tissue types which are in a single branch in an individual tree should stay within a single branch in a combined one. However, as mentioned above, it is expected that some degree of reshuffle may happen when a classification structure is not yet saturated.

The following section applies the methods and presents their results.

8.3 Case studies - continuation of Prague, Amsterdam

This section presents the results of the methods proposed in the previous one. It starts with the continuation of the work on Prague presented in the previous chapter, developing hierarchical clustering of initially identified urban tissue types. Further, it presents the indirect validation of the results using non-morphological data. The second part presents an inclusion of Amsterdam case study, as the delineation of tissue types, taxonomy and validation using historical origin. The final section discusses expandability of the proposed method testing its ability to form a methodological foundation of a general atlas of urban form.

8.3.1 Hierarchical clustering

The centroid values of each cluster, obtained as a mean value of each morphometric character, are used taxonomic characters within Ward's hierarchical clustering. The resulting relationship between centroids, representing the relationship between identified urban tissues, is illustrated on the dendrogram on figure 8.1. The horizontal axis represents each individual cluster, while the vertical axis captures the cophenetic distance, i.e. the similarity between observations. The lower the connection between two branches is, the more similar the tissues represented by these branches tend to be. The values under each connection represent the actual cophenetic distance of a connection and number of observations which belong to the link. The different branches of the tree are coloured to ease the interpretation of the tree itself and to provide the visual link between the dendrogram and the resulting spatial distribution of branches.

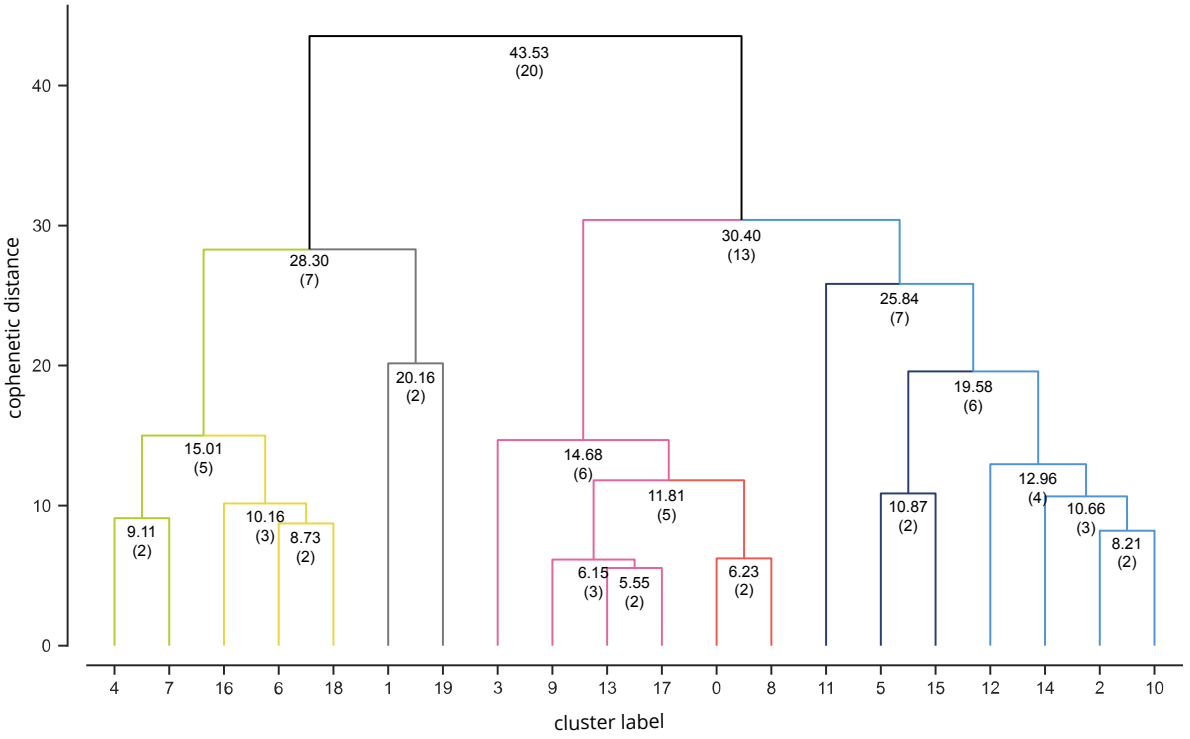


Figure 8.1: Dendrogram representing the results of Ward's hierarchical clustering or urban tissue types in Prague. The y-axis shows a cophenetic distance between individual clusters, i.e. their morphometric similarity. Branches are interpretative coloured - the colours are then used on maps illustrating spatial distribution of these branches.

The dendrogram shows several major bifurcations on different levels of cophenetic distance, indicating several distinct groups of urban tissues. However, the exploration and interpretation of each branch require the projection of the results into the geographical space. To allow that, each cluster is coloured according to the branch of the dendrogram it belongs to, using different lightness of the same hue to distinguish between individual clusters. The spatial distribution of hierarchically represented cluster in the whole Prague is illustrated on the figure 8.2, and the detail of the city centre is shown on the figure 8.3.

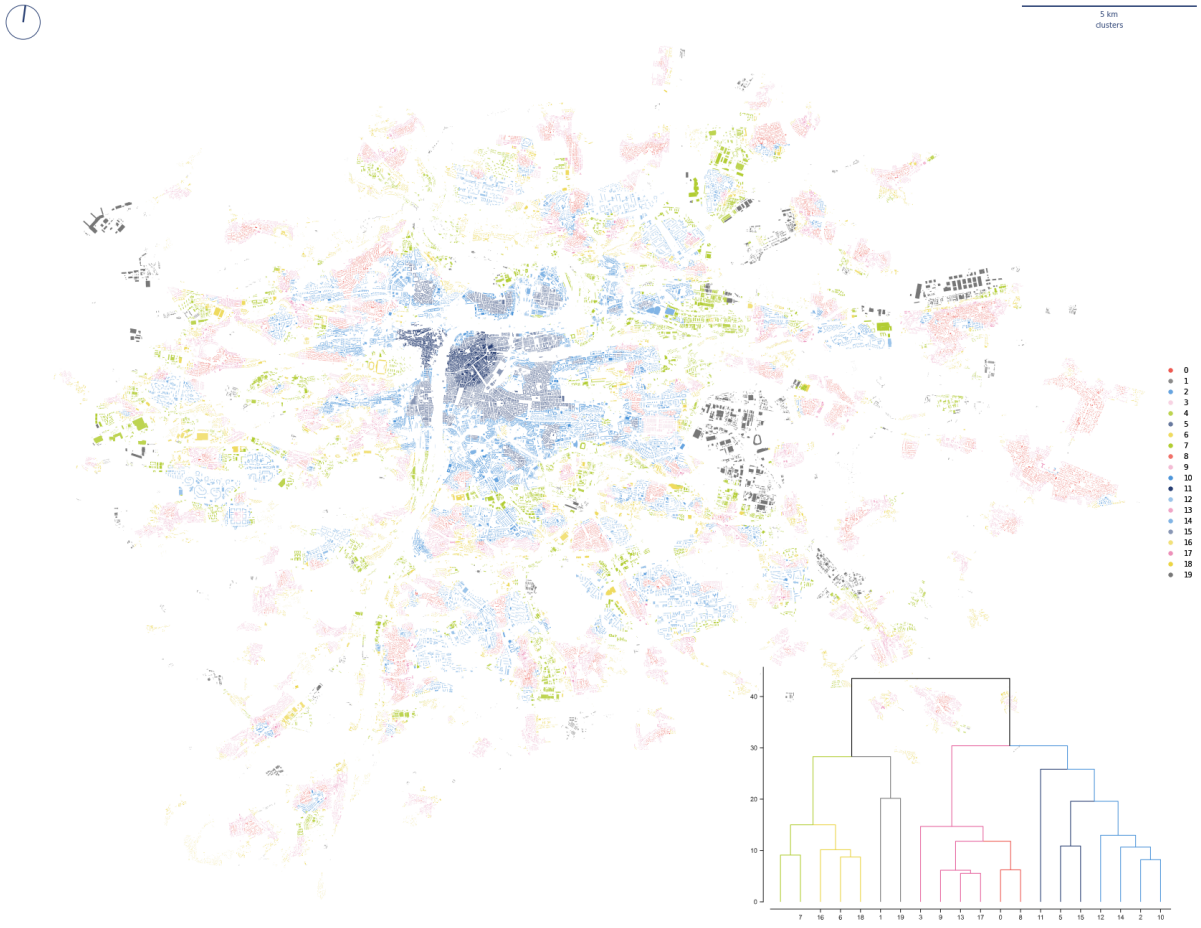


Figure 8.2: Spatial distribution of different branches of the dendrogram. Each tissue type is coloured according to a branch it belongs to, with a minor differences in colour intensity to allow for distinguishing of individual clusters.

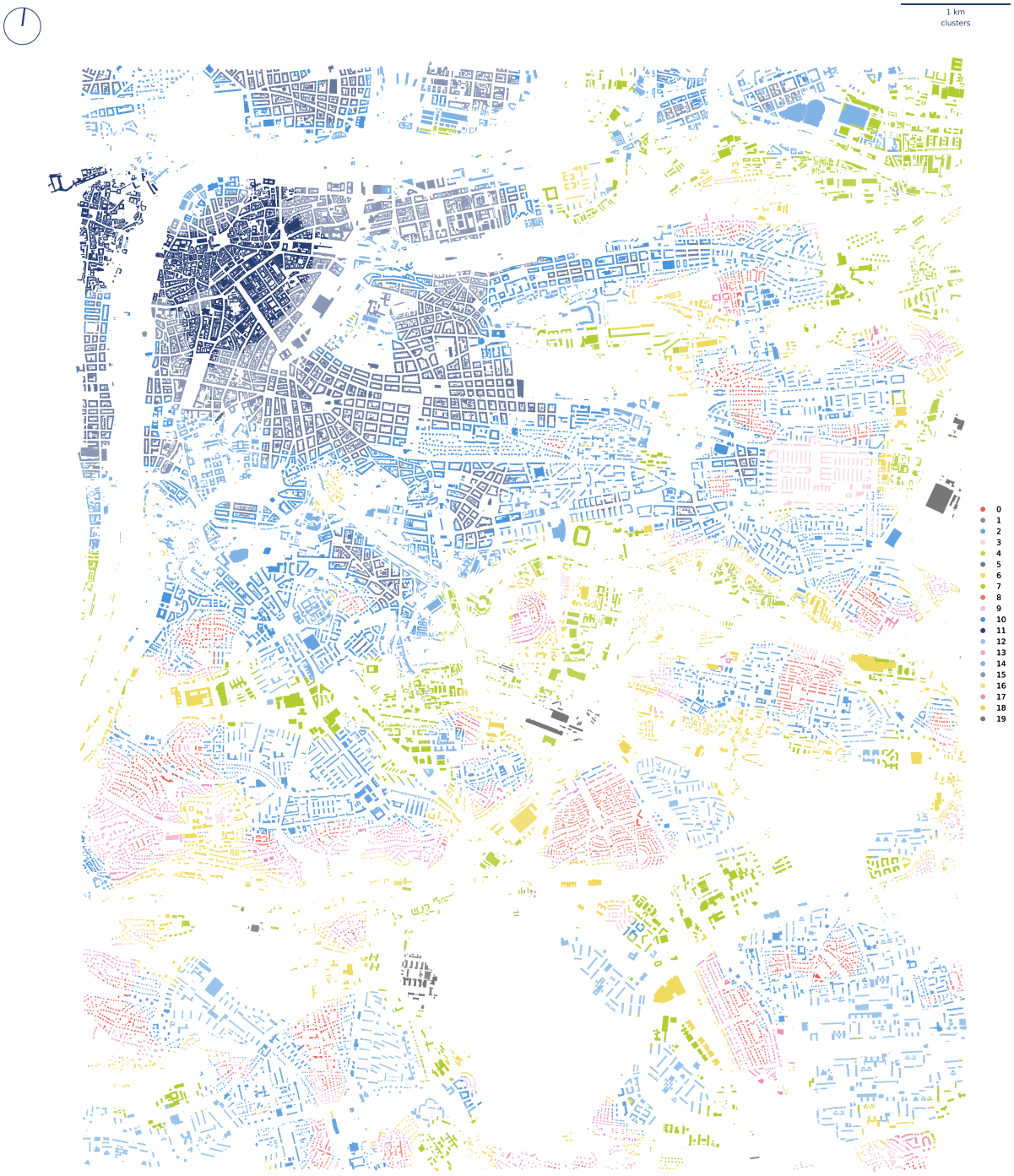


Figure 8.3: Spatial distribution of different branches of dendrogram zoomed to the central area of Prague. Each tissue type is coloured according to a branch it belongs to, with minor differences in colour intensity to allow for distinguishing of individual clusters.

Examining the dendrogram, we can highlight the different branches to understand their spatial distribution. Starting from the top of the dendrogram, from the bifurcation with the higher cophenetic distance (43.53), we can divide Prague's urban form into two major taxa. The right side of the tree represent urban form we could call *organised city* and is illustrated in figure 8.4. It consists of areas of mixed origin, spanning from the historical core to modernist and contemporary developments. The common characteristic is predominantly residential nature of all tissues.

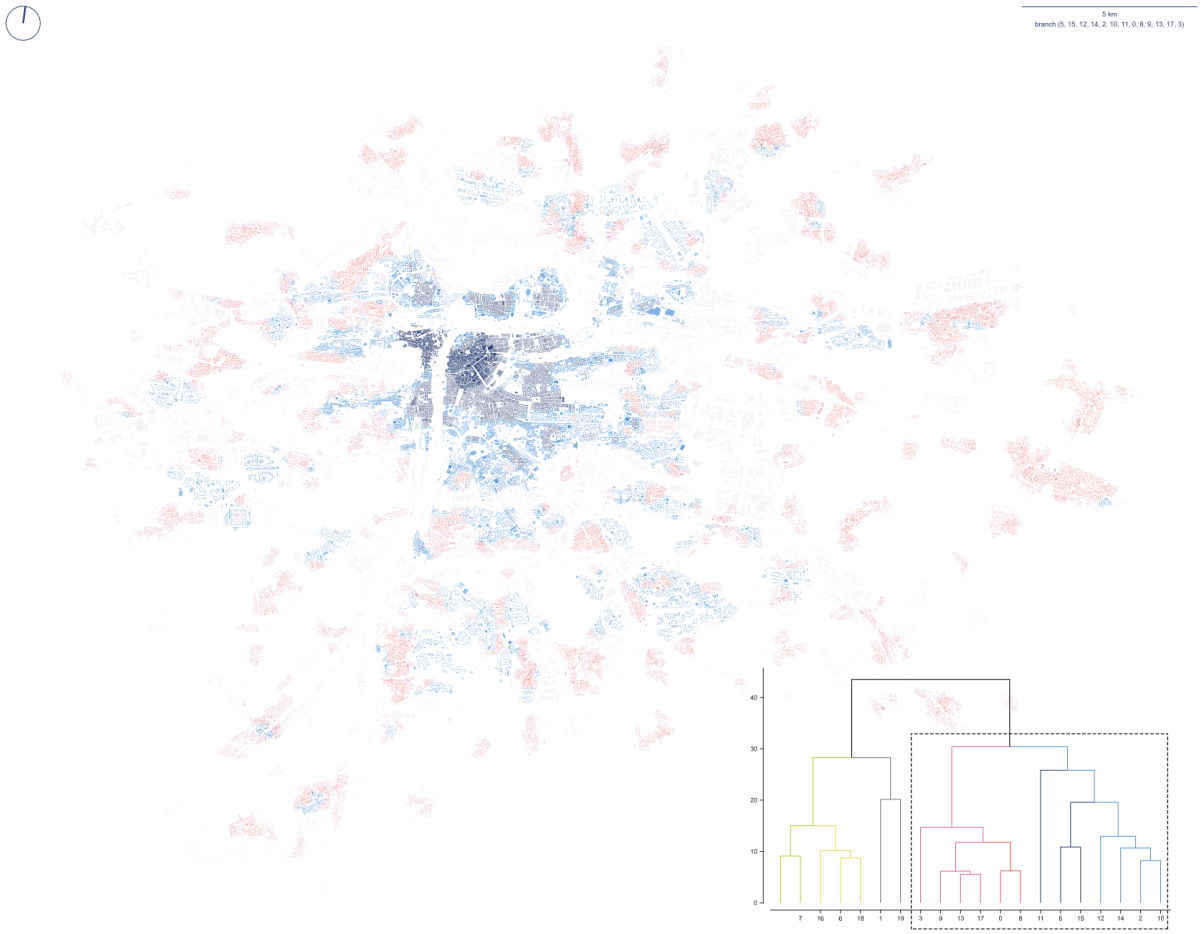


Figure 8.4: Spatial distribution of clusters within a branch representing organised city. We can see a relative contiguity in the city centre but scattered discontinuous areas in the periphery.

On the other side lies an *unorganised city*. It contains both industrial and fringe areas as well as contemporary office parks. The apparent shared logic is a relative disorder of patterns and high heterogeneity.

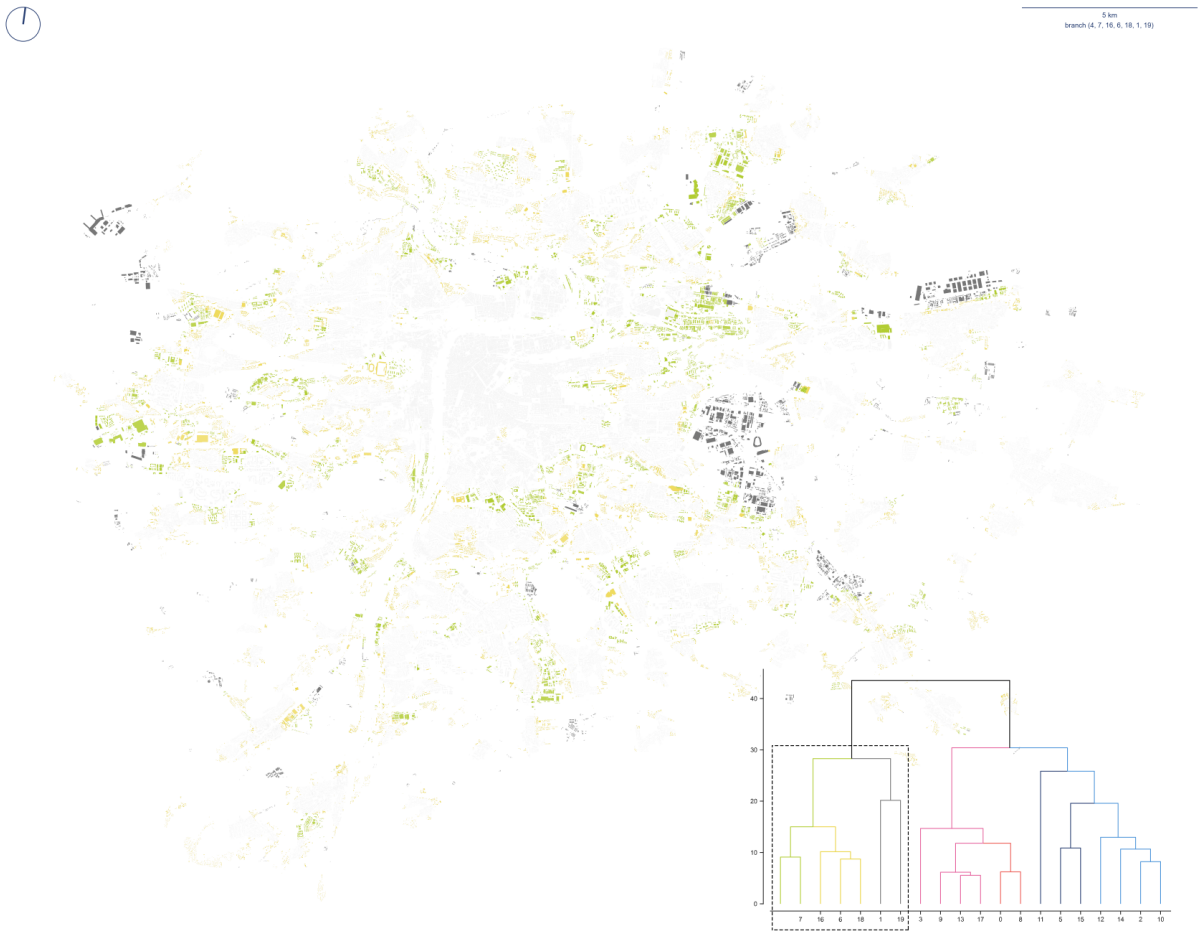


Figure 8.5: Spatial distribution of clusters within a branch representing unorganised city, located mostly on the outer ring of the city.

Going deeper into the right side of the dendrogram, we reach another major bifurcation happening at a distance 30.40 dividing the branch into two, representing mostly the intensity of the development. Left side, illustrated on figure 8.6 captures urban tissues which could be characterised by single-family housing of all sorts, spanning from villages to garden city-like neighbourhoods. To illustrate the density, the mean floor area ratio of the branch is 0.38.

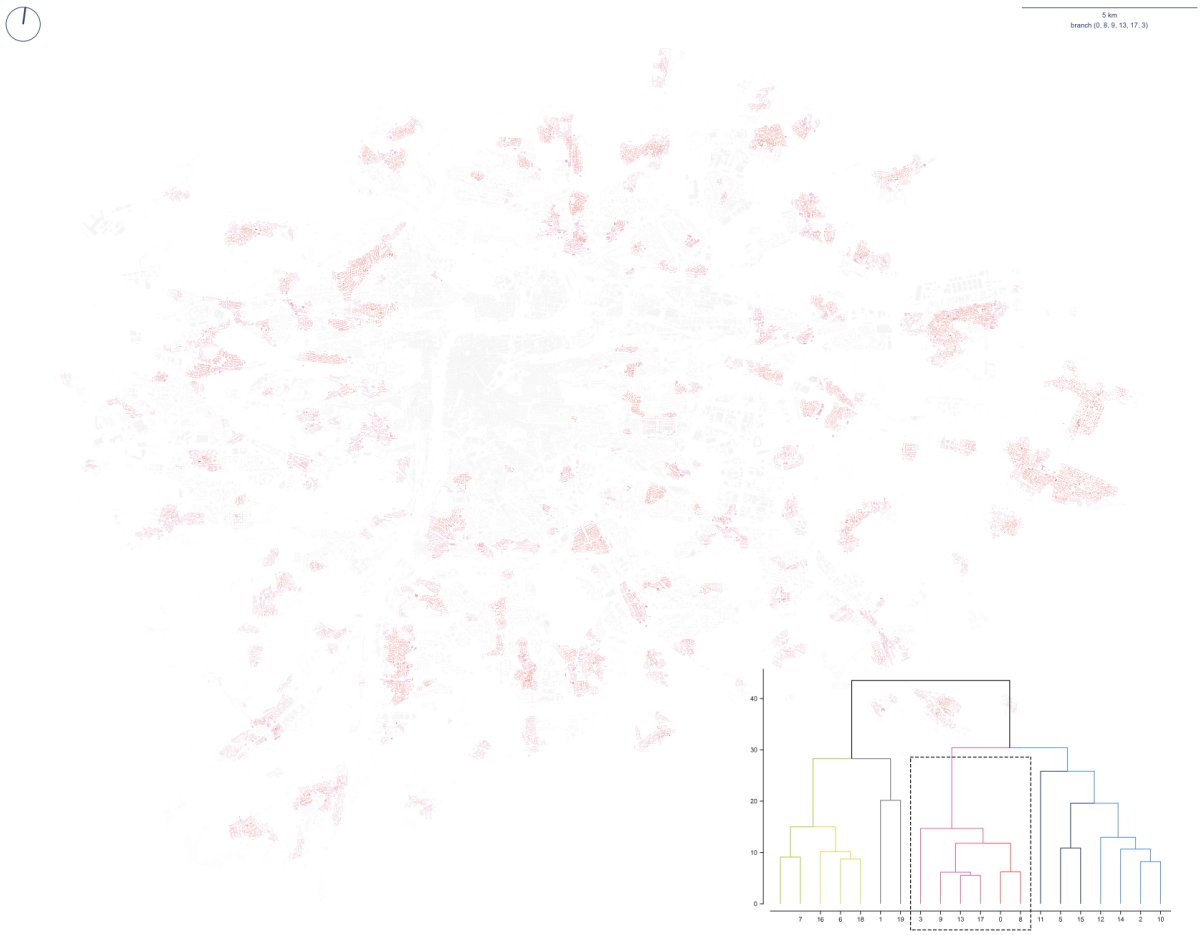


Figure 8.6: Spatial distribution of clusters within a branch representing low density, organised development.

The other side of this branch represents the *dense city* of all sorts (figure 8.7), from the medieval core, through historical perimeter block tissues to modernist housing estates. All these share the high volumetric density (mean floor area ratio of the branch is 1.76).

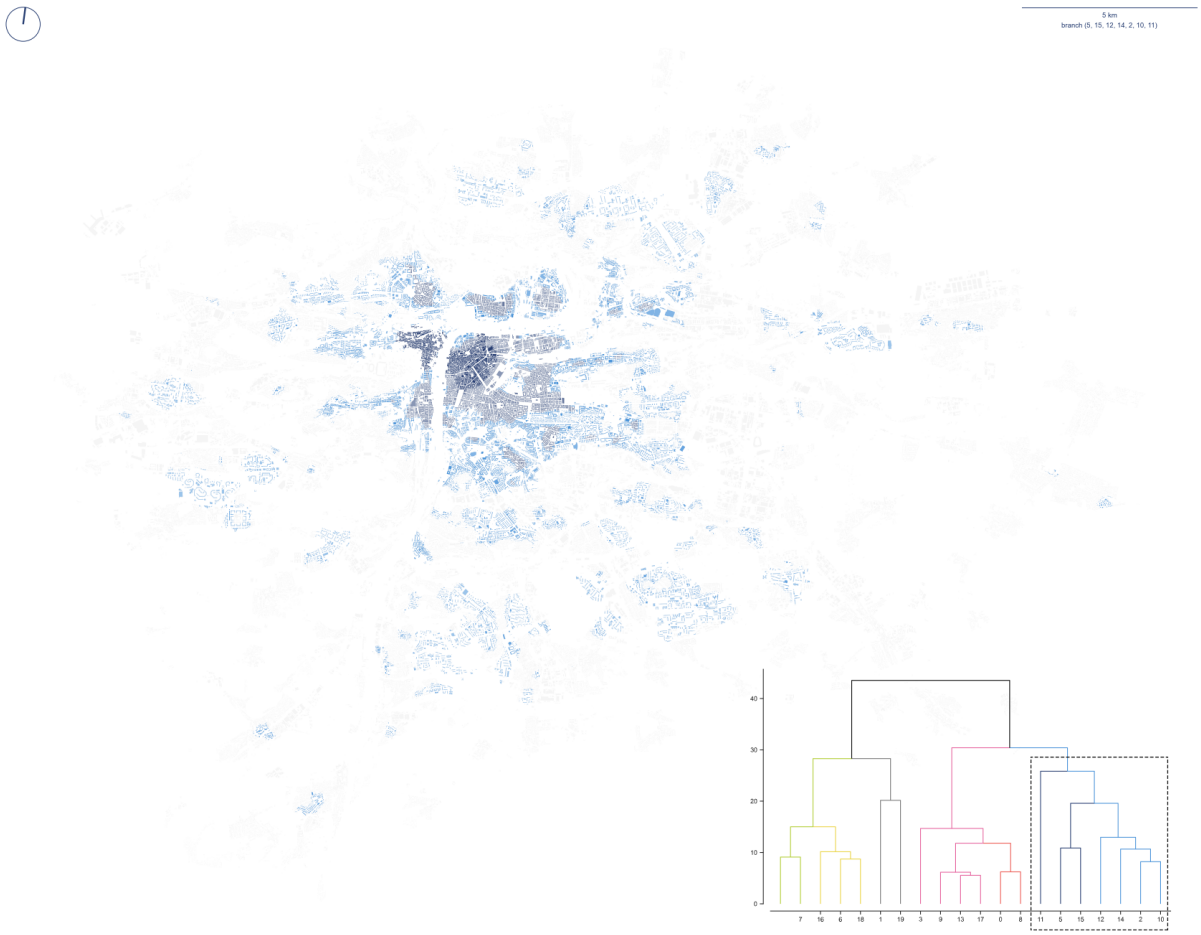


Figure 8.7: Spatial distribution of clusters within a branch representing high density, organised development. It is capturing two distinct phases of development of Prague - the historical period in the central areas and modernist development (mostly) in the periphery.

Further reading of the branching shows the bifurcation dividing the medieval city from the rest (c.d. 25.84) and further one splitting *ordered grid-like city* from disordered one (figure 8.8), composed of fringe areas adjacent to *ordered city* and modernist housing estates.

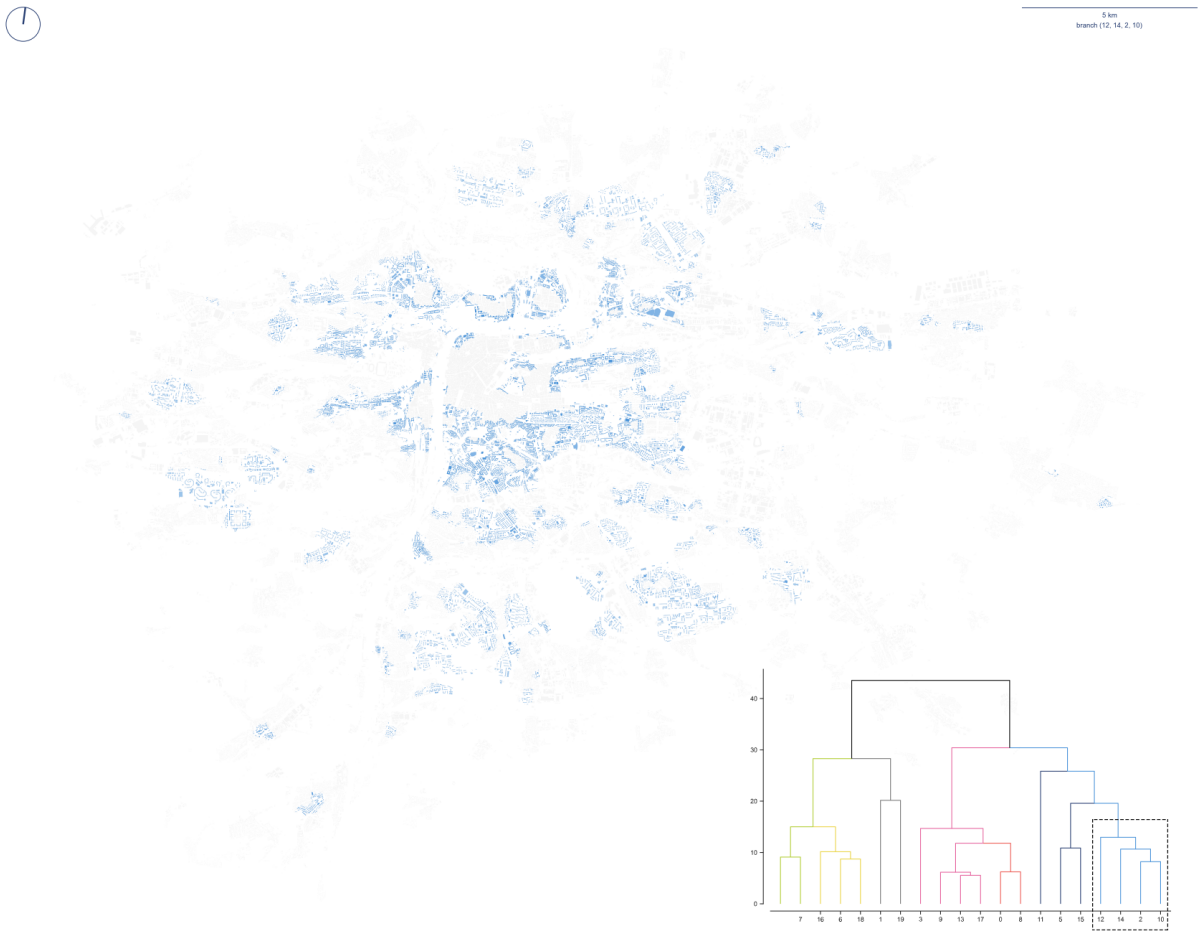


Figure 8.8: Spatial distribution of clusters within a branch representing high density, organised development with a fringe-like disorder in their patterns.

This basic description of branching shows that the top-level structure of the taxonomy reflects the spatial logic of the structure of Prague to a high degree of interpretability. As the key aim of this study is not to provide deep insight into Prague's structure but to develop the method itself, this section does not go into further detail. However, the rest of the maps of individual branches is available as an Appendix 8.1.

8.3.2 Validation

Validation results are divided into several sections. The first one focuses on the relation of delineation of urban tissue types and their taxonomic branches, and additional data, which should have the ability to reflect the differences in built-up patterns. That entails historical origin, land use patterns and municipal typology of urban form, all in the context of Prague. The second section outlines the results of cluster analysis and subsequent hierarchical classification of Amsterdam. To validate the Amsterdam case itself, local tissue types are assessed against data of historical origin. The third and last section focuses on results of combined taxonomy and potential of the expandability of the method and potential future development of the taxonomy of urban form.

8.3.2.1 Relation to additional data

Relation to additional data reflects the results of clustering and hierarchical classification compared to the historical origin, land use patterns and qualitative municipal typology of urban form. All these are assessed based on the contingency table and visual assessment of spatial distributions. The contingency table is used to calculate Chi-square statistic and Cramer's V, and to examine the composition of each cluster.

8.3.2.1.1 Origin Data for historical origin are illustrated in figure 8.9. There are some significant patterns which should be reflected in the clustering, notably historical core and modernist belt. However, not all differences in origin have their counterparts in differences in clustering, as there are patterns which are consistently built across time frames (e.g., low-density single-family housing).

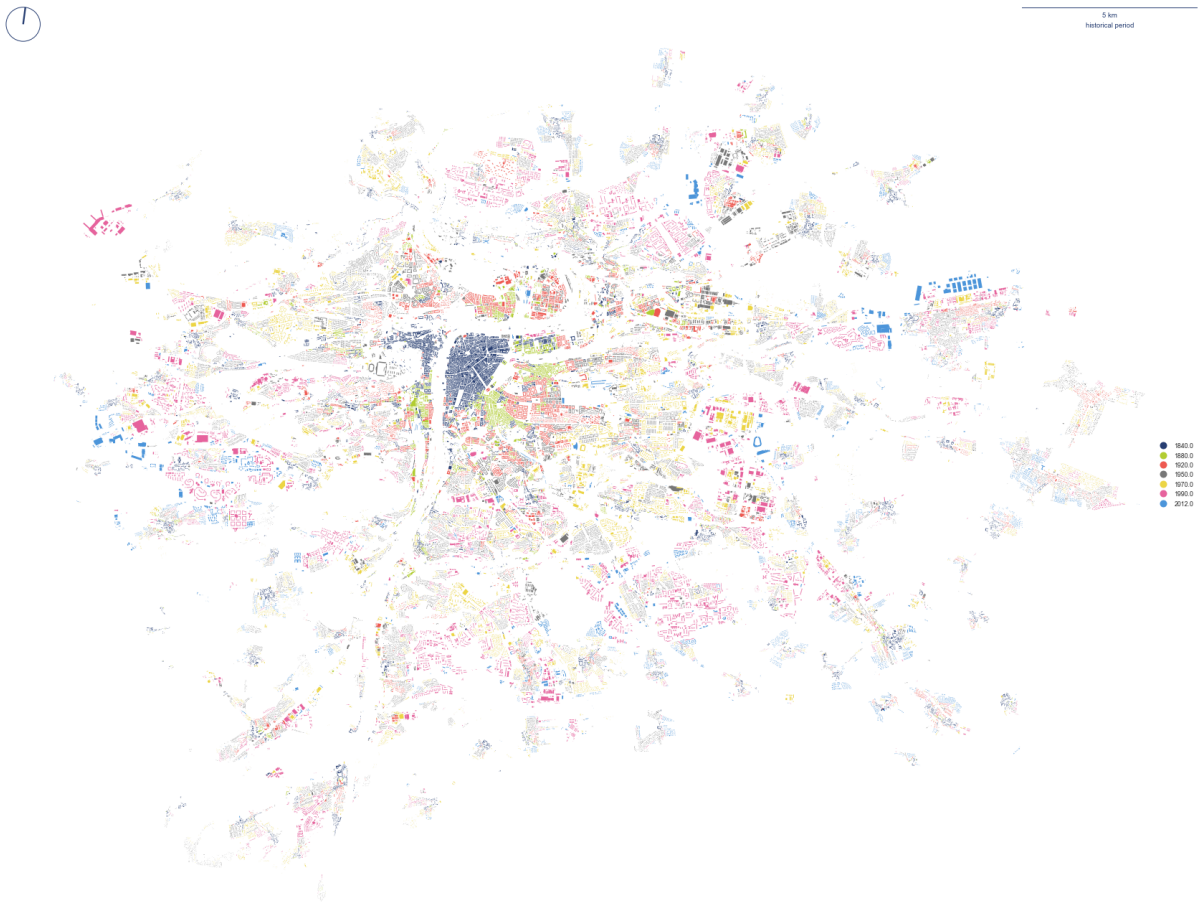


Figure 8.9: Spatial distribution of different periods of historical origin.

Contingency table 8.2 shows the distribution of buildings within clusters and time periods. Obviously, there is a relation, especially when it comes to the larger historical grouping into pre-WW2 and post-WW2 macro groups.

Table 8.2: Contingency table showing the counts of features per historical origin within individual clusters.

cluster	1840	1880	1920	1950	1970	1990	2012
0	251	110	2004	9310	2602	542	501
1	171	41	90	425	523	431	233
2	461	241	952	4585	3668	1816	224
3	6	16	27	1625	439	1217	749
4	302	208	728	1332	1359	842	299

cluster	1840	1880	1920	1950	1970	1990	2012
5	752	1707	2522	926	11	2	1
6	1146	406	763	3193	2040	1870	729
7	244	150	281	834	639	1188	594
8	107	255	923	5125	636	287	509
9	3048	516	912	1766	581	421	573
10	825	1284	2630	2034	279	66	52
11	2097	9	17	26	0	0	1
12	42	7	85	255	919	5220	290
13	1028	234	1021	6227	2611	1284	2472
14	868	656	1179	1466	345	139	189
15	1514	880	468	171	5	4	5
16	417	214	337	1002	601	571	288
17	740	354	1298	6229	1959	925	576
18	1346	204	544	2887	1686	719	1178
19	1	0	20	356	412	620	198

Reported results of a Chi-square test, based on the contingency table, assessing whether there is a significant relationship between two variables (origin and clustering), are $\chi^2(104, N = 140315) = 106700.51, p < .001$, which indicates a significant relationship. Cramérs-V assessing the strength of the relationship is 0.358, indicating moderate association. Since we cannot expect complete match because the data as the theoretical relation is only partial, both results are indicating the high performance of cluster analysis.

Looking at the composition of each cluster extracted from the contingency table (figure 8.10), we can see the relationship in a more interpretative way. Where the relationship is expected based on literature, there is a clear association (historical core, modernism) with the majority even falling into a single category. The striking difference between the structure of pre-WW2 and post-WW2 urban patterns partially confirms previous results reported by Dibble *et al.* (2017) and Porta *et al.* (2014).

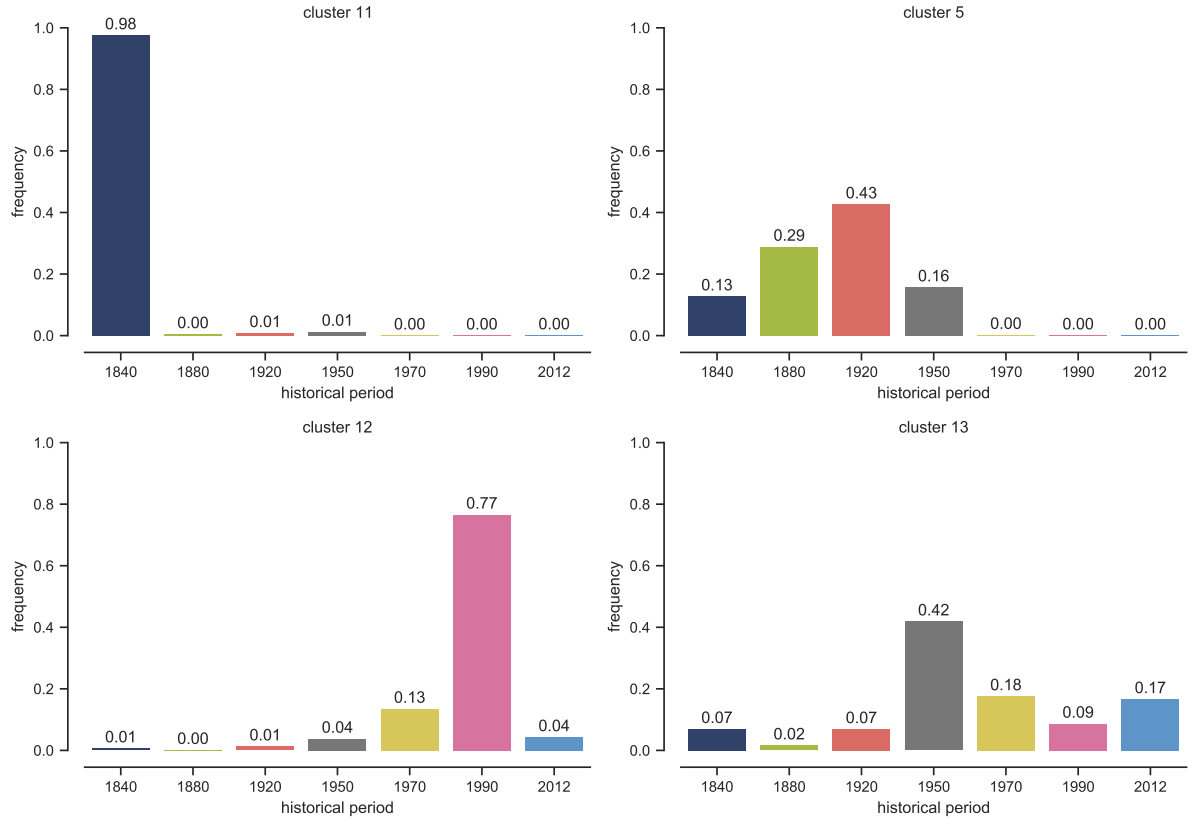


Figure 8.10: Illustration of composition of selected representative clusters from the perspective of historical origin.

Branches of the local taxonomy show similar relation (where is feasible to expect one), especially regarding the tendency of significant changes of development patterns after the Second World War (figure 8.11). However, it is important to note that unlike other European cities, Prague was not significantly damaged during the Second World War and the difference cannot be then interpreted as post-war regeneration and redevelopment.

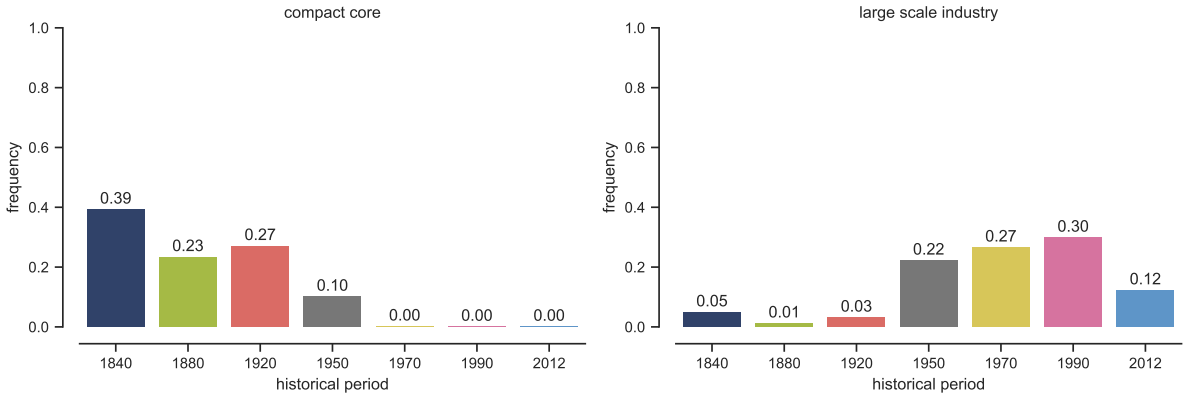


Figure 8.11: Illustration of composition of selected representative branches from the perspective of historical origin.

Spatial distribution tells us what do all these statistical values mean on the ground and how good is the coincidence there. The historical core of Prague in figure 8.12 captures the relation if cluster 11 to the historical origin. However, due to the limitation in data grouping, all development build pre-1840 into a single category, some differences need to be explained. Jewish quarter, the pre-1840 area but not part of the cluster on the north of the centre, has been demolished and rebuilt in early 1800. Therefore it is not of medieval origin and is correctly excluded from the cluster.

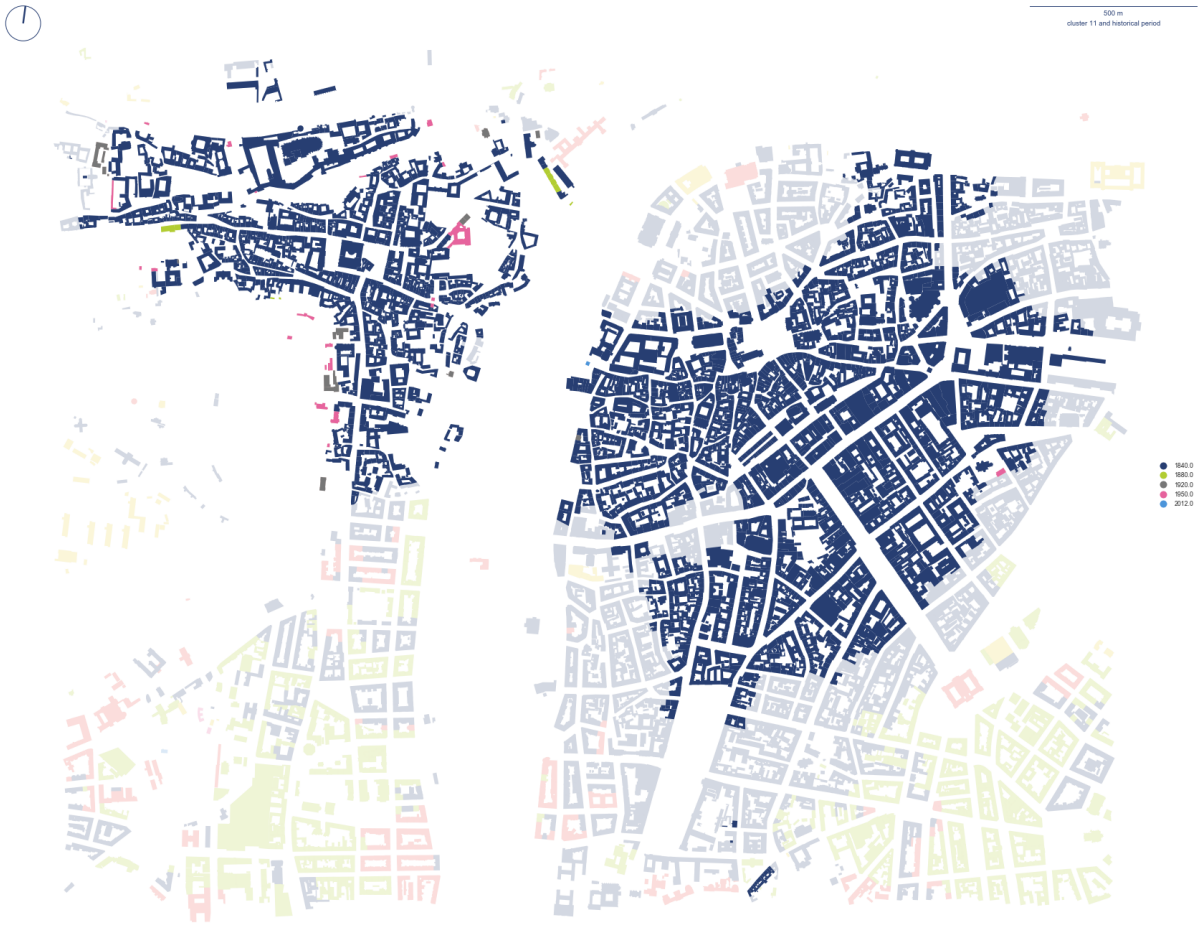


Figure 8.12: Illustration of the overlap between cluster 11 representing historical core and periods of historical origin. Colours represent period and the saturation the extent of the cluster.

Cluster 5, the one we can call a compact dense city was built according to the same principle over the years, but generally tends to coincide with relevant origin categories as shown on the figure 8.13. Notice the split in historical origin in the central part into 1880 and 1920 categories. Similar split has been identified by sub-clustering in section 7.2.3.4.1 shown in figure 7.59. However, there can be two reasons for the similarity - one is the historical period and the other topography, as the areas shown in green are mostly built on slopes.

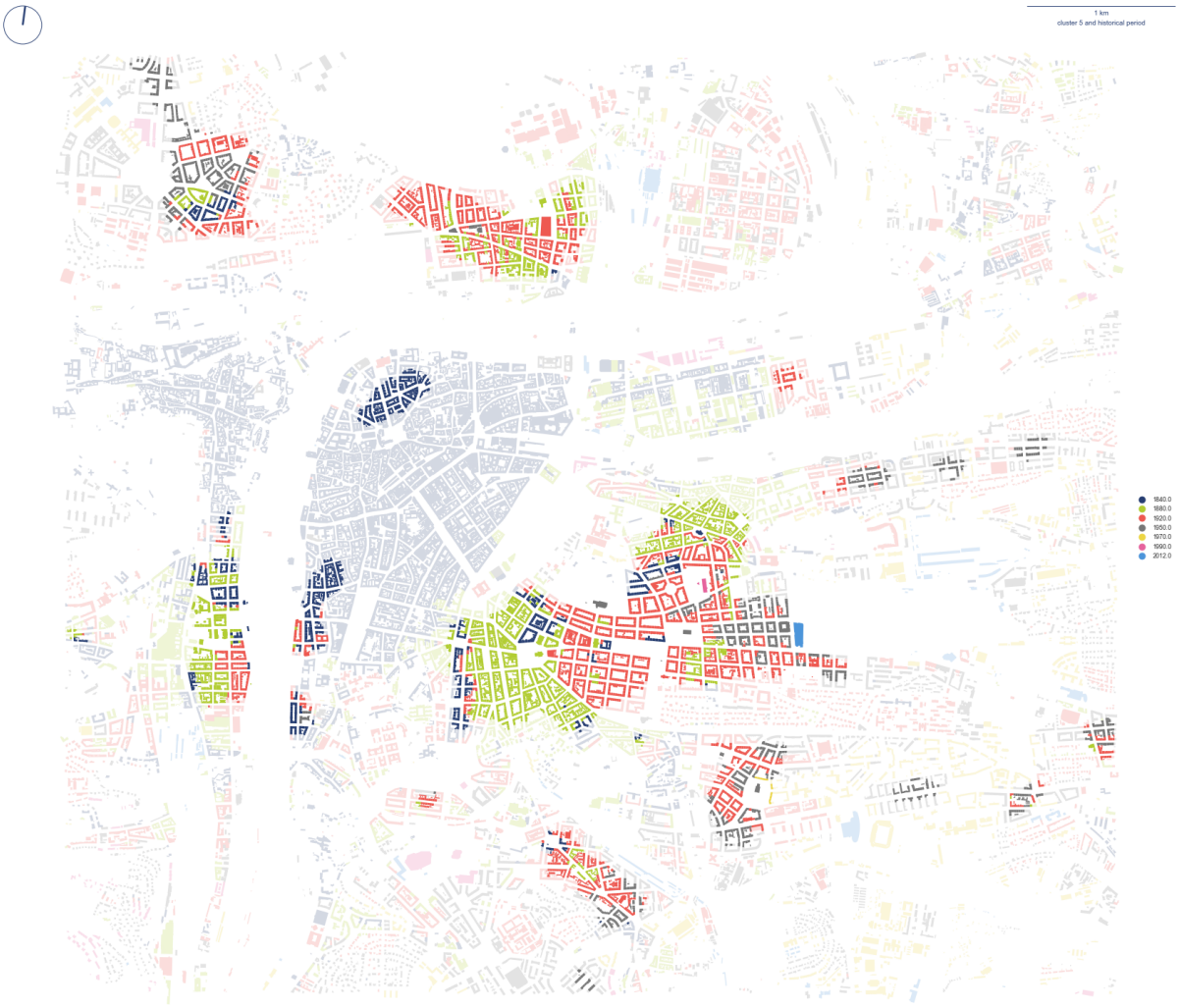


Figure 8.13: Illustration of the overlap between cluster 5 representing historical compact development and periods of historical origin. Colours represent period and the saturation the extent of the cluster.

A major part of the modernist housing belt around the city (cluster 12) was built between the 60s and early 90s, which is very nicely shown by the overlaps of distributions as well as on the figure 8.14 below.

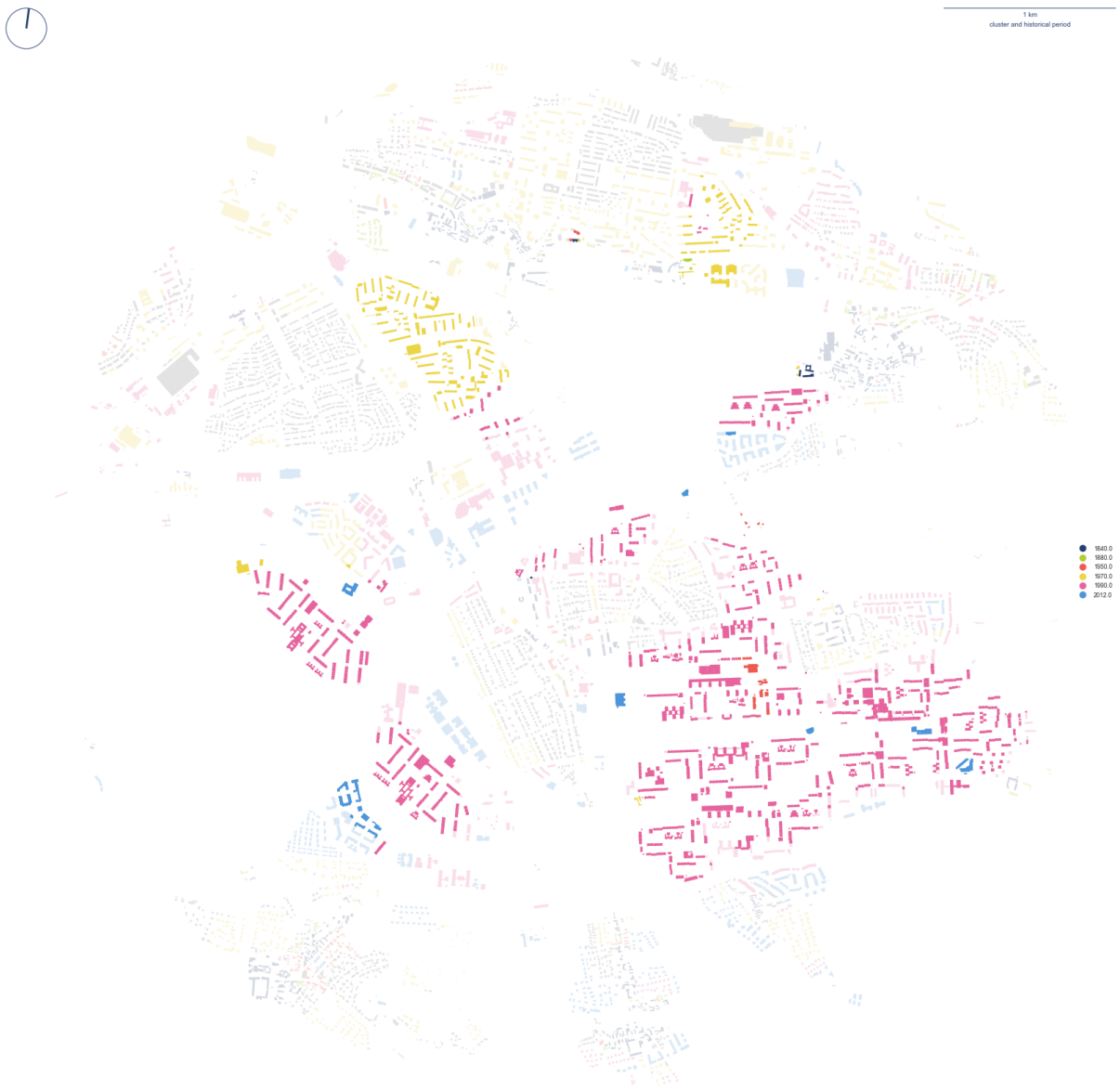


Figure 8.14: Illustration of the overlap between section of cluster 12 representing modernist development and periods of historical origin. Colours represent period and the saturation the extent of the cluster.

Looking at the combination of different branches, if we focus on a sample of the ordered part of a dense city branch, we see almost perfect overlap with pre-1950s development.

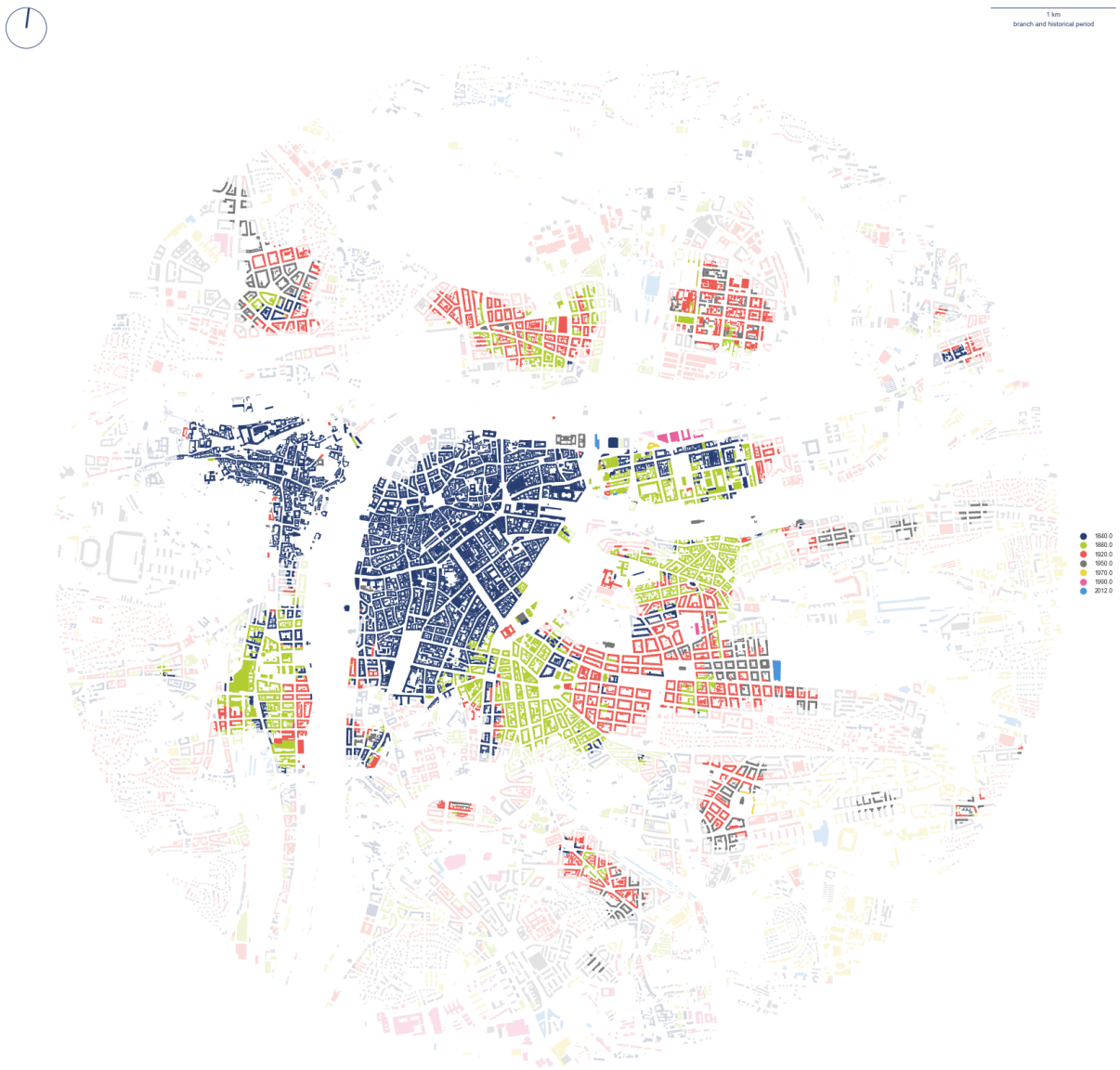


Figure 8.15: Illustration of the overlap between branch representing dense compact development and periods of historical origin. Colours represent period and the saturation the extent of the branch.

As far as it is possible to link these two aspects of form, origin and patterns, there is a significant connection validating the results of clustering and classification.

8.3.2.1.2 Land use Patterns of predominant land use are illustrated in figure 8.16. The large areas of the city are covered by single-family housing and multi-family housing. The latter spans from historical development in central areas to large developments during the second half of the 20th century, covering both dense compact development and modernist housing typology. There are only minor areas denoted as villas in the obtained land-use classification, mostly located in the north-west of the city. Moreover, the patches of the industry are often mixed with *other* land use categories, which indicates that the need to be cautious during the interpretation of results as this reflects the limitation of the data.

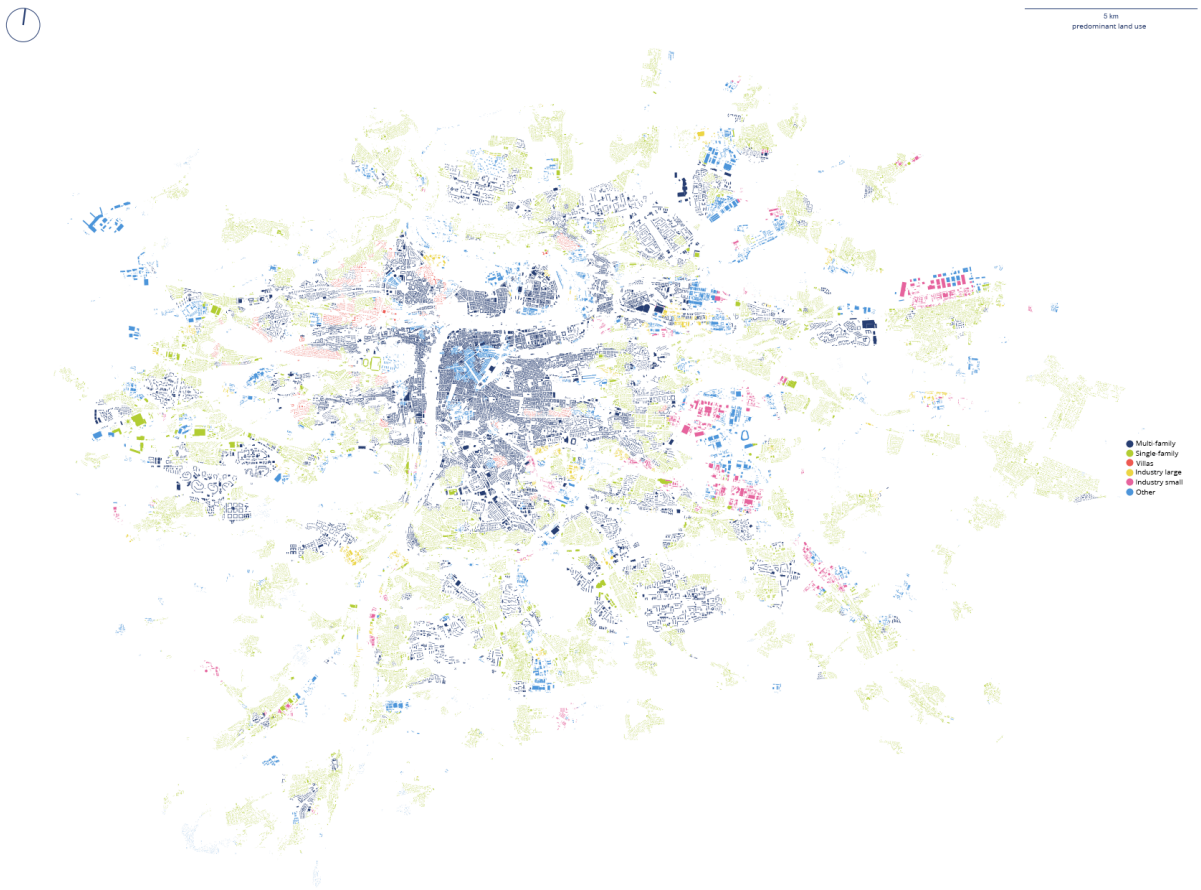


Figure 8.16: Spatial distribution of predominant land use categories.

Contingency table 8.3 shows even more evident patterns than in the previous example.

Table 8.3: Contingency table showing the counts of features per predominant land use within individual clusters.

cluster	Multi-family housing	Single-family housing	Villas	Industry small	Industry large	other
0	91	14412	819	0	0	15
1	30	304	2	72	287	1343
2	6019	5426	487	3	3	78
3	201	3817	0	0	0	115
4	312	1065	1	733	546	2624
5	5905	0	0	0	0	25
6	2176	7038	255	214	146	500
7	2489	573	15	221	151	691
8	287	7321	236	0	0	1
9	191	7292	160	73	20	126
10	6609	389	196	1	1	7
11	1461	0	0	0	0	706
12	6684	199	0	0	0	2
13	6	14794	130	8	0	54
14	3775	281	59	104	64	701
15	2731	0	0	0	6	323
16	431	2632	78	49	97	261
17	505	11127	466	11	2	34
18	42	7384	61	41	6	1230
19	18	31	0	27	706	874

Reported Chi-square results, based on the contingency table, assessing whether there is a significant relationship between two variables (land-use and clustering), are $\chi^2(95, N = 140315) = 176165.83, p < .001$, which indicates a significant relationship. Cramérs-V value is 0.501, indicating high association, higher than in the previous case. The land-use typology is known to be related to the form, even though not always in the same manner (see compact blocks vs modernism being in the same class). The results demonstrate that the clustering does reflect the similar subdivision as land-use presumes, hence indicate

the validity of the classification.

Detail of the composition of individual clusters (figure 8.17) shows in even more straightforward way the clear relationship between both variables. Cluster 11 (historical core) has a higher proportion of *other* uses caused by its central position where parts of the historical core are used as the central business district, therefore bringing different land uses to the same area.

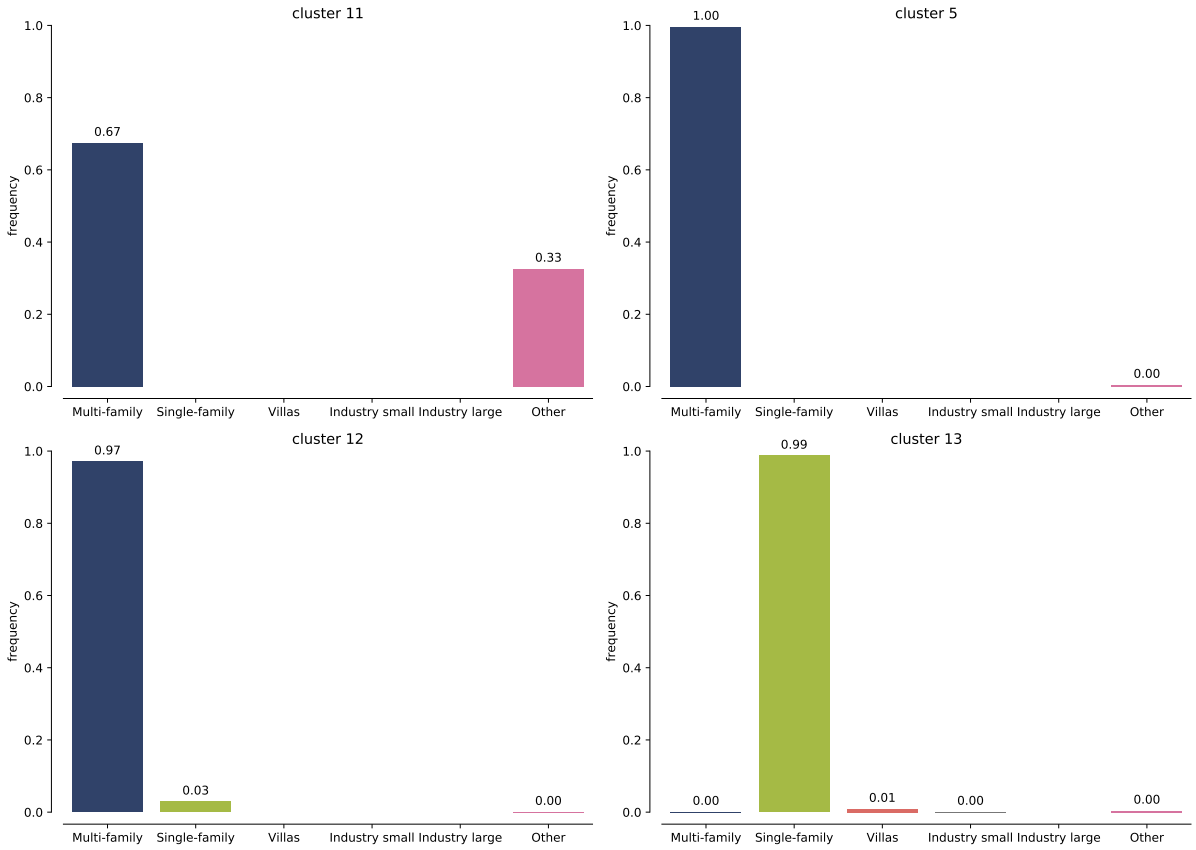


Figure 8.17: Illustration of composition of selected representative clusters from the perspective of predominant land use.

Similarly, a high degree of relation is present when we assess branches instead of individual clusters (figure 8.18). What can be seen as the highest rate of imprecision is 0.09 ratio of single-family housing in the industrial city branch, which shows that the actual precision is more than 90%.

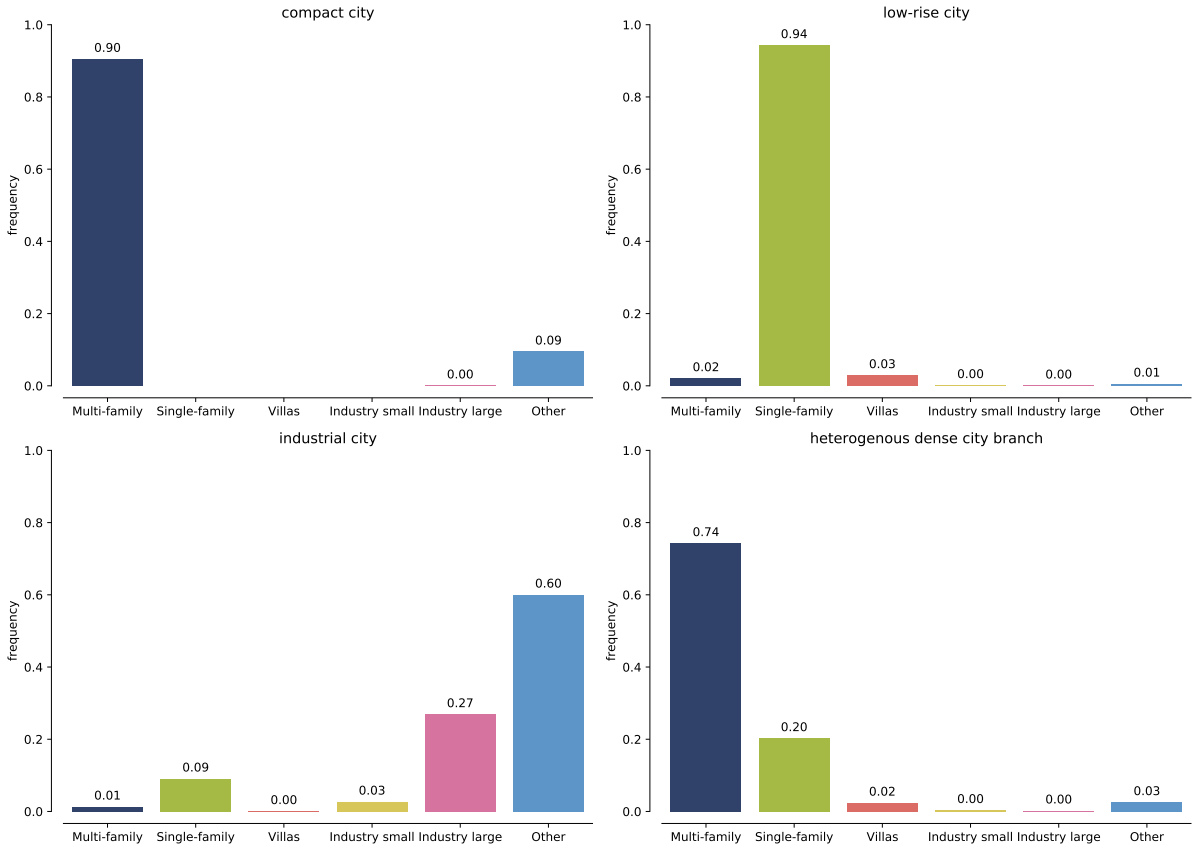


Figure 8.18: Illustration of composition of selected representative branches from the perspective of predominant land use.

Illustration of spatial distribution tells the same story and it is not necessary to go into the details. The same branch of the dense city (figure 8.19) as above shows what is already explained in numbers, i.e. the high level of overlap between land-use and clustering.

Figure 8.19: Illustration of the overlap between branch representing dense compact development and predominant land use. Colours represent land use and the saturation the extent of the branch.

The results of validation of clustering and taxonomy based on land use data show even higher levels of similarity between the variables than in the case of historical origin, clearly indicating that the morphometric tissue types can be relevant.

8.3.2.1.3 Municipal typology Comparison of the clusters and the qualitative municipal typology is the last validation using additional proxy data. As described in the section 8.2.2.1.2.3, there are certain limitations when it comes to the municipal typology itself which results in the necessity to drop a fraction of features (20960 are dropped, therefore 119355 features are used) before the actual analysis. The data used for the

analysis and their respective types are illustrated in figure 8.20. The types are clearly distinguished and in a sense, reflect the combination of origin and land-use categories.

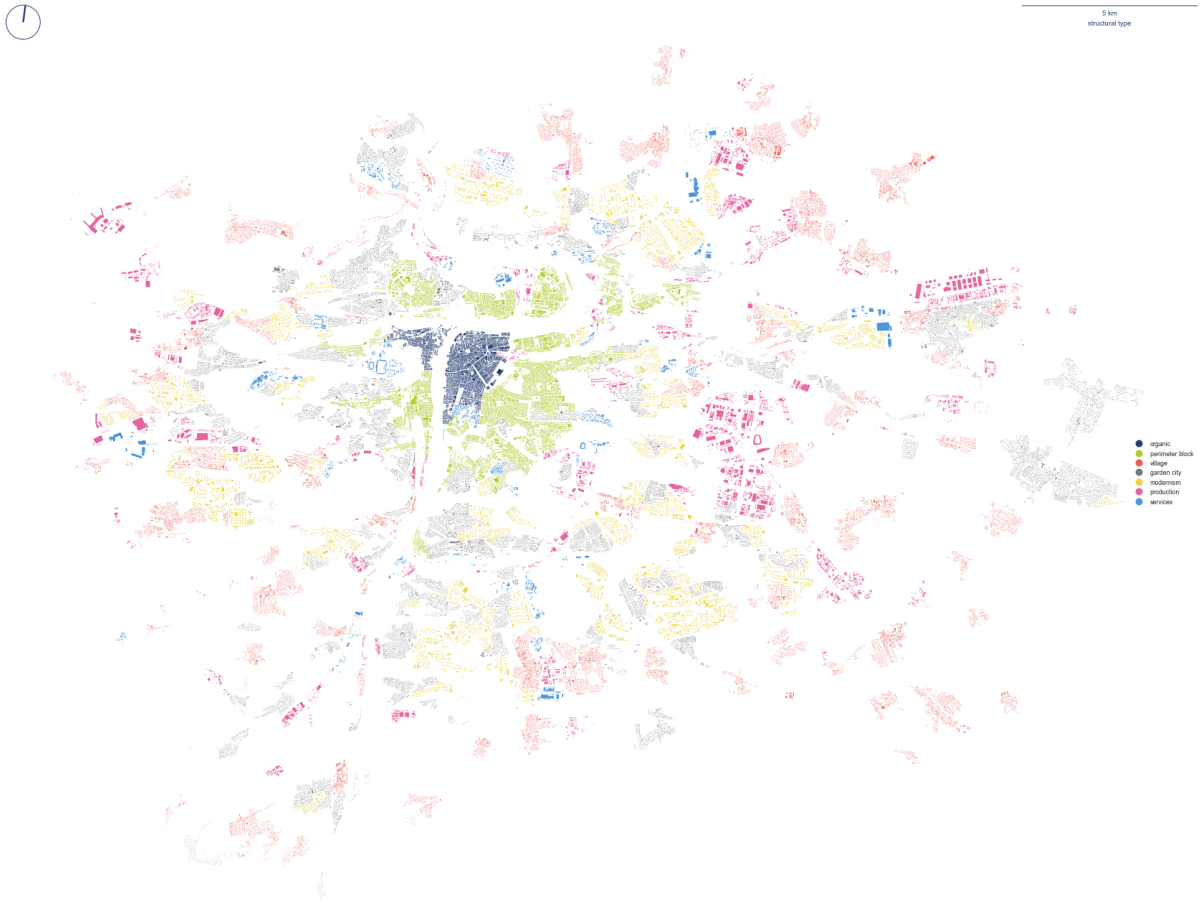


Figure 8.20: Spatial distribution of individual classes of qualitative municipal typology in Prague.

Contingency table 8.4 shows even more evident patterns than in the previous examples.

Table 8.4: Contingency table showing the counts of features per municipal typology classes within individual clusters.

cluster	organic	perimeter block	village	garden city	modernism	production	services
0	0	0	5158	9386	3	0	0
1	0	0	252	87	0	960	92
2	0	305	965	3802	3600	13	49

cluster	organic	perimeter block	village	garden city	modernism	production	services
3	0	0	381	1338	616	0	51
4	1	31	660	269	59	1997	543
5	375	5350	0	7	0	0	0
6	0	207	2551	3050	1042	445	283
7	0	158	178	174	1331	479	434
8	0	0	1941	4655	101	0	0
9	0	2	5659	1569	3	0	0
10	115	5248	70	853	46	2	3
11	2137	0	0	0	0	2	11
12	0	6	79	101	6540	4	13
13	0	0	8805	5640	2	2	5
14	174	2825	22	258	102	294	269
15	1283	1734	0	0	0	10	9
16	0	3	1198	1103	129	257	53
17	0	8	3538	6300	301	25	15
18	0	0	4327	3153	34	34	59
19	0	0	6	1	1	1461	73

Reported Chi-square results, based on the contingency table, assessing whether there is a significant relationship between two variables (municipal typology and clustering), are $\chi^2(114, N = 119355) = 325595.20, p < .001$, which again indicates a significant relationship. Cramérs-V value is 0.674, indicating high association, the highest of all tested datasets. That is no surprise, as both layers (clusters and municipal typology) are trying to describe the same aspects of the city. Considering the MAUP-related imprecision of municipal dataset and error margin of clustering, the results offer a strong indication of the validity of the clustering output.

The relation is also clearly present in the composition of each cluster (figure 8.21) with the only difference of distinction between village and garden city typology. That is poorly specified in the original dataset and the difference is unclear as it mixes historical origin and ideology with the morphological description. So the difference between the two should be taken with care as it might just be misleading.

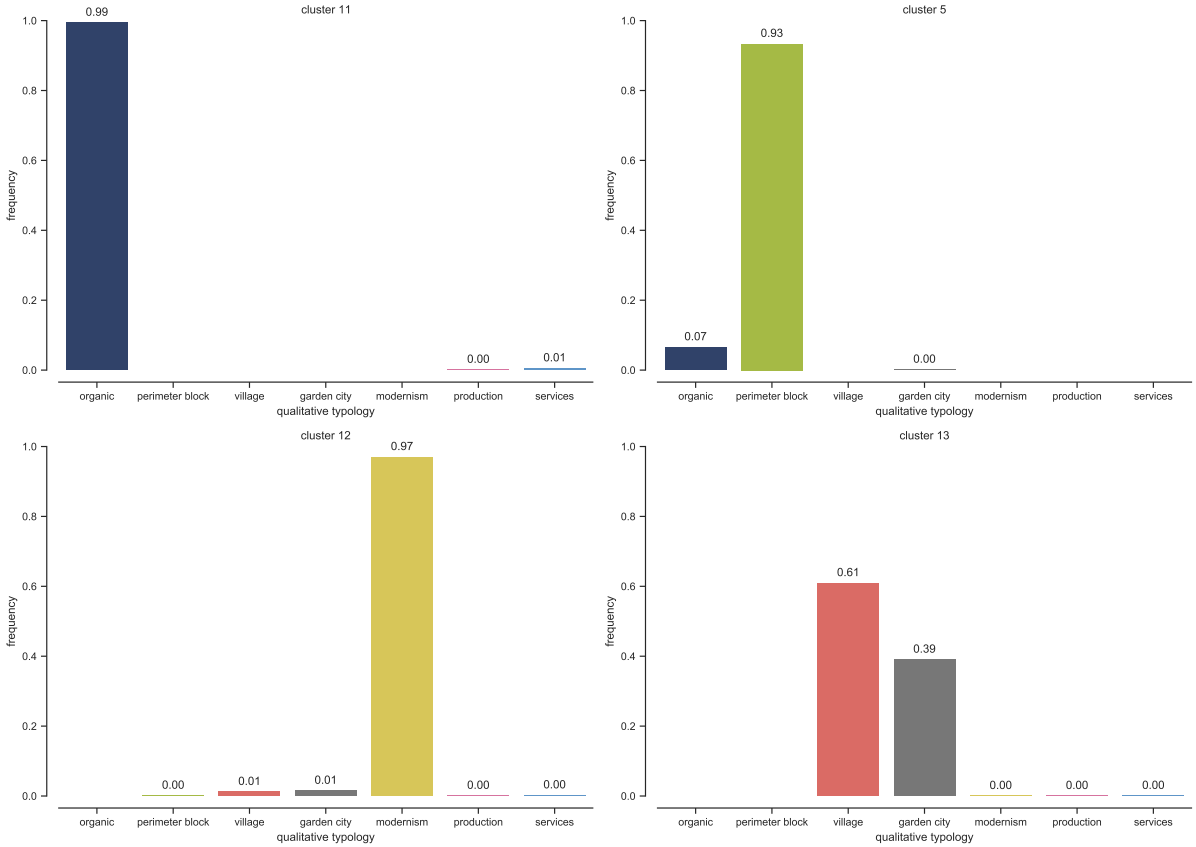


Figure 8.21: Illustration of composition of selected representative clusters from the perspective of municipal typology.

Branches are combining different groups of municipal typology in a similar manner they combine clusters (figure 8.22). We can see some deviations, but the general tendency is clear and tells a very similar story as taxonomy itself.

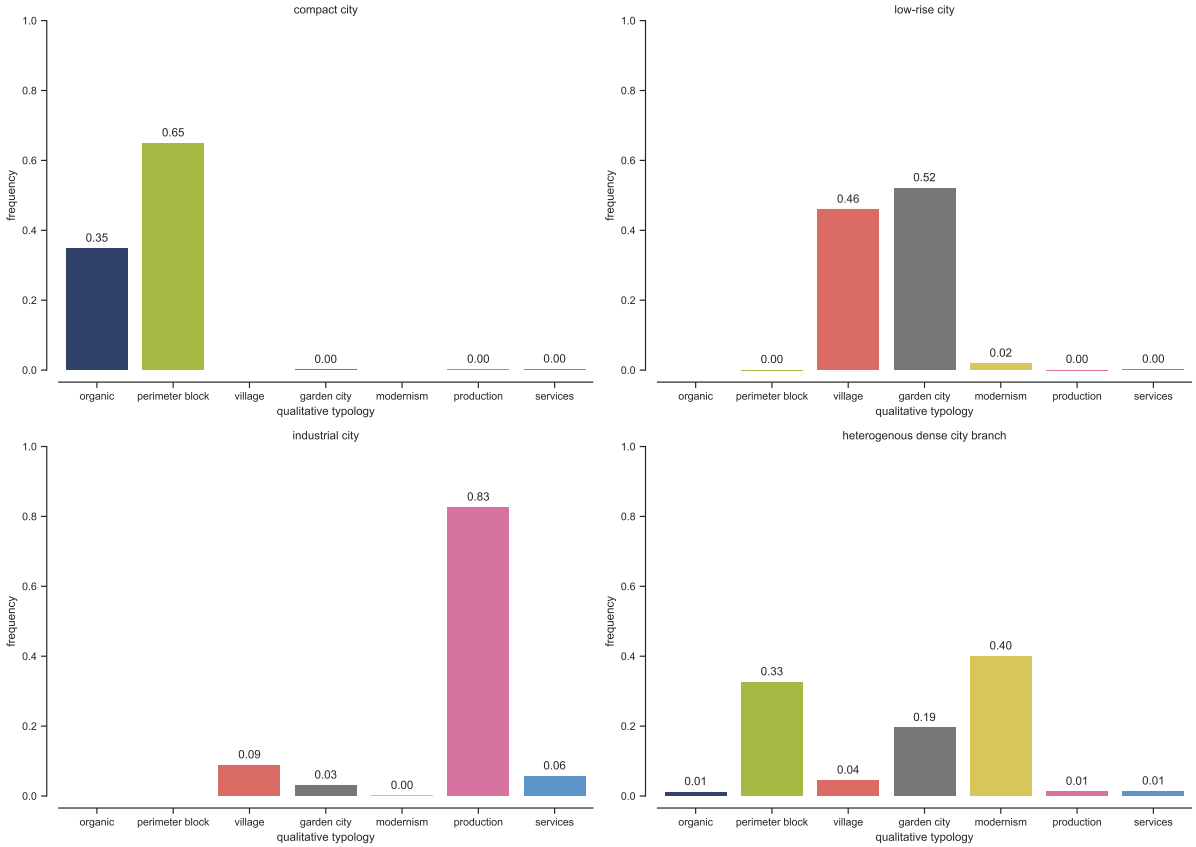


Figure 8.22: Illustration of composition of selected representative branches from the perspective of municipal typology.

A single illustration of spatial distribution on the branch of the dense historical city shows almost precise overlap with municipal typology (figure 8.23). The differences, e.g. the areas marked as *perimeter blocks* in municipal typology not captured by clusters, are mostly incorrectly classified in the typology itself due to MAUP (Dejvice University Campus in the north-west, Industrial belt of Holesovice in the north).

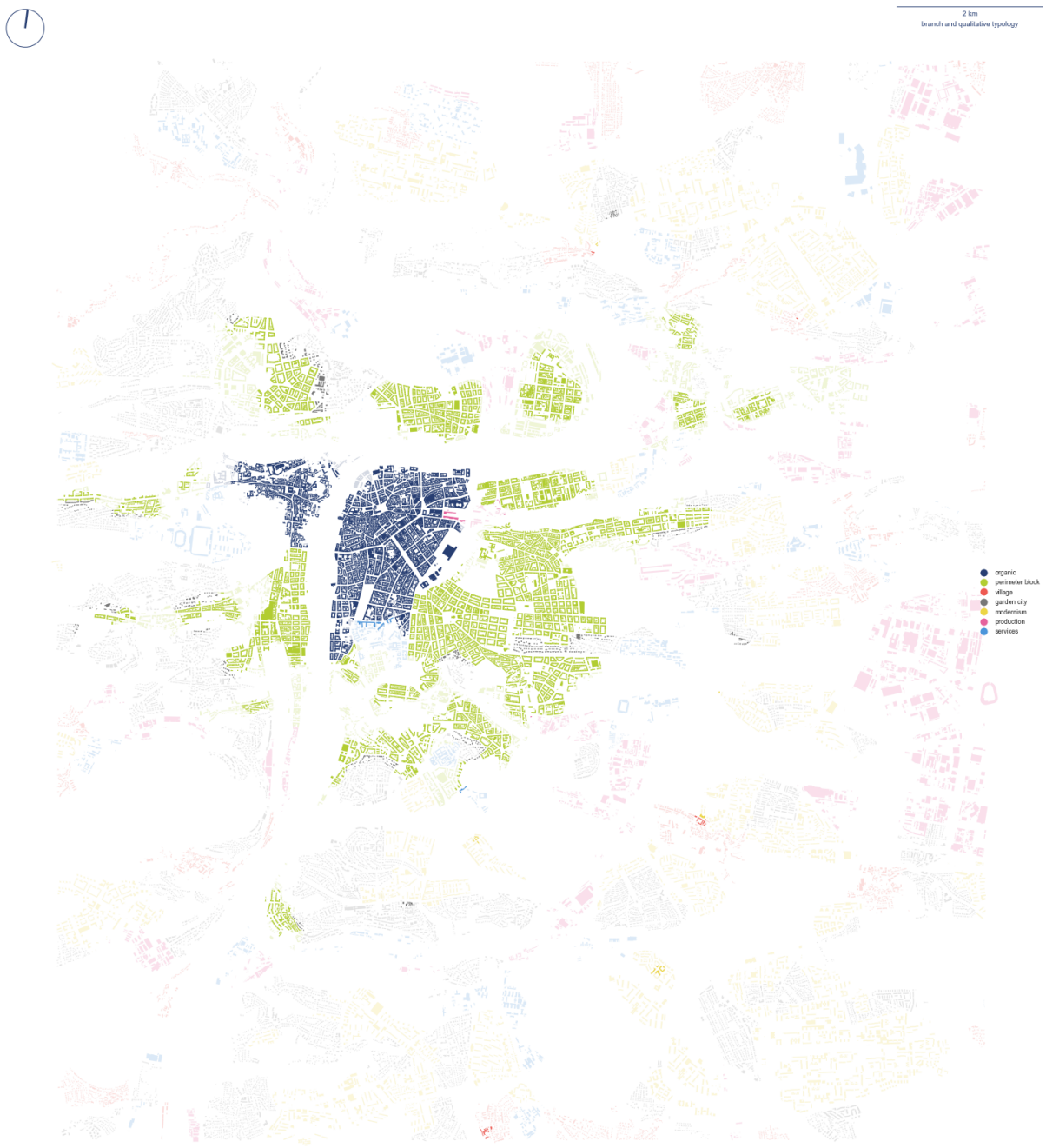


Figure 8.23: Illustration of the overlap between branch representing dense compact development and municipal typology. Colours represent classes of typology and the saturation the extent of the branch.

Municipal typology shows the highest similarity with the morphometric urban tissue types

from all three proxy layers. This layer should be similar as it tries to capture a similar perspective of the city. The fact that it does shows such high significance tells that the clustering method and classification does deliver results capturing meaningful results.

The aim of the section was to validate the results of clustering using additional data. The clusters would be seen as validated if they showed a significant relationship to all tested layers. The results of the validation show significance and moderate (origin) to a high association (both land-use and municipal typology) based on the employed tests. These results clearly indicate that the method of identification of urban tissue types proposed in the previous chapter and related hierarchical classification of urban tissues are both reflecting the morphological reality and could be seen as a valid method of urban morphology analysis.

8.3.2.2 Transferability to other places

Even though the method is valid in the context of Prague, it is unclear whether it is transferable to other contexts. It was designed with universality in mind, so it is assumed that it should be able to capture a similar level of information in other geographic and historical contexts. To test this hypothesis, the whole methodology is applied to the case study of Amsterdam, including one layer of validation using historical origin data.

Amsterdam dataset tests not only the transferability of the method but also its scalability. The number of buildings on the input of clustering is 252,385 compared to 140,315 buildings in Prague.

The results of primary and contextual characters are not presented in the main body of the text, and their distributions are available as Appendix 8.2. Since the method is following the steps defined in sections 7.1 and 8.2.1, results report the selection of an optimal number of components, results of clustering and hierarchical classification.

8.3.2.2.1 Clusters Bayesian Information Criterion shows different curve than we have seen in Prague case as it does not culminate to indicate the optimum (figure 8.24). That is a situation which may happen with BIC and indicates overfitting of the model, which BIC is unable to correct.

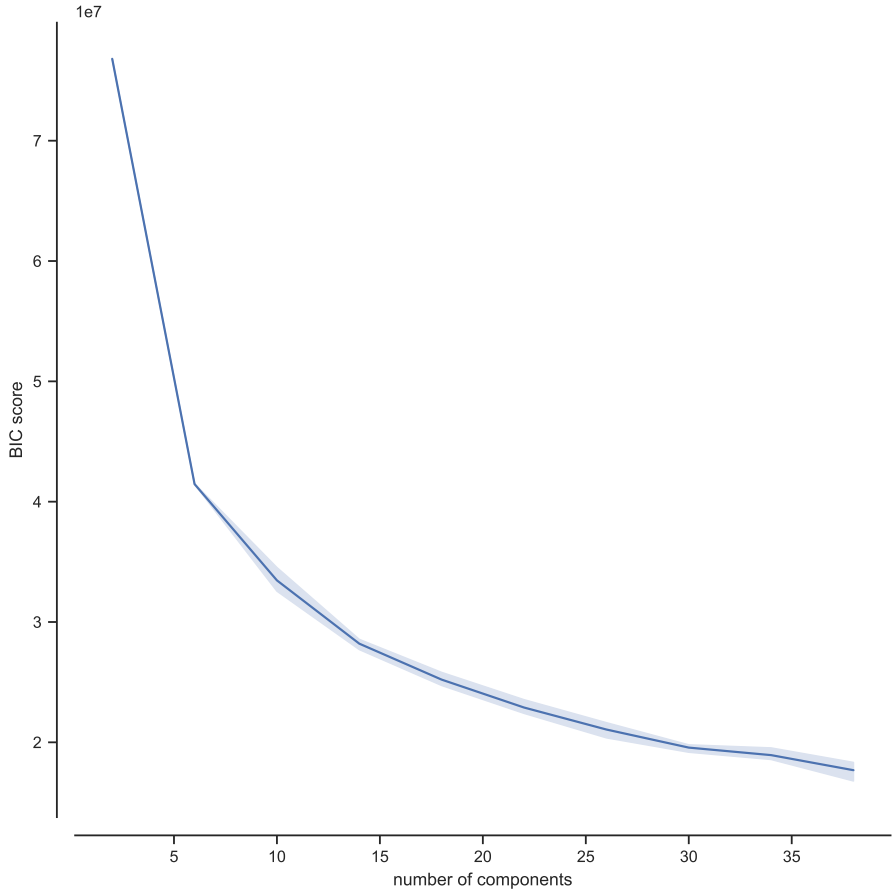


Figure 8.24: Bayesian Information Criterion score for the variable number of components. Shaded area reflects .95 confidence interval.

In cases like this, it is recommended to follow a different principle of identification of the optimal number of components derived from the gradient of the curve (figure 8.25). The resulting number is then the smallest value on the stabilised gradient. Once it flattens, i.e. the change starts to become more linear, it is not expected that a larger number of components will significantly improve the classification. For that reason, the number of components used for identification of clusters in Amsterdam is derived from the gradient curve as 30.

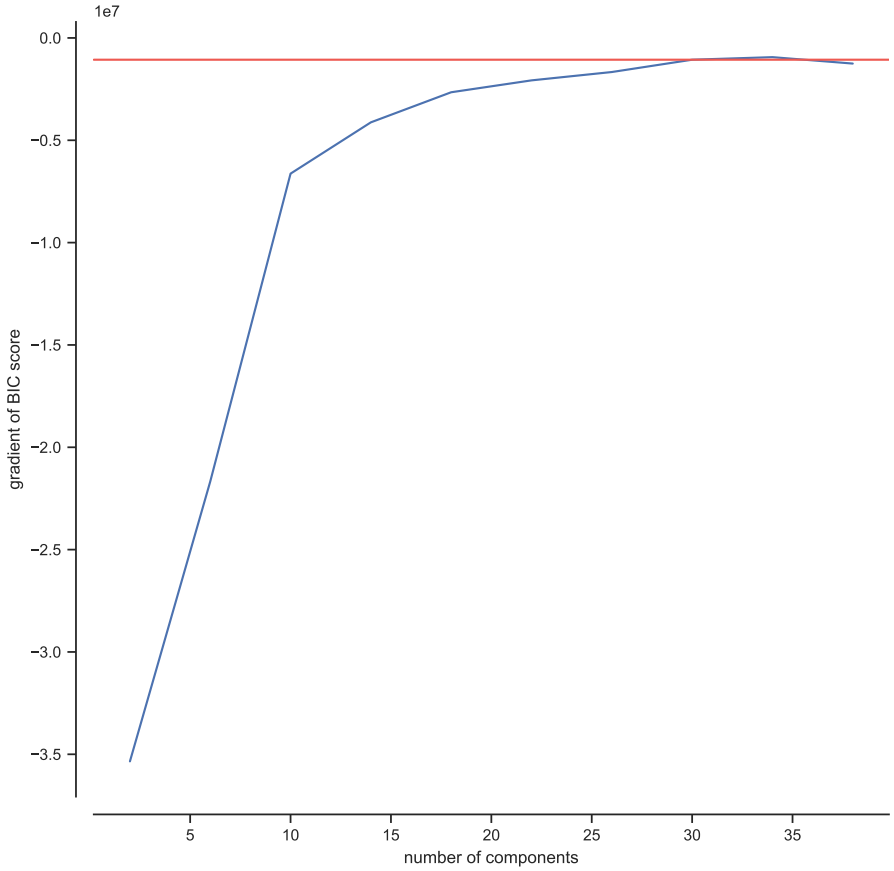


Figure 8.25: Gradient of Bayesian Information Criterion score for the variable number of components. The shaded area reflects a .95 confidence interval. The red line marks the culmination of the gradient at about 30 components.

Results of the cluster analysis are shown visually on figures 8.26 and 8.27. Spatial distribution on the next two pages (the full extent of the case study and the detail of the city centre) show potentially meaningful clusters. Especially clusters in the city centre show a high degree of legibility initially seems to reflect different phases of the development of Amsterdam. From this first perspective, it looks promising, and there is no reason to think that the method identification of clusters is not transferable yet.

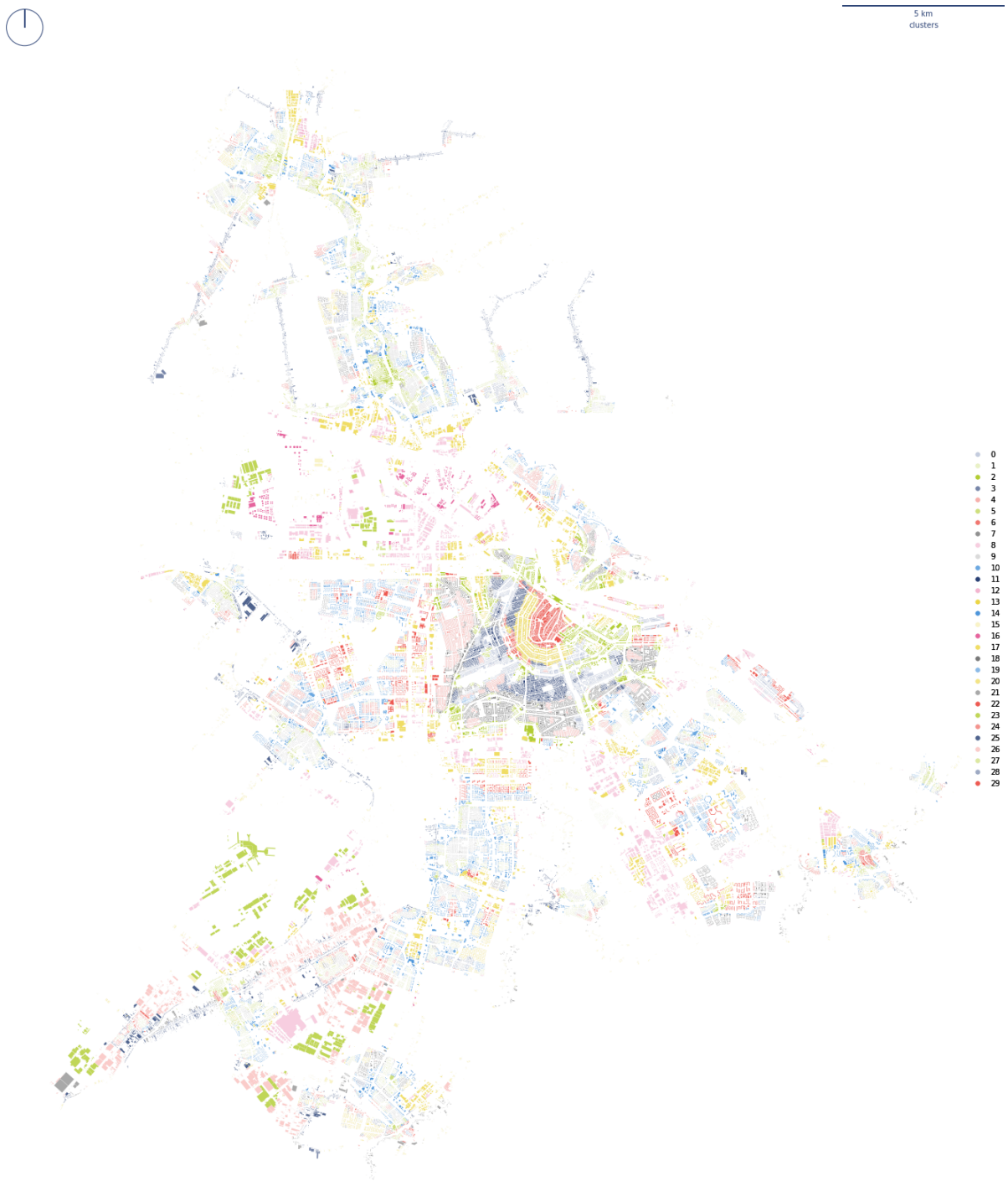


Figure 8.26: Spatial distribution of 30 clusters as identified by GMM based on morphometric data.

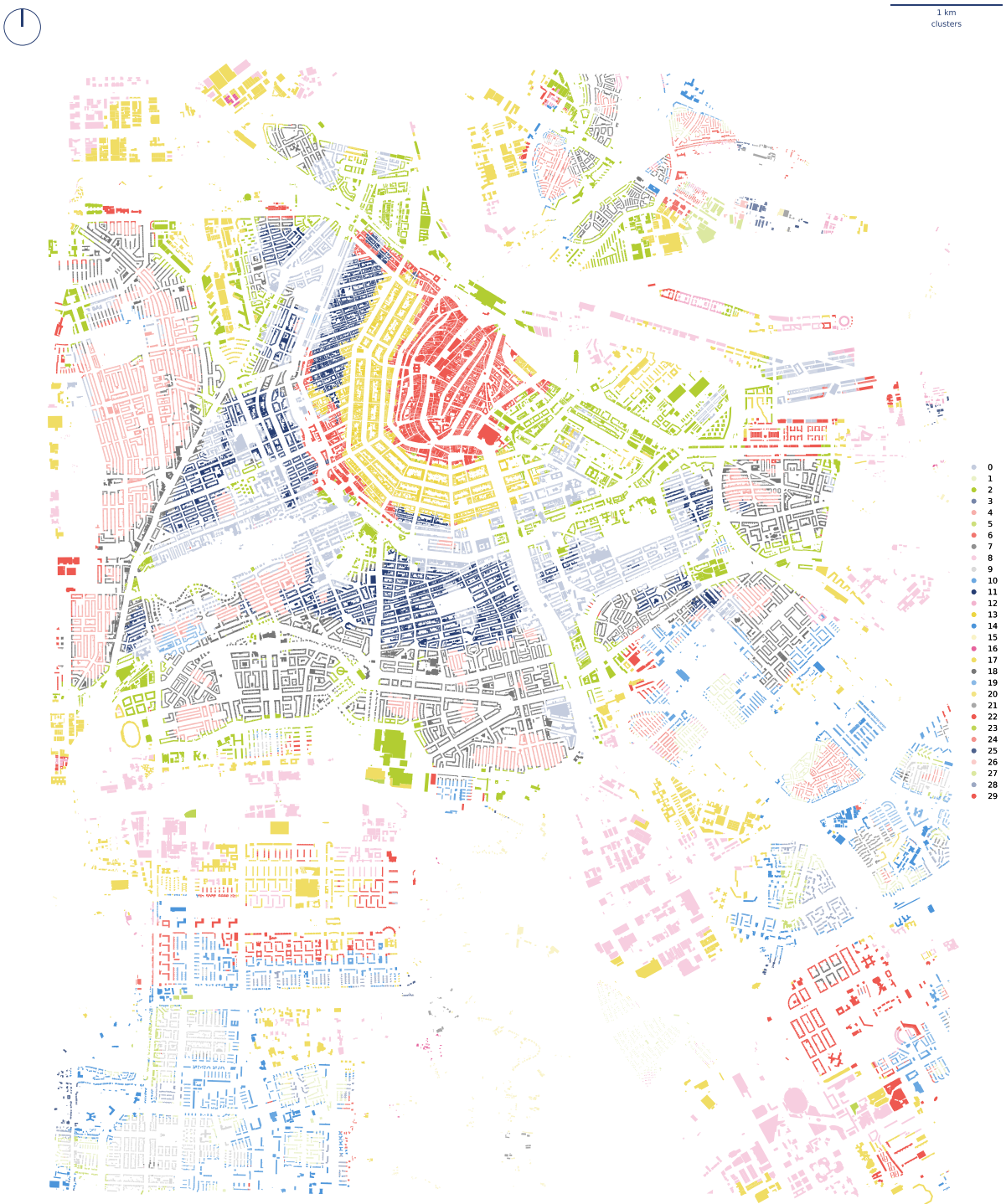


Figure 8.27: Detail of spatial distribution of 30 clusters as identified by GMM based on morphometric data.

Consider cluster 29 in the previous figure. It, with some margin of error, captures the original historical core of Amsterdam. The next concentric belt is captured as cluster 13. The difference between dark blue (11) and grey/pink areas (4, 18) reflects the change in the planning paradigm with the rise of New Amsterdam School (Panerai *et al.*, 2004), captured especially by grey cluster 18. On the other hand, the results in peripheral parts of the city show a certain degree of fuzziness, which again can be a sign of potential overfitting, indicating that the actual number of clusters might need to be smaller. However, that may be resolved by taxonomy, by linking individual clusters together to branches and assessing the urban form via branches.

It is important to tests the method in additional geographical contexts than initial Prague case study to ensure that it is transferable. Applying the method on the case of Amsterdam, NL, first results indicate that the method is transferable. However, the case also raised questions regarding the stability of the method of determination of the number of components for Gaussian Mixture Model clustering as BIC curve did not culminate. That should be further explored in further research, and additional ways of identification of the optimal number might have to be introduced.

8.3.2.2.2 Hierarchical tree The hierarchical tree (figure 8.28) derived from morphometric data of Amsterdam shows similar characteristics as we have seen in Prague, with the significant bifurcation into two branches distinguishing predominantly industrial and housing tissues and then consequent bifurcations lower in the tree distinguishing different rules of the organisation.

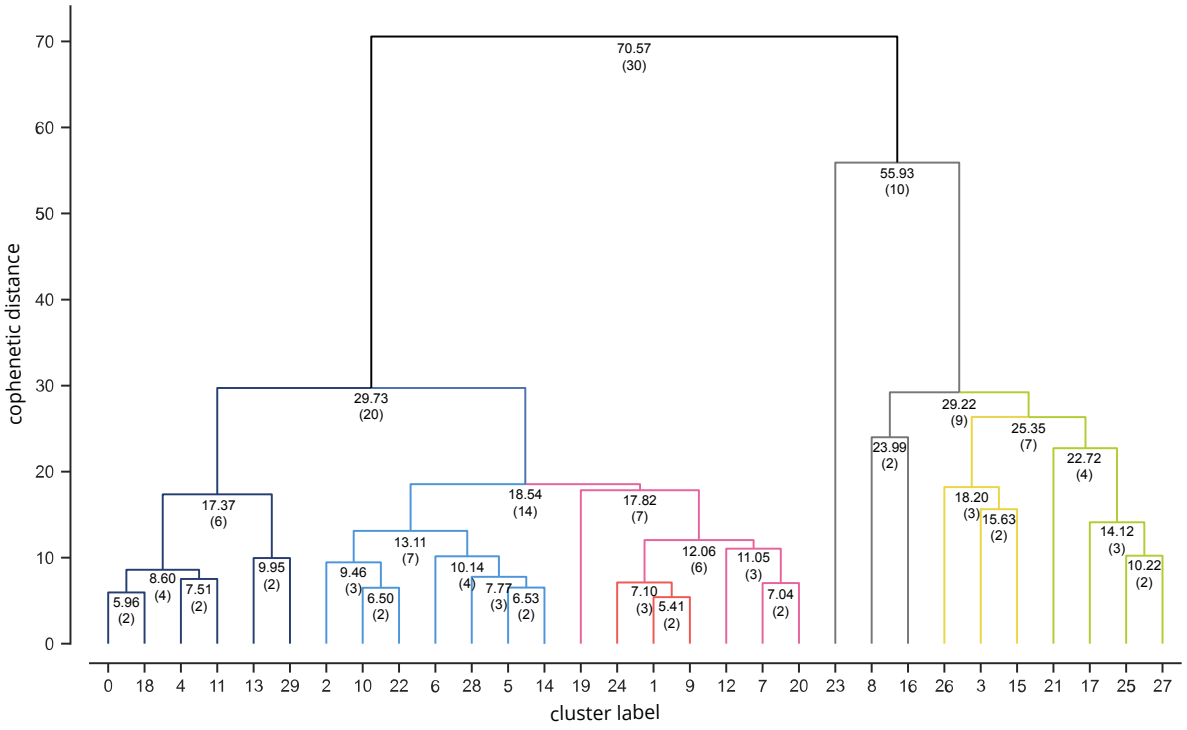


Figure 8.28: Dendrogram representing the results of Ward's hierarchical clustering or urban tissue types in Amsterdam. The y-axis shows a cophenetic distance between individual clusters, i.e. their morphometric similarity. Branches are interpretative coloured - the colours are then used on maps illustrating spatial distribution of these branches.

The spatial distribution shows what the branches actually mean. See the whole case study and its detail coloured according to the dendrogram on figures 8.29 and 8.30. Again, each cluster has a different shade of the same colour when it belongs to the same branch.

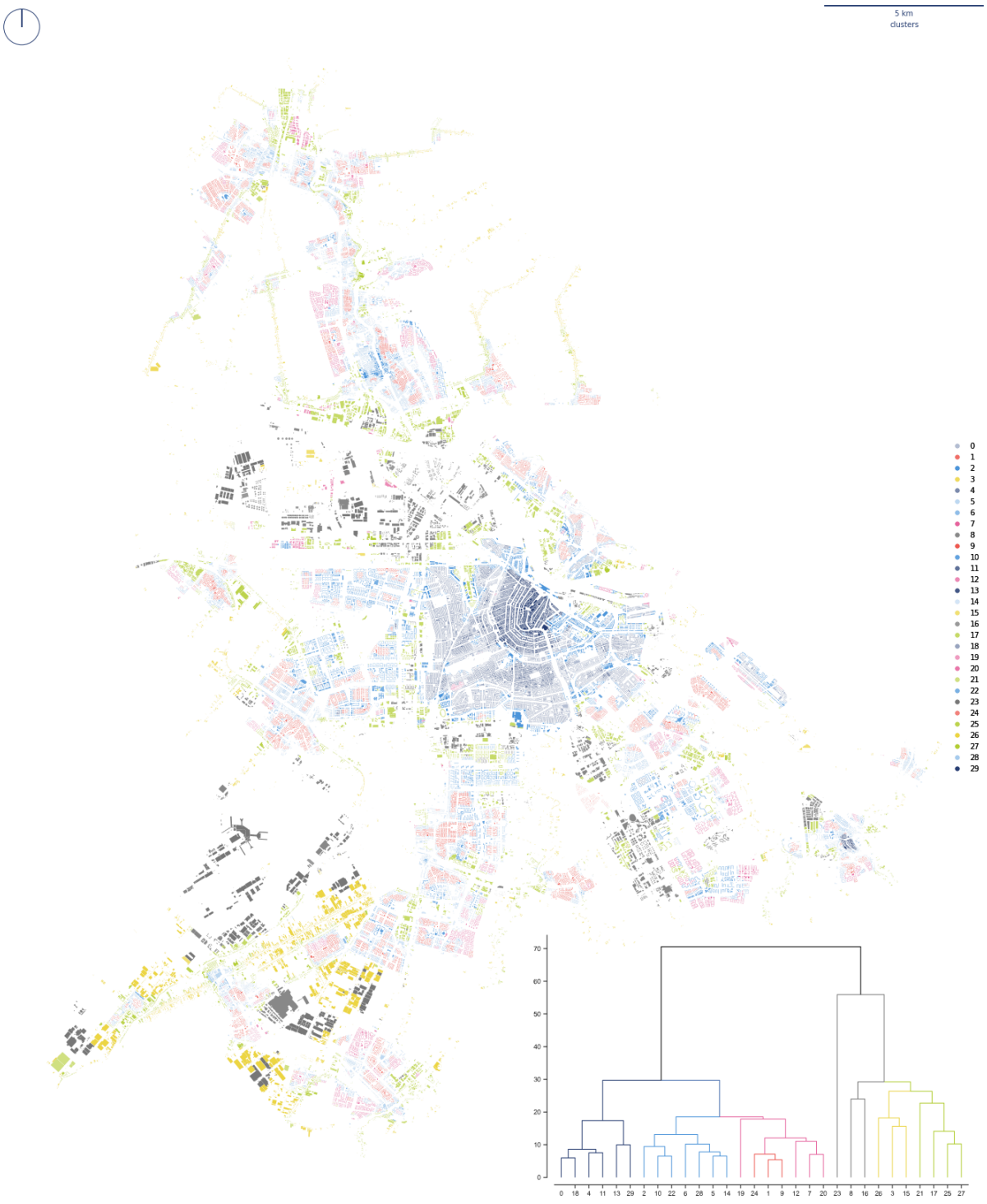


Figure 8.29: Spatial distribution of different branches of dendrogram in Amsterdam. Each tissue type is coloured according to a branch it belongs to, with a minor differences in colour intensity to allow for distinguishing of individual clusters.

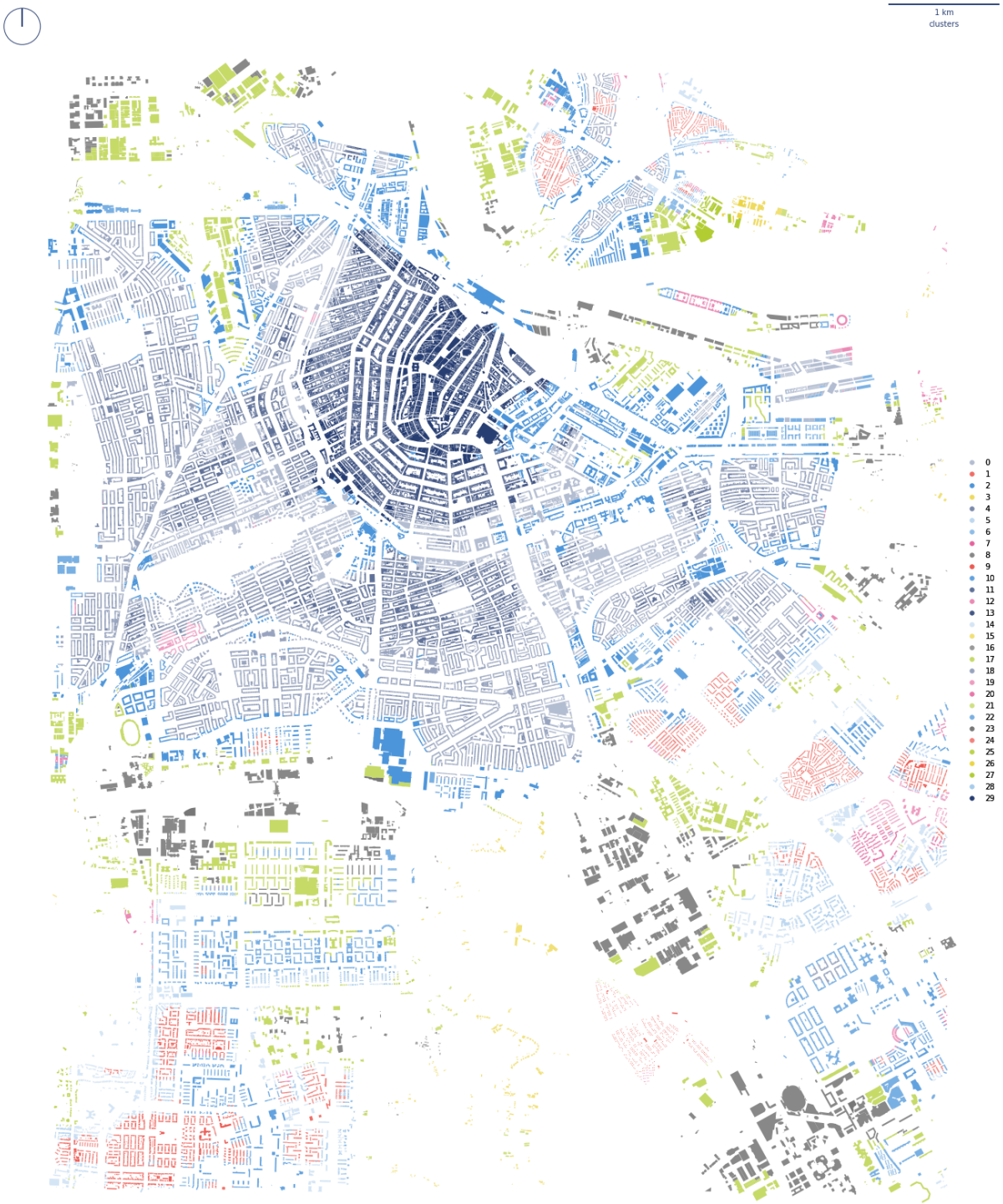


Figure 8.30: Detail of spatial distribution of different branches of dendrogram in Amsterdam. Each tissue type is coloured according to a branch it belongs to, with a minor differences in colour intensity to allow for distinguishing of individual clusters.

The first bifurcation is analogous to one identified in Prague and divides the city to the organised (figure 8.31) and unorganised (figure 8.32) parts. Organised is composed of predominantly residential use. It entails the whole historical core and major parts of residential housing. However, there are major gaps between contiguous areas caused by the high presence of all sorts of industry and other uses. After all, Amsterdam is historically a port city.

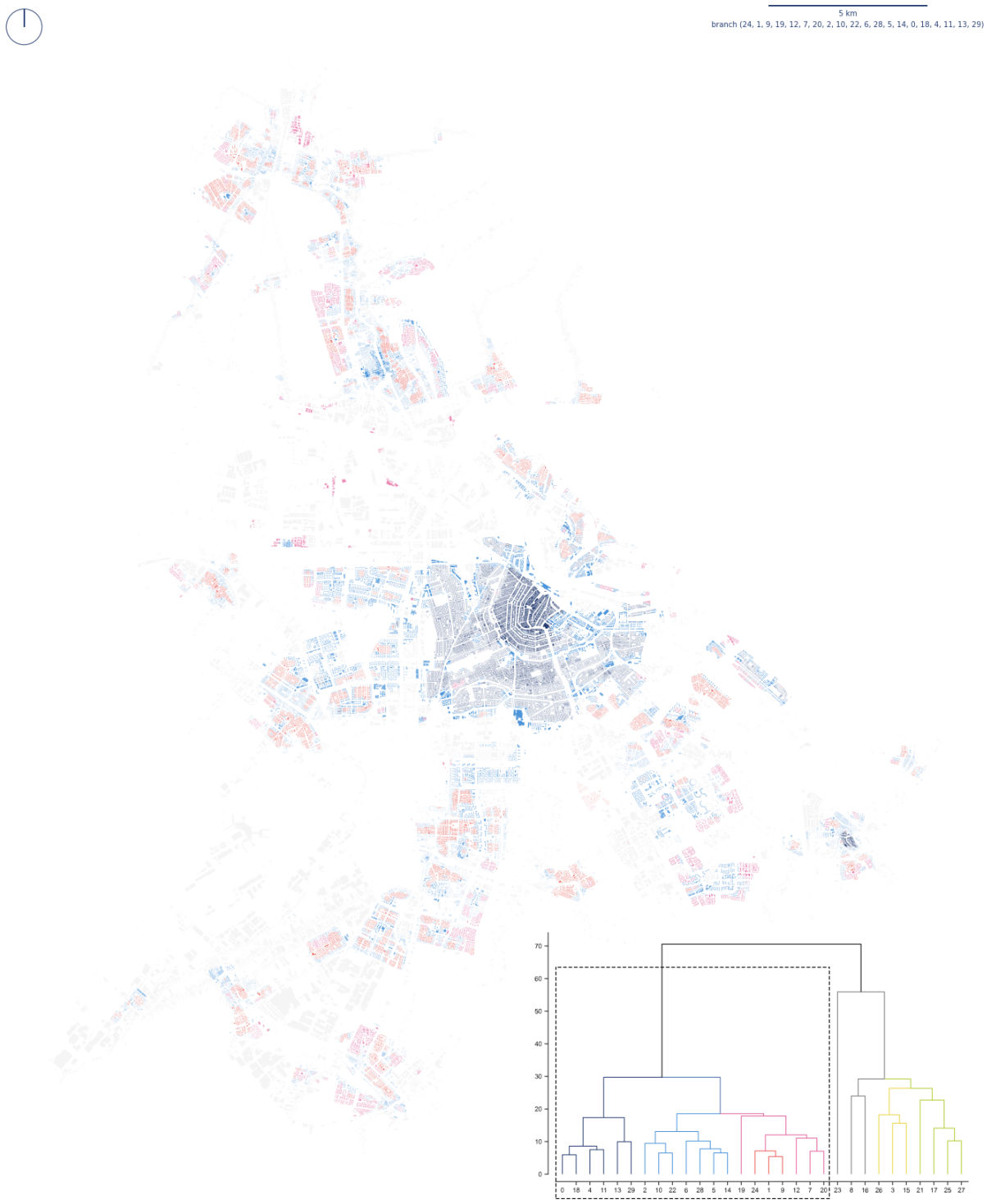


Figure 8.31: Spatial distribution of clusters within a branch representing organised city in Amsterdam. Although there are significant gaps, the overall distribution of tissues seems to be more contiguous than in Prague.

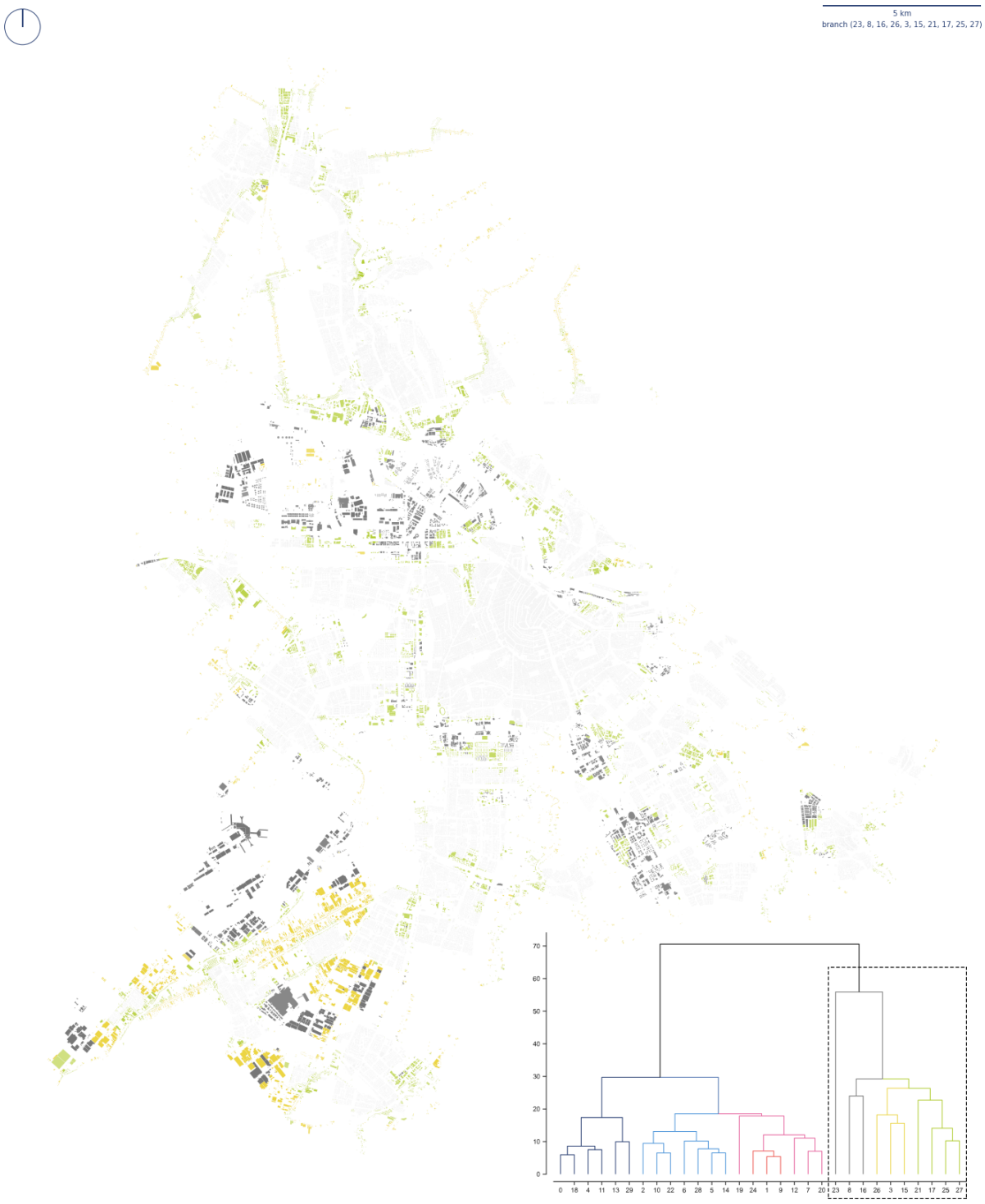


Figure 8.32: Spatial distribution of clusters within a branch representing unorganised city in Amsterdam. Similarly to the previous figure, clusters belonging to this branch tend to form either contiguous areas or follow main roads in between.

The branch comprising the organised city splits into high density and low density in a similar manner as Prague's case does. High density is mostly historical core 8.33, in this case, the one of Amsterdam and also Weesp on the east side of the city. Moreover, parts of the modern development of former port sites tend to show similar characteristics. Further bifurcation within lower density development is readable and follows sorts of the compactness and homogeneity of patterns, so we have organised residential areas, their more heterogeneous counterparts and fringe areas around compact high-density areas (figure 8.34). The tree also resolves the initial fuzziness of individual clusters (notice that light blue, red and pink branches are very similar to each other).

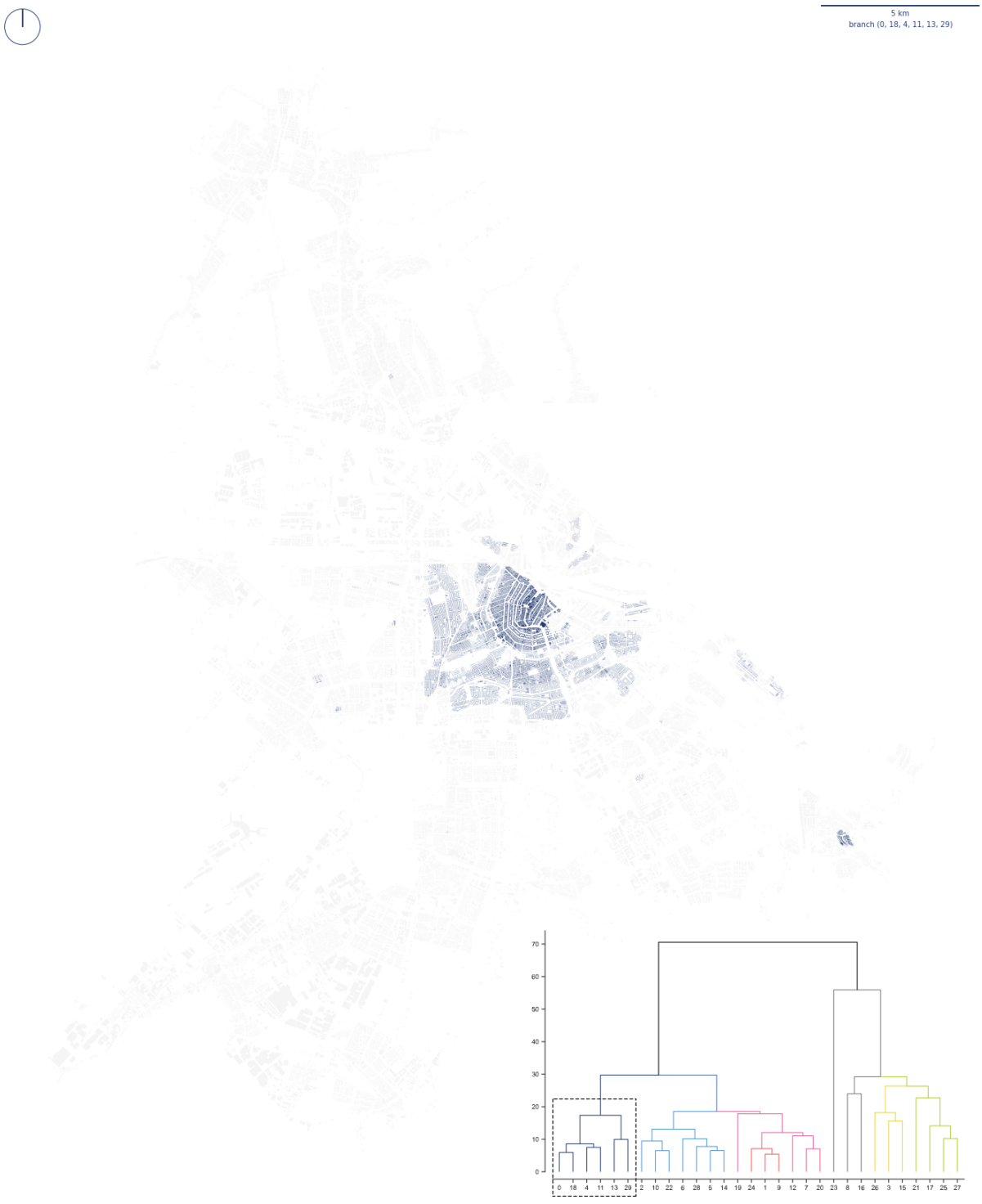


Figure 8.33: Spatial distribution of clusters within a branch representing high density, mostly historical development of in Amsterdam.

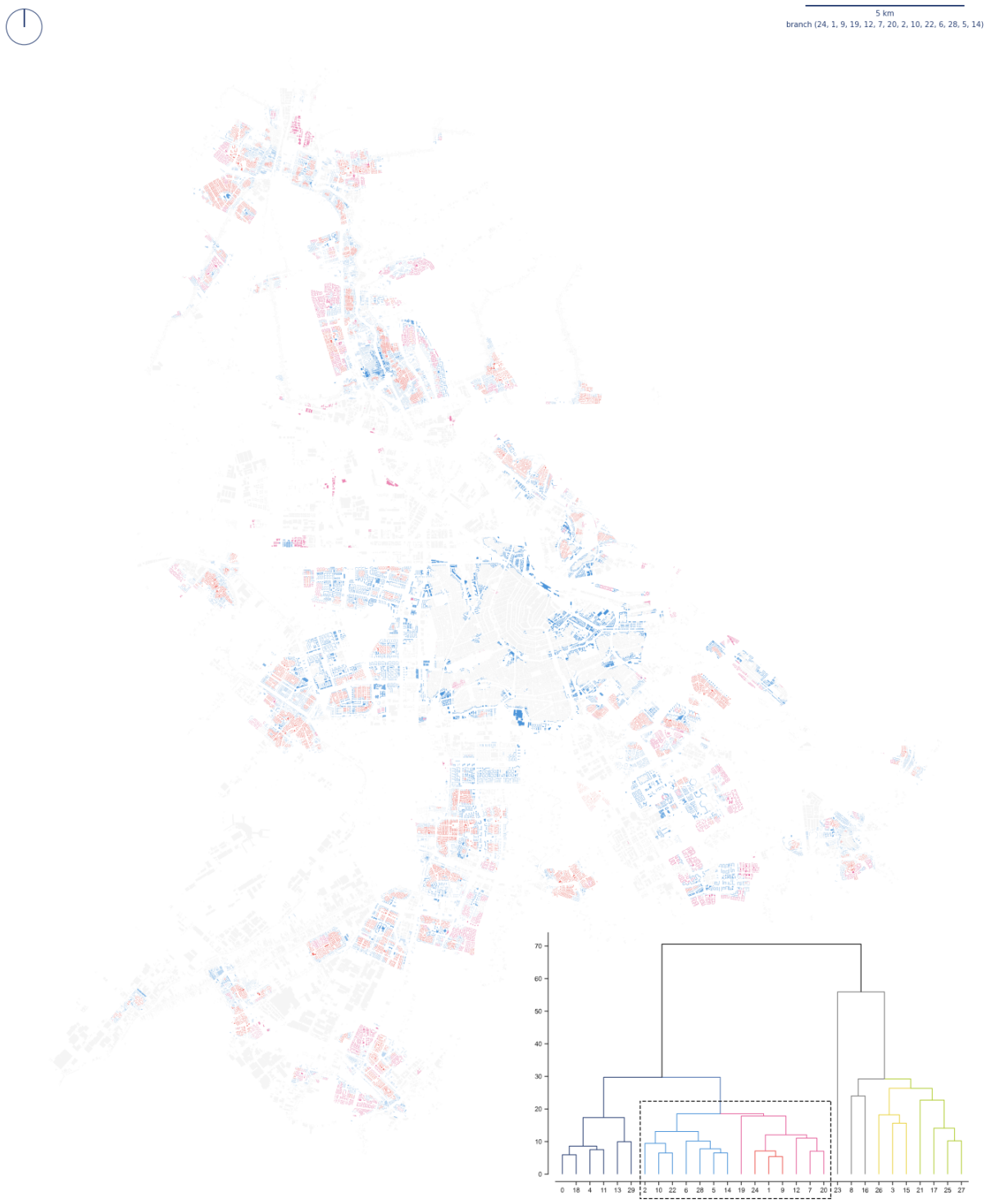


Figure 8.34: Spatial distribution of clusters within a branch representing low density development in Amsterdam.

Hierarchical taxonomy shows very similar character as was observed in Prague and visually works in a coherent way. It seems safe to conclude that the proposed method similarly behaves in both cities, which indicates its universality and transferability.

8.3.2.2.3 Validation using historical origin To further assess the validity of clustering in Amsterdam, one layer of validation based on additional data, historical origin, is used as a proxy. The initial input data are illustrated on figure 8.35.

It should be noted that it captures the origin of individual buildings, not the time when the plot was first built as in Prague. As with the data on origin for Prague, we can expect a certain degree of association, but not full as not all patterns are time-dependent. The test will be done using the same method based on the contingency table, chi-squared and Cramér's V.

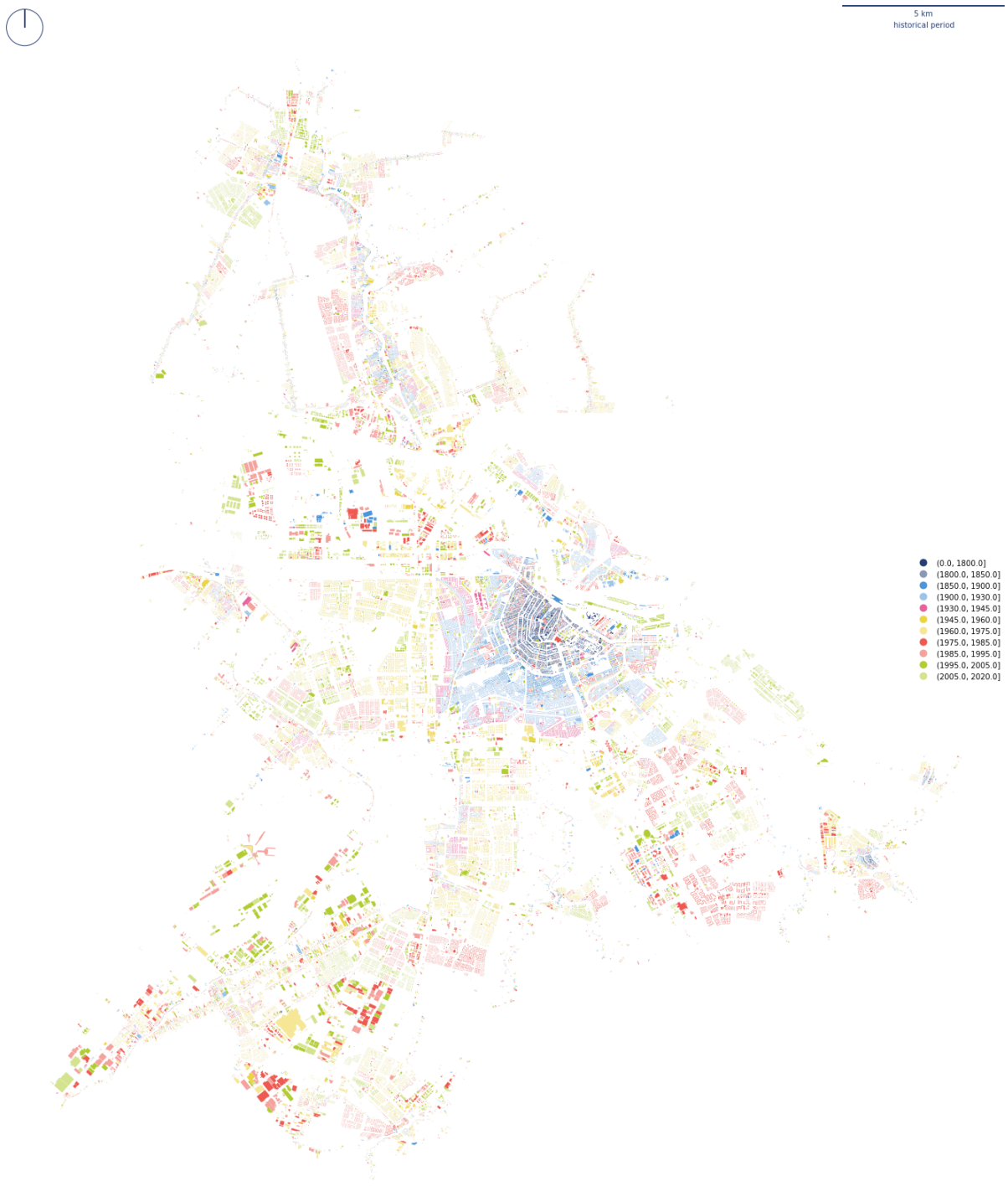


Figure 8.35: Spatial distribution of different periods of historical origin in Amsterdam.

Table 8.5: Contingency table showing the counts of features per historical origin within individual clusters in Amsterdam

cluster	1800	1850	1900	1930	1945	1960	1975	1985	1995	2005	2020
0	827	4	3012	2560	484	154	86	311	471	1048	854
1	2	1	23	752	748	5678	5841	2048	5582	2136	3210
2	761	10	625	2307	839	431	207	495	593	590	181
3	14	21	116	584	251	381	546	385	337	460	328
4	38	0	526	5743	2771	75	6	6	20	23	5
5	80	52	698	4214	2126	2132	2034	1208	932	920	763
6	6	4	83	821	490	679	1521	693	1993	1365	1287
7	0	0	1	0	7	604	5455	3455	3282	503	367
8	29	6	50	38	36	149	449	360	442	516	399
9	4	6	136	3151	4444	5095	1579	578	1154	791	1279
10	1	0	13	498	340	5379	3086	839	2949	1940	1796
11	980	0	3526	4324	272	38	50	200	452	204	58
12	2	0	35	272	114	252	630	178	718	1582	1267
13	2815	41	287	462	170	42	57	68	124	121	23
14	35	14	165	1525	911	2398	4149	1916	2511	1529	1305
15	12	9	154	498	209	302	408	369	390	375	354
16	0	0	23	19	0	277	321	171	136	65	93
17	50	19	225	553	238	691	1578	998	1197	1228	962
18	33	0	312	8359	3280	573	45	357	344	102	34
19	14	0	16	270	37	208	499	504	560	9	145
20	3	0	3	30	19	77	2031	4203	1475	325	167
21	18	11	96	323	94	207	340	228	219	243	165
22	1	0	83	152	317	1901	1108	1160	1023	709	1143
23	0	0	11	10	2	12	51	84	158	158	117
24	0	0	0	2696	942	2457	932	107	3806	768	624
25	46	14	142	1459	796	632	1112	597	675	692	609
26	1	1	38	229	164	344	569	315	329	371	228
27	48	16	128	687	319	336	421	308	325	638	309
28	0	0	31	255	155	685	859	2430	2032	1453	402

cluster	1800	1850	1900	1930	1945	1960	1975	1985	1995	2005	2020
29	2971	62	367	703	233	48	65	145	279	92	50

The contingency table 8.5 is shown above. Reported Chi-square results, based on the contingency table, assessing whether there is a significant relationship between two variables (origin and clustering), are $\chi^2(290, N = 252385) = 312903.31, p < .001$, which indicates a significant relationship. Cramérs-V value is 0.353, indicating moderate association. The value is almost the same as reported in Prague, indicating that the relationship of clustering to historical origin is consistent across both cases, and it is not case-dependent. The further exploration of the contingency table is excluded here as it is only illustrative, and the numerical values are significant.

Results of morphometric cluster analysis, consequent taxonomy and validation using historical origin data in the case of Amsterdam, NL, indicate the transferability of the method to other geographical and planning contexts, at least within the European region. The results are consistent with what was reported above for Prague, leading to the conclusion that the method can be seen as valid even from the perspective of transferability.

8.3.2.3 Expandability and compatibility

The last question which remains to be answered is whether these two cases are compatible with each other. On other words, whether we can build combined taxonomy and further extend it by adding other cases.

The extension is tested in a relatively simple way. Identified tissue types from both cased are mixed to form a single pool of clusters, all represented as cluster centroids and used as an input of hierarchical clustering. That means that while the first step, the cluster analysis is done locally for each city independently on the other, the second, taxonomy is combining them. Hence, the resulting hierarchical taxonomic tree should identify the similarity of urban tissue types across both cities.

The dendrogram from the combined pool of clusters (figure 8.36) shows the similar structure as both individual ones have, with a major bifurcation dividing unorganised/industrial areas from the organised city. Further bifurcation of organised city splits into the dense, compact city and the rest and then further into sub-types of development.

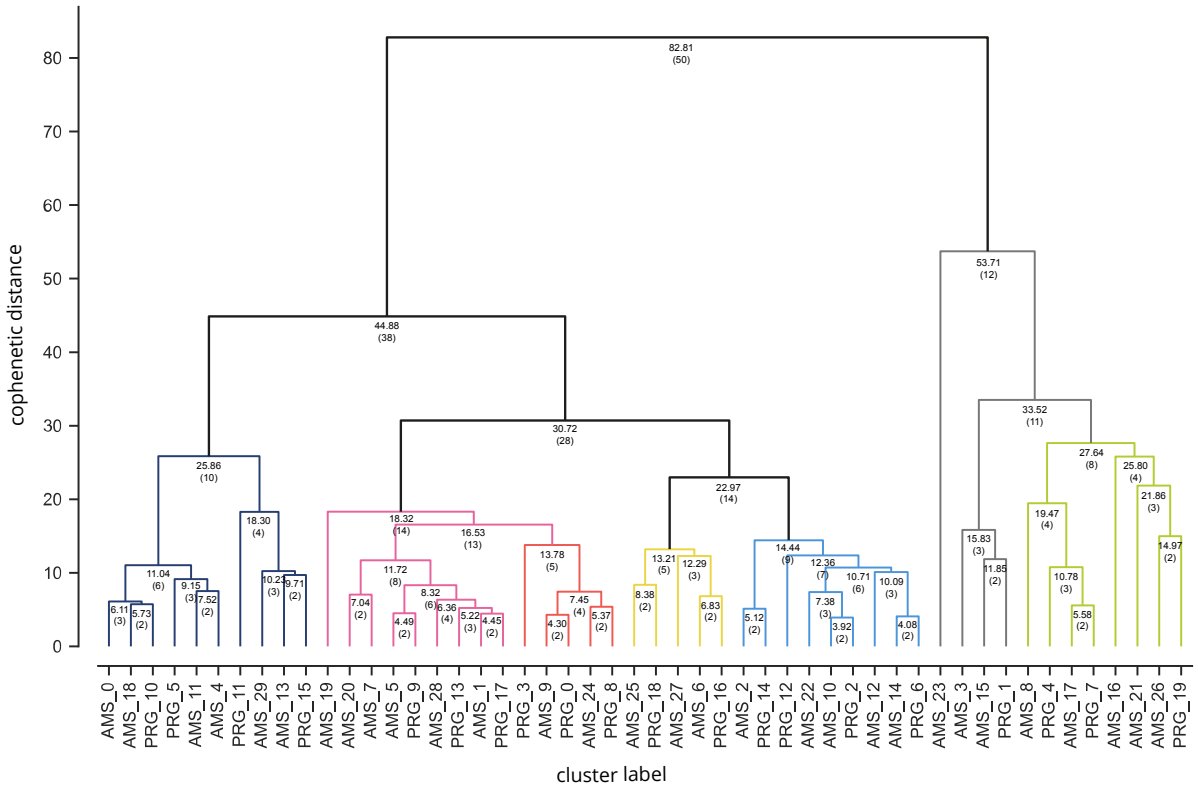


Figure 8.36: Dendrogram representing the results of Ward's hierarchical clustering or urban tissue types from a combined pool of Prague and Amsterdam. The y-axis shows a cophenetic distance between individual clusters, i.e. their morphometric similarity. Branches are interpretative coloured - the colours are then used on maps illustrating spatial distribution of these branches.

Spatial distribution of resulting branching (figures 8.37 and 8.38) tells the same story as individual classifications in both Prague and Amsterdam. What is important here is the ability to compare similar tissue types across cities, and as shown in the dendrogram, there are some which are really close to each other.

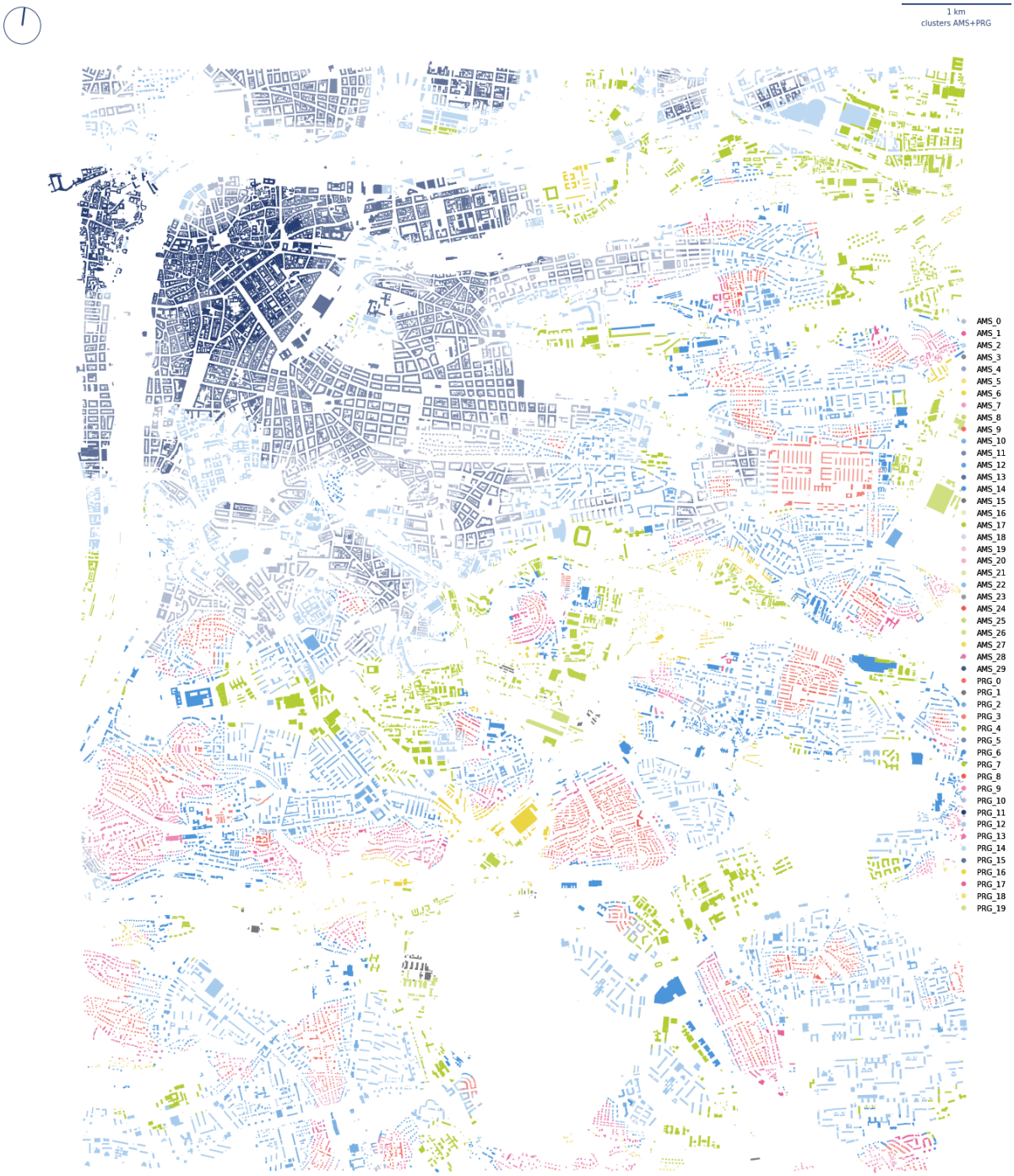


Figure 8.37: Detail of spatial distribution of different branches of a the combined dendrogram in Prague. Each tissue type is coloured according to a branch it belongs to, with a minor differences in colour intensity to allow for distinguishing of individual clusters.

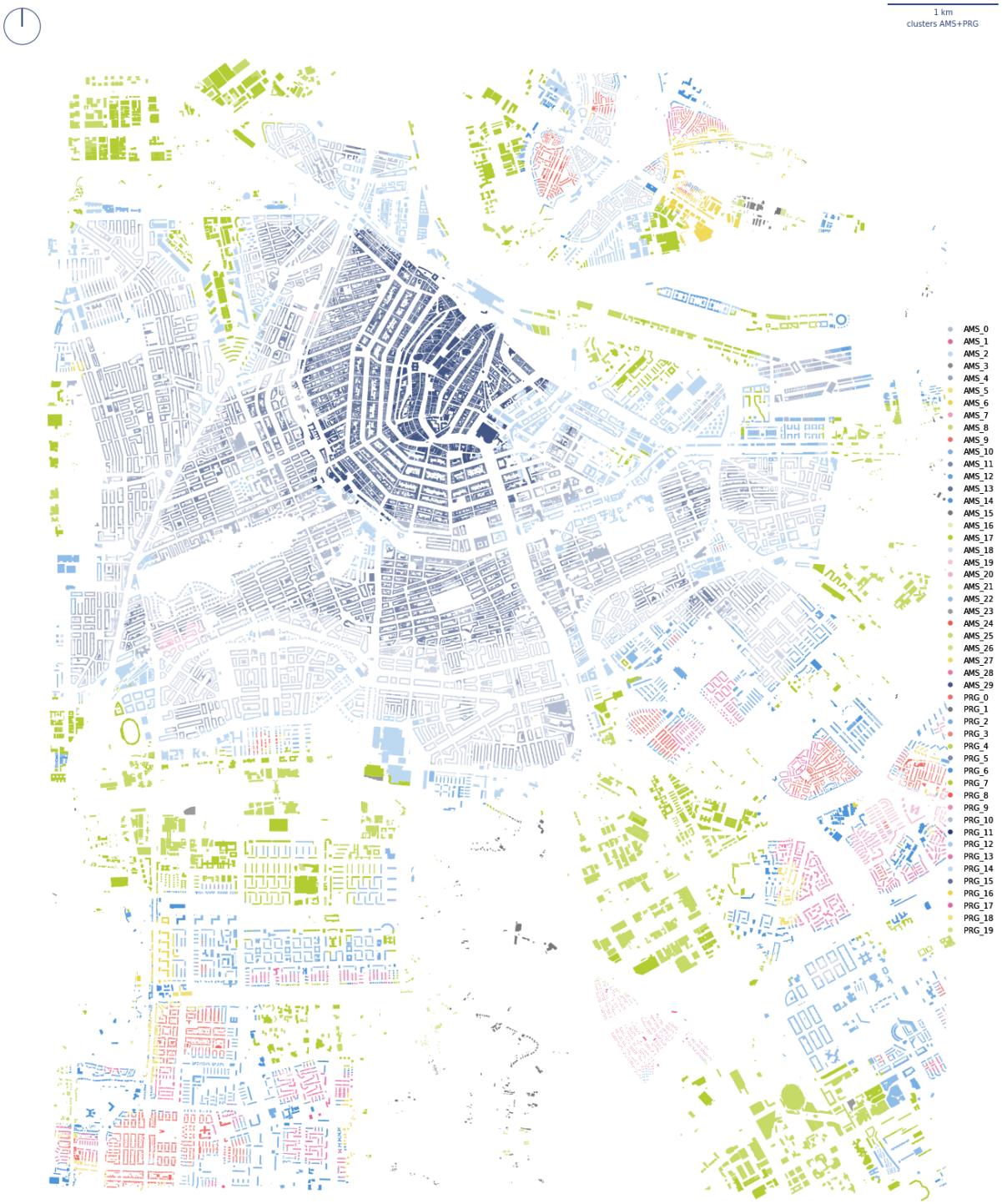


Figure 8.38: Detail of spatial distribution of different branches of a the combined dendrogram in Amsterdam. Each tissue type is coloured according to a branch it belongs to, with a minor differences in colour intensity to allow for distinguishing of individual clusters.

We can further look into the spatial distribution of the same major branches in both cities (figures 8.39, 8.40, 8.41, 8.42). It is a great tool to study their structure and compare them. One clear outcome of this first comparison is the observation that industrial areas are much larger in Amsterdam due to its port nature compare to traditionally mercantile Prague, but such an analysis could be done in a very detailed manner. However, that is not within the scope of this study.

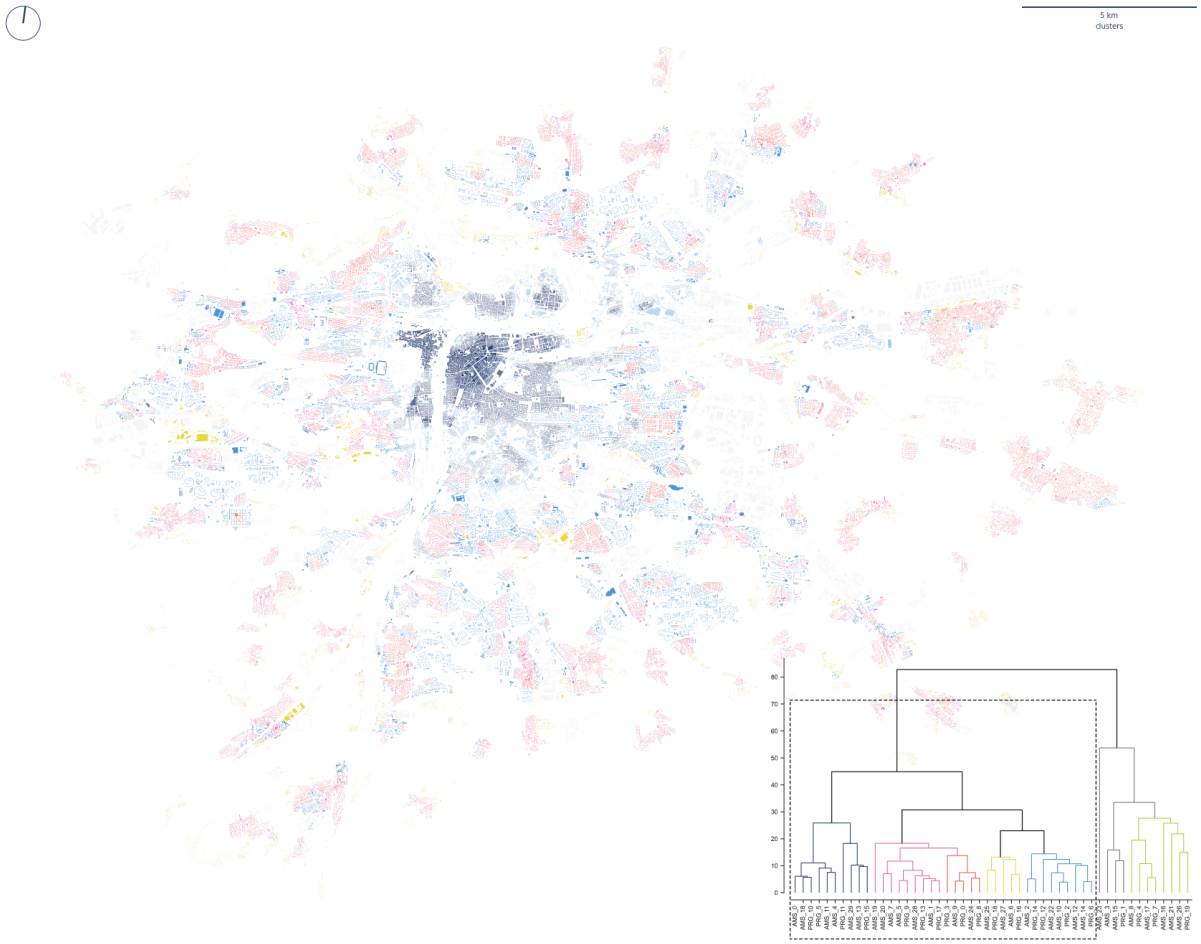


Figure 8.39: Spatial distribution of clusters within a branch representing organised city in Prague, derived from a combined dendrogram.

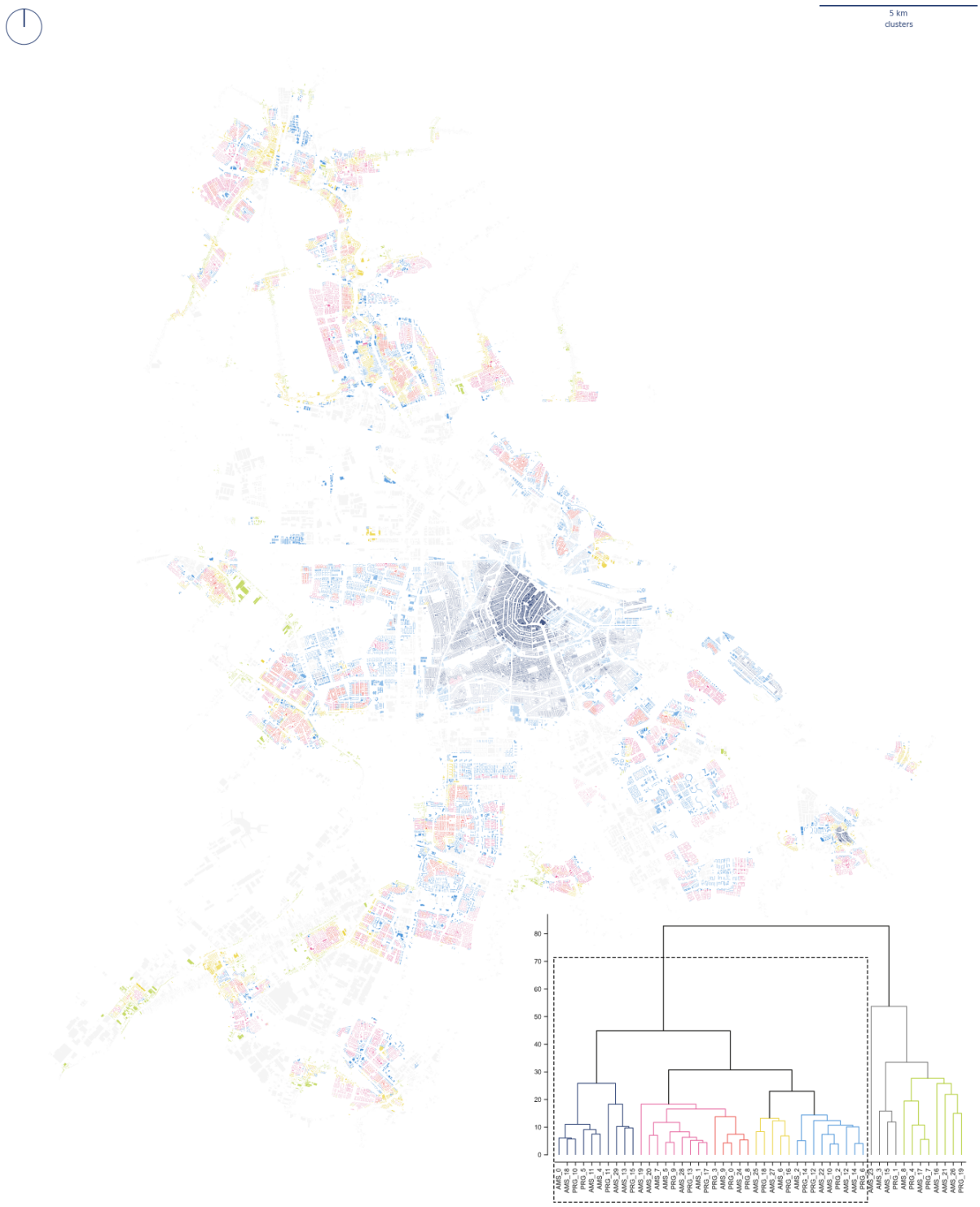


Figure 8.40: Spatial distribution of clusters within a branch representing organised city in Amsterdam, derived from a combined dendrogram.

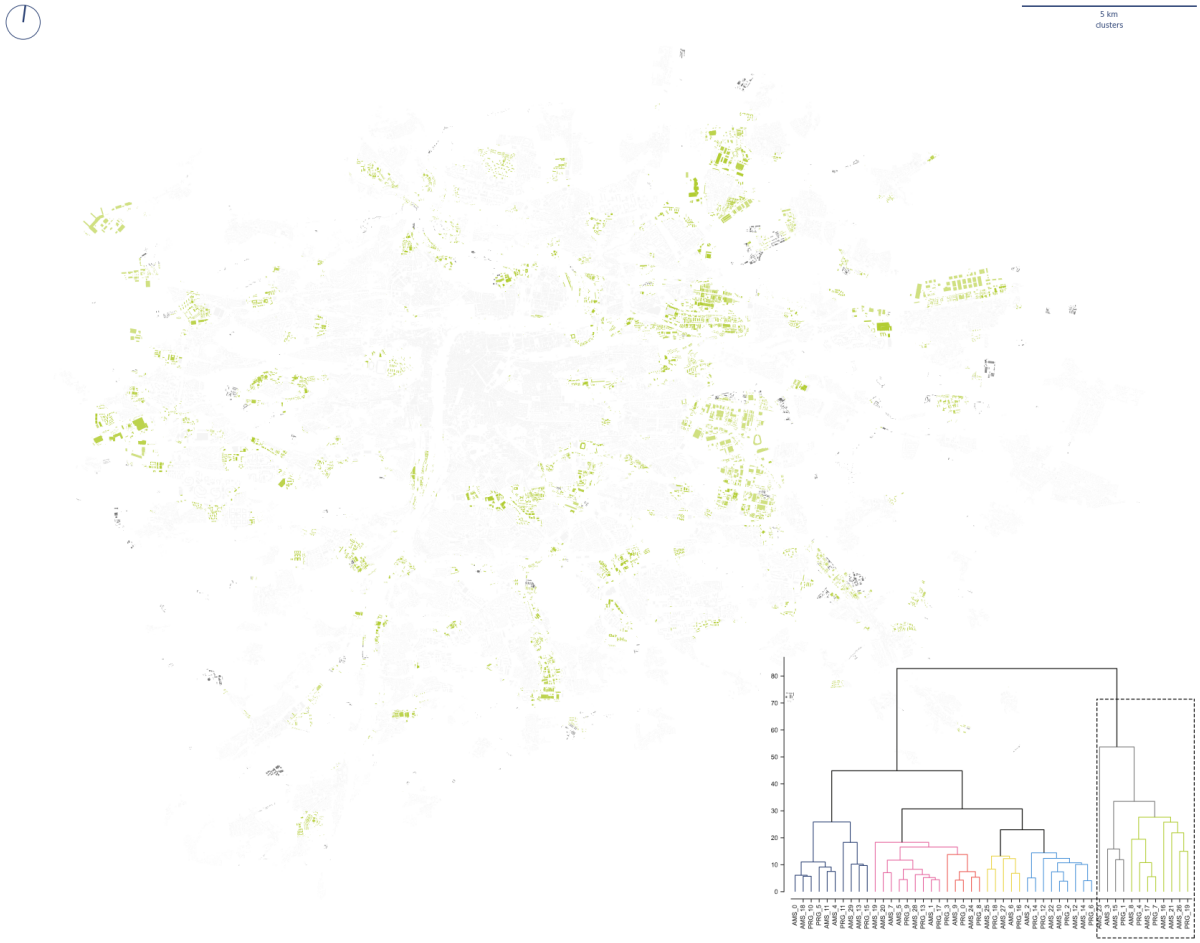


Figure 8.41: Spatial distribution of clusters within a branch representing unorganised city in Prague, derived from a combined dendrogram.

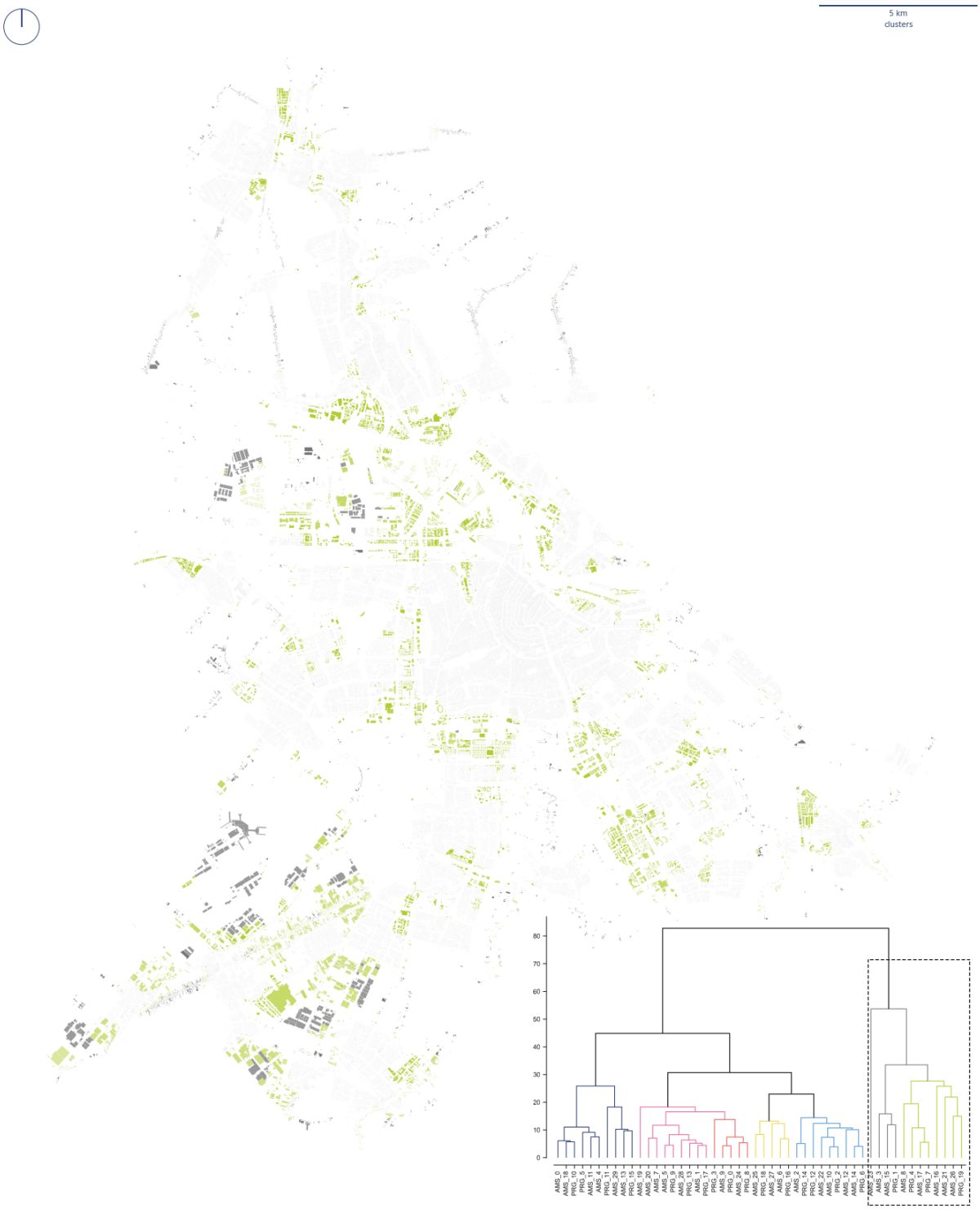


Figure 8.42: Spatial distribution of clusters within a branch representing unorganised city in Amsterdam, derived from a combined dendrogram.

However, there are some differences in branching. The resulting tree reshuffled few of the clusters and slightly reorganised branches. It will likely happen when we add more cases until the taxonomy will get more saturated. As shown on figure 8.43, the reshuffle is, however, relatively minimal and clusters which were associated to the same branch in individual dendrograms tend to stay together in the combined one as well. Furthermore, the overall structure of the dendrogram and the relationship between different branches remains relatively consistent.

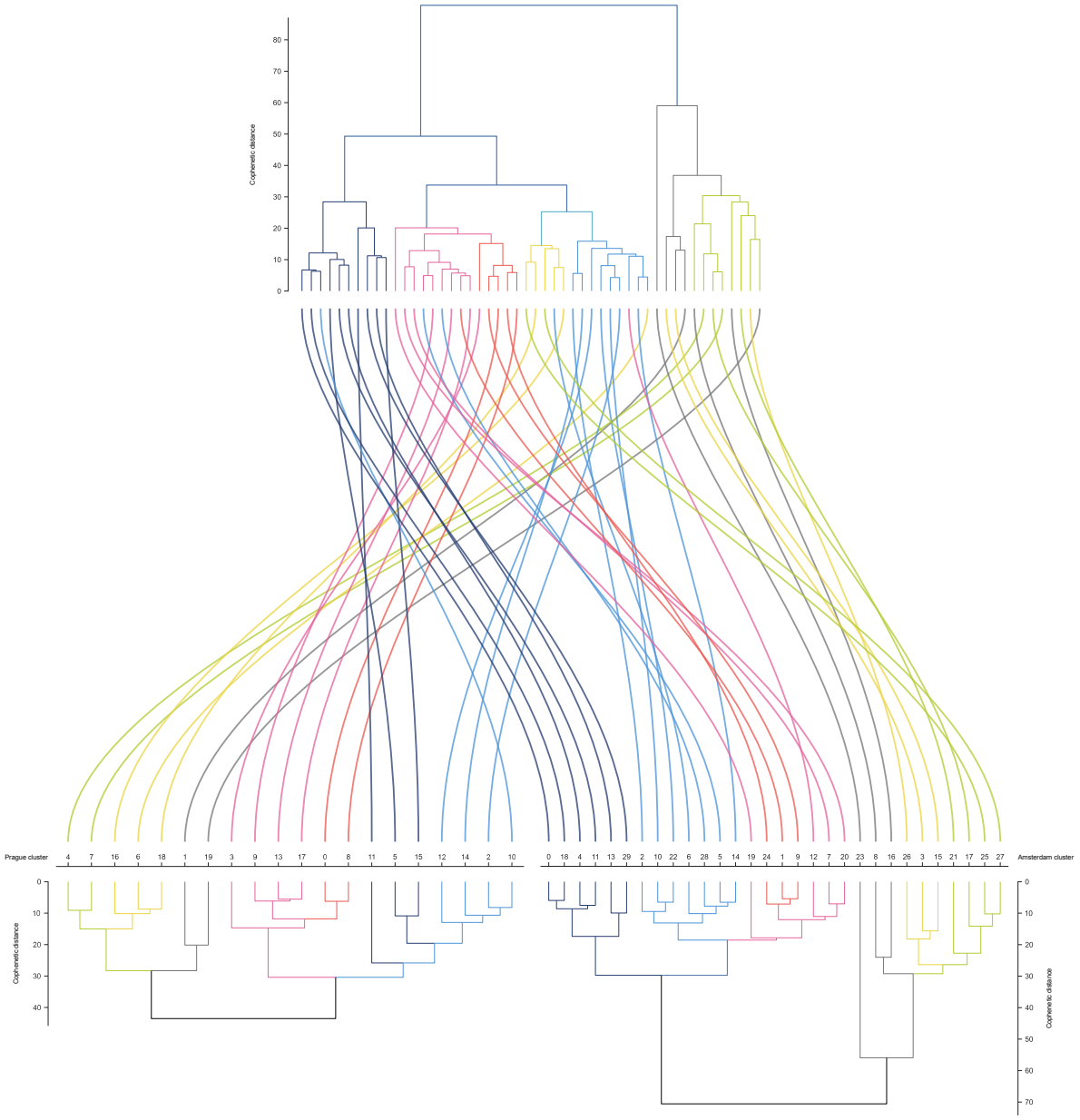


Figure 8.43: Diagram illustrating the flow of clusters between branches of individual dendrograms and the combined one.

There are a few differences worth noting. The Prague clusters 6 and 16 has moved from the unorganised city branch to organised. Both clusters are capturing mostly single-family housing built on the steeper hills causing the disruption of a standard pattern. The similar

shift happened to Amsterdam clusters 27 and 35, composed of low-density development around roads. All are now classified among other single-family tissue types, which seems to be a more appropriate place. Prague's cluster 10 (fringe-like edges of compact development) is now closer to the homogenous compact development making together more reasonable branching. There is a reshuffle in low-density branches of Amsterdam, but as these are generally similar, it is not a big issue. The inclusion of Prague's low-density cluster most likely caused new regrouping based on additional information. Generally, it looks that those few reshuffles are actually making the taxonomy more robust and eliminate the potential issues due to the small number of OTUs in a single case study.

As shown by the results above, there does not seem to be an issue with combining morphometric tissue types identified independently of each other into a singular taxonomy. The differences are expected as the current number of tissue types, i.e. OTUs for the derivation of taxonomy is still low, and the taxonomy itself is not yet saturated. However, it is assumed that further expansion of the pool of case studies could eventually lead to the stabilisation.

8.4 Summary of taxonomy and validation

This chapter, the last of the core chapters, focused on the development of a method able to derive a taxonomy of urban form and validation of the whole morphometric assessment proposed in this research.

The development of taxonomy takes the form of Ward's hierarchical clustering of initial morphometric tissue types defined in Chapter 7, represented by their centroids (i.e. mean values of contextual morphometric characters). The results show a very meaningful identification of relationships between individual tissue types and allow a rich analysis of the structural composition of the city of Prague.

The validation of the morphometric assessment was done in three steps. The first one compares the spatial distribution of clusters and taxonomic branches to additional proxy data which are known to reflect the patterns of urban development. This research used data on historical origin, predominant land use patterns and finally, qualitative municipal typology of urban form. Data on all three showed a significant relationship between them

and identified clusters, indicating that from the perspective of the ability of classification to reflect expected patterns of these proxy layers, morphometric urban tissue types are a valid concept.

The second validation step was the application of the whole method on the case study of Amsterdam, NL, which on the one hand shares the morphological richness of Prague and on the other was built in a different planning context. The resulting urban tissue types and their taxonomy tend to behave in a similar manner as was observed in Prague, indicating that the method could be transferable from one context to the other.

The final part of the chapter is assessing the proposed method from the perspective of its expandability and compatibility of results from Prague with those from Amsterdam. In other words, it tested whether it can become a methodological foundation of the taxonomy of urban form. The combined taxonomy of tissue types shows a high degree of consistency if compared to individual ones. Furthermore, the certain reshuffle which happened mostly reflected the previous misalignment of tissue types, rendering the final taxonomy more robust than those done independently.

Chapter 9 will discuss what the implication of these results and where the research should head next is.

Chapter 9

Synthesis

Where chapter 1 introduced the issue of classification of urban form, chapters 2, 3, and 4 provided background knowledge allowing formulation of hypothesis and research questions in chapter 5 to be answered in chapters 6, 7, and 8, the final chapter aims to put the whole research back into the broader context. Therefore, the following sections provide reflections and discussion of the value of the thesis and its components, their limitations and potential further research.

The chapter first links the proposals of this thesis to the context described in the background chapters and specific gaps of knowledge and limitation identified there. Then it discusses the relation of presented results to the hypothesis and each of the research questions. Having summarised the research itself, the next section discusses its limitations followed by the proposals for further research.

9.1 Reflections and discussion

The background chapter (2, 3, and 4) provided the context of specific aspects of urban morphology to date. It is hence only fair to examine where the method proposed in the core chapter (6, 7, and 8) sits in relation to all three chapters and the key gaps of knowledge they identify.

9.1.1 Classification of urban form

Classification can be done in a multitude of ways, as shown in chapter 2. However, a classification which reflects the aim of this thesis should follow seven principles of the *Optimal Classification Model* (see section 2.1.2 for details). As shown in Table 2.1, none of the existing models presented in the published literature reflects all seven principles. The method proposed in this research attempts to be the first one which does.

The method, as a combination of cluster analysis delineating urban tissue types and subsequent hierarchical classification deriving a taxonomy, is *exhaustive* as it covers all observations within the dataset (i.e. all tessellation cells) leaving none unclassified. Furthermore, classification is *mutually exclusive* - none of the features is member of more than one class at the same time. The method is unsupervised, data-driven, hence purely *empirical*. The final structure of the taxonomy is hierarchical, allowing flexible interpretation following different branches, enabling classification of form from 2 to N classes, depending on the requirements of each subsequent study. The data input for cluster analysis has 296 variables, reflecting the spatial distribution of 74 primary morphometric characters, based on relational framework of urban form, which indicates that the method can be considered comprehensive. That, in turn, minimises the selection bias and attempts to reflect the complexity of urban form. At the same time, the classification is done on the level of individual buildings/tessellation cells. That brings high granularity of the result, making the method *detailed*. Finally, due to its algorithmic nature and Python backend (see section 5.3 and Annexe 2), the method is *scalable*. While it is currently shown on metropolitan areas composed of up to 250 000 features, the method itself has the potential to scale further up.

There are similarities between the proposed method and some of the existing. The principle of tissue type delineation is similar to Araldi and Fusco (2019). Both methods measure primary characters on selected elements (street segment in the case of Araldi and Fusco (2019)), then include contextualisation layer (LISA/ILINCS patterns in the case of Araldi and Fusco (2019)), which is used in cluster analysis determining types. Although the direct quantitative comparison of Araldi and Fusco's method with the presented one has not been done, the proposed method is expected to be able to capture more granular differences in urban patterns due to the smaller unit of analysis (street segment vs tessellation cell), provides a higher granularity of the typology (9 types in French Riviera vs 20

in Prague and 30 in Amsterdam) and the hierarchical layer of taxonomy (compared to the flat model used by Araldi and Fusco (2019)). Similarly to Dibble *et al.* (2017), Dong *et al.* (2019), Li *et al.* (2020) or Serra *et al.* (2018), the resulting classification is hierarchical, but there are generally more differences between mentioned works and the current one than similarities, primarily due to the automatic delineation of tissue types used as the OTU. Granularity and extent are similar to Berghauser Pont *et al.* (2019), who do not combine elements into a single classification and generally use only a small number of variables. Whilst missing the direct comparison of results, the proposed method provides a higher granularity of typology (i.e., more types) and the recognition of similarity between the types (hierarchical classification), not present in the assessed research. In the case of Amsterdam, as a case study shared by this research and Berghauser Pont *et al.* (2019), the clustering in the city centre area presented in this thesis follows the historical development of Amsterdam more closely (while comparing with plot types). However, it is important to acknowledge that the work of Berghauser Pont *et al.* (2019) is based on different conceptual assumptions and has a different aim; therefore the direct comparison is not entirely possible. Notably, the proposed method includes more morphometric characters than other studies available in the literature to date.

9.1.2 Numerical taxonomy

The proposed method adheres to the seven principles of numerical taxonomy outlined in chapter 3, where applicable to urban morphology. Although some results may indicate the possibility to infer the phylogenetic relationship between urban tissue types from the taxonomy, it would be a bold statement at this point.

The classification of urban form is, in principle, classification on the population level. Where in biology population would reflect a group of individuals belonging to the same species within the same geographical area, the population in case of urban form is analogous to urban tissue. Therefore, the whole process is seen as a *mixture problem* (see section 3.1.2.2). The Operational Taxonomic Unit is an urban tissue type, which itself is recognised as a population of fundamental elements of urban form defined by inner morphometric homogeneity. Such an approach limits the effect of MAUP as no predefined artificial aggregation or boundary is in play. Moreover, it allows capturing patterns in different planning contexts, following different structural principles.

9.1.3 Urban morphometrics

The context of chapter 4 and classification of quantitative studies in urban morphology (section 4.2), this work can be categorised into *comparison* based on purpose, using 296 characters, with the smallest scale of grain (1) and the largest scale of extent (10), occupying the bottom right corner of figure 4.2, as shown in figure 9.1.

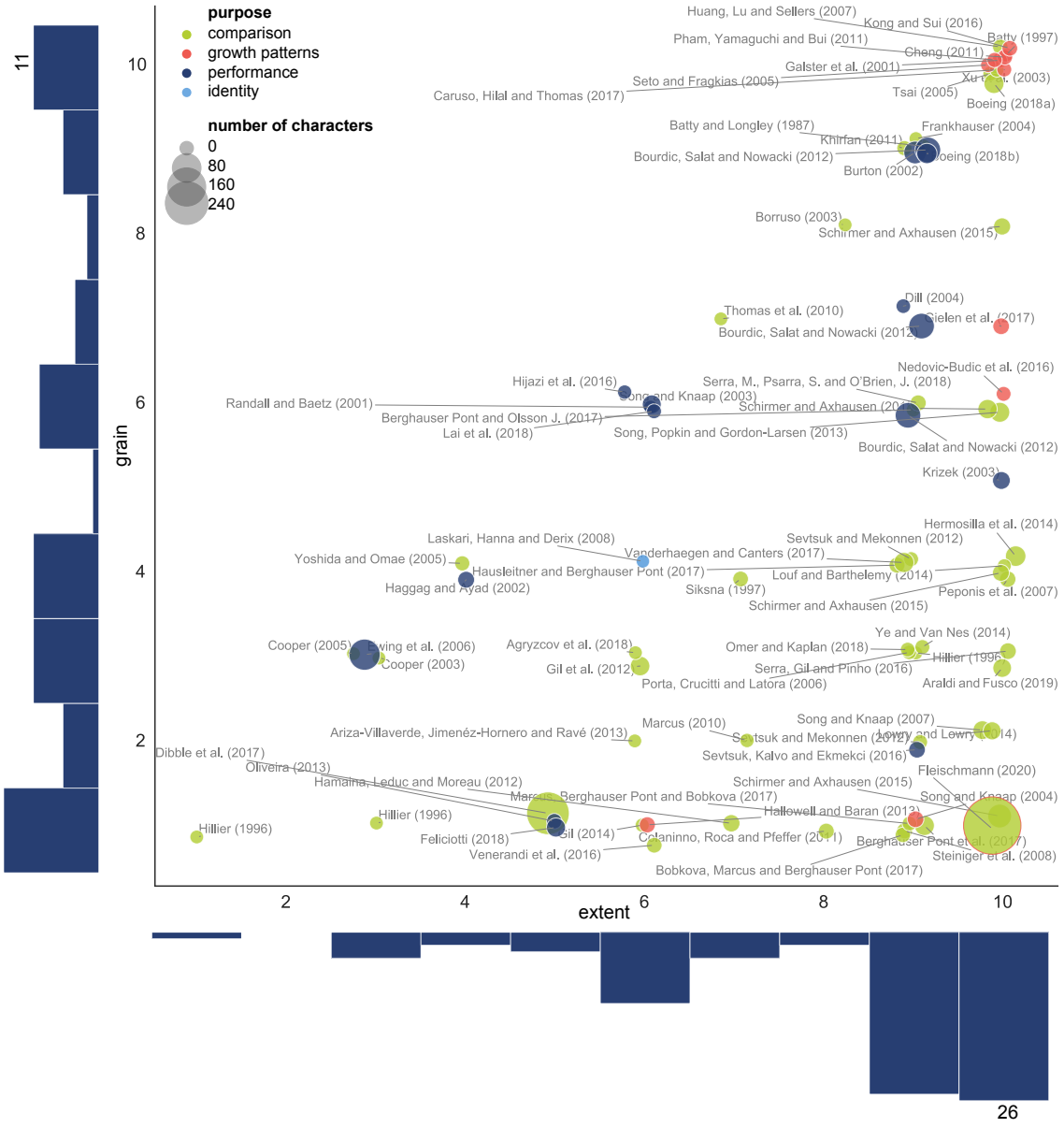


Figure 9.1: Classification of Literature with an inclusion of this research (bottom right - highlighted). Predominantly quantitative studies in urban morphology classified according to grain scale (Y axis), extent scale (X axis), purpose (colour) and number of urban form characters (size). The histograms show a relative balance in terms of scale of grain and a tendency towards large scales of extent. Note: placement of points is jittered to minimise overlaps.

Terminologically, the whole research follows *Index of Element* principle in the naming of morphometric characters, limiting ambiguity and *nicknaming* to a minimum. The

classification of characters into categories is illustrated in figure 9.2. The most common category of primary characters (24 characters) is a dimension, the most simple one. Shape (16), spatial distribution (15) and intensity (11) are relatively balanced. Connectivity is naturally sparse since the required scale of characters does not allow for characters derived from large networks. Diversity is present in two examples only, but that is due to the research design. Looking at the distribution of contextual characters used in the cluster analysis (figure 9.2b), the fact that interquartile range, interdecile Theil index and Simpson's diversity index are all meta characters belonging to diversity category changes the balance with a significant prevalence of diversity over other categories (see section 7.1.2.2 for details).

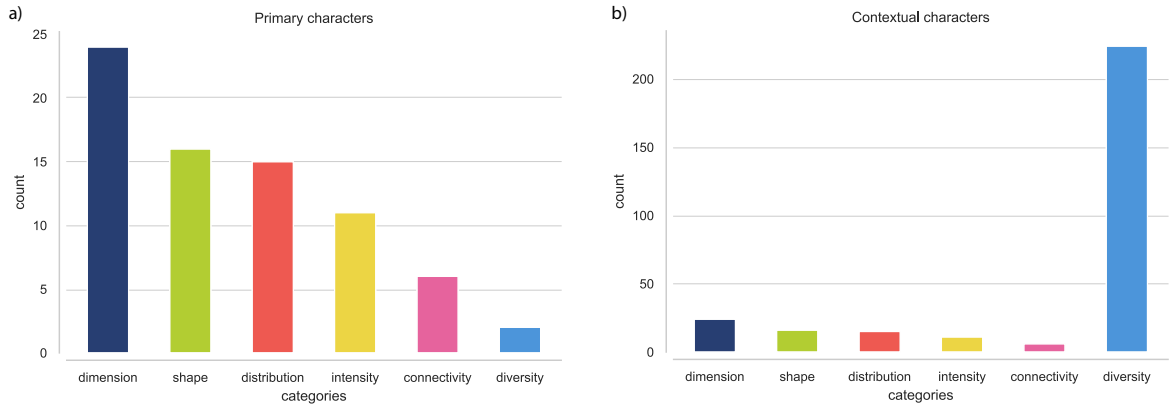


Figure 9.2: Number of morphometric characters per category as used within different stages of this study. a) shows counts of primary characters per category; b) shows counts of contextual characters per category.

9.1.4 Hypothesis and research questions

Whilst chapter 5 formulated the hypothesis and a series of research questions, the following chapters tried to answer them and validate the results.

The main hypothesis of the research related to the applicability of methods of morphometrics and numerical taxonomy in the context of urban morphology remains plausible. The results do not indicate that there is a ground for its rejection, and it is possible to conclude that the hypothesis still holds. That means that there is clear potential in the application of such methods and further research in the topic as the initial work of Dibble *et al.* (2017) as well as this one are explorative. Both works, although using different implementations,

conclude the same - numerical taxonomy is applicable in urban morphology.

The overall method proposed in chapters 6, 7 and 8 provides an answer to the main research question, *how to adapt methods of numerical taxonomy to study of urban form.* Subdividing the answer into relevant supplementary research questions:

> SRQ1 - *What are the fundamental morphometric elements and how to model their relationship?*

Chapter six proposes three fundamental elements based on two sources of input data. Buildings, represented as footprint polygons and street network, represented as centreline geometry, which can be further split to nodes and edges capturing intersections and streets, are direct inputs. Furthermore, buildings can be used to generate morphological tessellation, the smallest spatial unit taking the role of the smallest spatial unit of analysis within the morphometric assessment. The relationships between all elements are then captured by the relational framework of urban form, structurally following overlapping semi-lattice argued for in Alexander (1966).

> SRQ2 - *What is the optimal Operational Taxonomic Unit of urban form and how to identify it in continuous urban fabric?*

The Operational Taxonomic Unit, as understood within this study, is an urban tissue type for the level of taxonomy. The way of identifying it in continuous urban fabric follows the mixture problem, i.e. delineation of populations based on morphometric profiles of individual features (building/tessellation cell entity). The method itself then measures primary and contextual characters, which are then used within cluster analysis (employing Gaussian Mixture Model method) defining tissue types (Operational Taxonomic Units).

> SRQ3 - *What are the taxonomic characters describing urban form?*

Chapter 4 and Appendix A4 provide a comprehensive overview of possible morphometric characters which can be used in the study of urban form. The set of *taxonomic* characters used within this thesis is a smaller subset primarily designed to delineate urban tissue

types and characterise them. However, it is worth noting that the completeness of a set of taxonomic characters is always an illusion to a degree. An infinite number of characters can theoretically describe each urban tissue and even each individual element.

> SRQ4 - *How to determine the taxonomic relationship between OTUs to derive taxa of urban form?*

The answer to the last supplementary question, which completes the answer to the main research question is proposed in chapter 8. The application of Ward's hierarchical clustering in the derivation of the taxonomic relationship between urban tissue types seem to be plausible not only based on the results of the validation in chapter 8 but also considering relevant literature tending to employ the same algorithm (Dibble *et al.*, 2017; Serra *et al.*, 2018; Jochem *et al.*, 2020).

One question remains. What is the relation of morphometric urban tissue type (i.e. clusters) and urban tissue as it would be identified using qualitative methods of traditional urban morphology (Oliveira, 2016)? If we assume that clusters precisely capture what would be an urban tissue type, a single urban tissue would be a contiguous patch of morphological tessellation belonging to the same class. However, the question is still hard to answer as it would require a proper typo-morphological or historic-geographical study of one of the case studies and comparison of results. The closest available dataset is the municipal typology of Prague's urban fabric, which even though it has its own limitations (see section 8.2.2.1.2.3), indicates a high association between the two. What plays a significant role in this relationship is the definition of homogeneity. Tissue is *a distinct area* (Kropf, 2017), but the definition does not specify how much distinct is should be. Therefore it is not straightforward to link tissue and cluster in a definitive way. Is tissue a contiguous area of a cluster as delineated in chapter 7? Or is tissue rather a contiguous area of a branch of dendrogram defined in chapter 8? A theory does not give us an answer as both can be correct. It always depends on the required resolution of each study. However, the flexibility which comes with the dendrogram helps in finding this connection.

9.2 Limitations

Whilst the method proposed in this thesis and its components appear to be a promising new addition to urban morphology; there are some limitations.

The general limitation which affects all steps of the method is data availability. It is not complicated to find case studies offering the data in required quality and detail, but, indeed, data of this level of precision are not available everywhere around the world. That is true, especially for building height parameters. Having all data, as outlined in chapter 6, is the ideal situation. In the real world, the situation might be less optimal than that, so pre-processing procedures have to be employed before performing the analysis itself. Furthermore, in cases with sub-optimal data which can not be pre-processed to the required level of detail or which do not have known building height, the method needs to be adapted and further validated.

Going to the individual aspects of the proposed method, the first limit of an assessment of morphological tessellation presented in chapter 6 is that the cadastral parcels in Zurich, which were loosely treated as 'plots', are solely based on land-ownership. That causes a discrepancy between the generated tessellation and the cadastral layer, which includes multi-building plots. However, as only 21% of plots are affected and results are reported for both groups, it is believed that the presented method is robust enough to provide relevant results.

Furthermore, whilst it is true that a morphological tessellation can be generated directly from a building layer alone, it cannot be created from any building layer, as this needs to comply to certain quality requirements, which links back to the first limitation above. Notably, since the method sees every feature of GIS layer as an individual input for tessellation, it is important not to have buildings composed of multiple features each representing, for example, different heights or different parts of the same (as in the case of British Ordnance Survey). Similarly, it is important not to have different independent buildings collapsed into a single simplified feature (as in the case of vast portions of OpenStreetMap).

The key component of the method is the selection of morphometric characters used for both delineation of tissue types and hierarchical clustering. Although the set of primary

characters is defined based on a thorough procedure (see section 7.1.2.1.1), its verification comes only indirectly through the validation of the whole method. There are likely characters which are not included, but which would help in classification. Furthermore, due to computational demands, it was not tested whether the full set is required. However, the results indicate that the used set of characters could be perceived as valid for the purpose.

A similar situation is with contextual characters. The method uses four meta contextual characters to derive spatially lagged information on the spatial distribution of values of primary characters. Although the selection is able to deliver expected results, other options may be more suitable. However, that is left for further research.

Gaussian Mixture Model technique used to delineate morphometric tissue types does reflect the nature of the clustering problem but comes with certain limits. First, the method is non-spatial, which mean that it does not include any contiguity constraint. That can be an issue, since the resulting cluster may be in some cases scattered, especially along boundaries of distinct tissues or in heterogeneous areas. While this issue is mostly mitigated by the design of contextual characters which are an input of the cluster analysis, it is not perfect and there are some features which could be seen as mislabeled from the perspective of contiguity. However, available methods of spatially constrained clustering are not yet efficient enough to deal with datasets of this size.

The computational efficiency also limits GMM to a degree. Compared to more straightforward methods as K-Means, a probabilistic component of GMM significantly slows down the computation and it is a question how scalable the method will be when used on larger areas composed of millions of features. Although there are potential performant implementations based on distributed computing on GPU (Rocklin, 2015; Bingham *et al.*, 2019), it may become a weak link in the whole process of scaling up. However, as shown in chapter 7, there is still the potential to train model on a sample of the data.

Furthermore, similarly to K-Means, GMM uses an initial random seed, which affects the final result. While the effect of the seed is not significant, there is some, present mostly in areas where two tissues meet. The resulting boundary between tissue types should hence be seen as fuzzy, since different random seeds place is in a similar, but not the same position.

The validation has two limitations - one related to additional data, other related to the additional case study. As mentioned in chapter 8, there is no ground truth with which the results could be compared. Therefore, the first step of validation is using data which should indirectly reflect changes in urban patterns. Although the selection is representative, there could be other sources of data more appropriate, which would give us better, more direct, understanding of the performance of the proposed method. One such layer could be the qualitative study of the same case study already mentioned above.

The second limitation is affecting the validation of transferability and extensibility of the method. Since this research focuses solely on the European context and selects two cities of a historical origin, its validity in other contexts is still unknown. Although it is assumed that it should perform similarly in Asian, African or North American cities with different spatial logic, this remains to be tested.

Finally, the scope of this work restricted to minimal data input comes with the limitation regarding the ability to reflect other dimensions of the built environment. None of the characters used within the method captures green and blue space or land use, among other aspects of cities, all of which may affect the outcome of clustering and taxonomy. The significance of such an effect is unknown and will likely depend on individual cases. However, it is necessary to acknowledge that differences in urban patterns reflected, for example, by the variable density of natural features are not captured by the presented model. Another consequence of the restriction of data input is the exclusion of the plot layer, which is partially substituted by morphological tessellation. While morphological tessellation can provide analytical information on the scale of the plot, the method is not able to capture morphological characters reflecting the position of a building on a plot (among others).

9.3 Applicability of morphometric data

The morphometric data resulting from the proposed method offer three layers of applicability: 1) morphometric characterisation of urban environment based on 370 individual measurable characters, 2) delineation of urban tissue types, and 3) taxonomy of tissue types. Each offer different kind of information which can be used to answer (or to help

answer) different questions.

The raw morphometric values, linked to either morphological tessellation or relevant elements of urban form (e.g. street network nodes), provide a characterisation of the built environment from a broad range of perspectives. As such, they could become an input of various predictive models and semi-morphological studies, aiming at understanding relations between different aspects of urban form and other facets of life. Furthermore, some of the characters literature identifies as proxies for urban form resilience (Feliciotti, 2018) or other performance-based indicators (e.g. sustainability (Bourdic *et al.*, 2012)), allowing for a more narrowly focused assessment of urban fabric.

The delineation of urban tissues (i.e. flat classification) is the major step in complexity reduction since it combines a large number of fundamental elements into a small set of tissue types. The types can be consequently employed as a unit of analysis when asking about the relation of urban patterns and other aspects of the urban environment. Where the layer of raw morphometric values provides information on individual structural aspects of urban form, the layer of tissue types captures coherent patterns as their combination.

Finally, the taxonomy identifies similarity between tissue types which in turn brings flexibility to the classification. If we take the Amsterdam case study as an example, we may ask about the relationship between tissue types and AirBnB locations (leaving aside other aspects influencing such the locations for the sake of the illustration). It is likely that using the initial 30 clusters may prove suboptimal since they are covered by each type is relatively small and within this question hard to interpret. Therefore we may want to reduce the number of classes, which can be flexibly done by moving through the dendrogram upwards. The final analysis can therefore look into the location of Airbnb listings within, e.g., five macro taxa of tissue types. The level of aggregation into higher-order taxa depends on the research question asked and within the proposed model can be adapted to the specific needs.

9.4 Further research

The avenue of further research based on the findings presented in this thesis is wide. One direction could further study different aspects of the method to eliminate some of its cur-

rent limitations, while the other could focus on the applicability of urban morphometrics and numerical taxonomy in urban science.

Starting with the first direction and proposed morphological tessellation, further research should resolve the question of the external boundary of tessellation which is currently defined as a static 100m buffer around building footprints. This decision likely causes edge effect, where edge cells are significantly larger than they should be. The optimal solution could be a limitation of tessellation by the pre-defined known boundary of the built-up area. Alternatively, an adaptive buffer capturing the different scales of urban patterns could be developed. Moreover, further research should focus on the question of the exact meaning and variation of topological distance on morphological tessellation and its definition for specific purposes. The question of how many topological steps should be used for the analysis of urban form does not have a fixed answer. It is expected that it will vary depending on the scope of the research.

Inclusion (or even elimination) of additional morphometric characters into the set of primary characters is an expected evolution of the method. Furthermore, the set should be adapted to different data sources, e.g. with missing height attribute of different level of detail to expand its applicability. That also entails Earth Observation. One of the potential directions could attempt to derive morphometric profiles based on EO multispectral raster data, which would radically expand the applicability of the method.

Similarly, contextual characters should be further studied to ensure better interpretability and eliminate potential issues related to binning within Simpson's diversity index.

Considering the limitations of GMM outlined above, further research should focus on the efficiency of the implementation needed for studies on a larger scale. Also, as shown in the case of Amsterdam, Bayesian Information Criterion might not be the optimal method to determine the number of clusters and other options (e.g. clustergram (Schonlau, 2002)) could be tested alongside with other clustering algorithms besides GMM.

Finally, as already briefly discussed in the previous section, the method could be further validated using other additional data and especially using comparative analysis with qualitative typo-morphological and/or historic-geographical studies of urban form alongside the application of the method in different geographical contexts.

The second major direction of further research can build on the methodological foun-

dations presented in this thesis and focus on different questions using the information obtained from numerical taxonomy. One of them is the application if the method to eventually build an atlas of urban form, containing urban tissue types from a wide set of case studies, hierarchically categorised based on their phenetic similarity. Such an atlas could provide a global repository of urban patterns, a counterpart of atlases known in biology (e.g. an atlas of birds or an atlas of fungi).

When it comes to the application of the method on specific case studies, a large number of observations can be made based on morphometric values backing the whole method. Each recognised tissue type as well as each branch of taxonomy have rich morphometric profiles. The combination of all measured values within each cluster gives an abundant description of various morphological aspects which can be further analysed and eventually even turned into form-based regulatory plans directing further development.

Further work should also look into the question of character selection and importance. Very similar results can likely be achieved with a reduced set of characters instead of 296 applied to presented case studies. However, such a reduction would necessarily be case-specific, reflecting the local nature and peculiarities of urban form. In other words, it is assumed that character importance will vary across cases and using, for example, the top 20 characters identified based on the Prague dataset to delineate clusters in both Prague and Amsterdam would likely affect the latter significantly more than the former. Such an effect would be even more pronounced in a different geographical context. Therefore, further work should expand the set of case studies to cover different historical, cultural and geographical contexts and analyse the character importance within the complete pool as well as independently per each case. The result could provide insight of two types. First, it can limit the set of morphometric characters that should be measured, which would lower computational demands. Second, it may identify differences between urban fabrics of different cities and indicate the variation between underlying rules influencing their formation and transformation.

The exclusion of certain data inputs within this research, notably plots and natural infrastructure, opens a potential pathway of research building on top of the proposed framework and including measurable characters based on additional data. That may overcome the limitations induced by the scope of the work pointed out above and develop a complex characterisation of the built environment that goes beyond the morphometric description

of minimal features representing urban form.

9.5 Conclusions

There are two different perspectives when it comes to the study of patterns of development of cities - relational and descriptive. This research belongs to the second one, to urban morphology, trying to describe urban form as it is and as it was formed. At the same time, it aims to support the other perspective. For any study of relation of urban form and other aspects of life, from economy, wellbeing, happiness, to social deprivation, the core component is the description of urban form, i.e. the way we are able to capture and characterise it. This work offers one approach, exploiting the abundance of geospatial data, building tools to make sense of it and proposing a detailed quantitative description of individual elements of urban form, patterns they form and similarity between them.

The introduction stated that urban morphology or urban studies in general face three problems. The field is unable to describe form comprehensively enough, its methods of identification and systematisation of homogenous areas lack either detail, granularity or scalability, and data-driven classification is in its infancy. This work proposes methodological framework which may eventually lead to the minimising or even elimination of all three of them.

The overall aim of this thesis was to propose a method of derivation of data-driven taxonomy of urban form patterns. Considering the contents presented in previous chapters and results of validation, it seems to be fulfilled. The method of analysis of urban form based on urban morphometrics and numerical taxonomy allows for unsupervised delineation of urban tissue types or using the generalised terminology of the first chapter, urban form patterns. Furthermore, it defines the phenetic relationship (similarity) between all of them forming the basis for the taxonomy of urban form. All that within the scope defined in section 1.3, thus using minimal data input (building footprints with a height attribute and street network) capturing solely the fundaments of urban form.

Based on the two case studies, the method seems to be transferable to other geographical and planning contexts. The question remains whether the same performance will hold outside the European tradition. With the proposal of morphological tessellation as the

smallest spatial unit and relational framework of urban form, the work builds a foundation for the transfer of numerical taxonomy to urban morphology and a broader application of urban morphometrics as such. That is further supported by the release of software tools, allowing the reproduction of the whole work in a flexible form of an open-source Python package, allowing further development within the research community.

Although the quantitative science of urban form is rapidly evolving, we still have a long way to go. I believe that this thesis can be a small but valuable addition to the pursuit of a comprehensive quantitative description of urban form and one step towards capturing its inherent complexity.

Appendix A4: Supplementary material for chapter 4

A4.1 Table of Urban Form Characters

The full version of the Table 4.2 (see section 4.3.2).

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Dimension	Area	Block	S	S	Dibble et al. (2017)
Dimension	Built-up area	Block	S	S	Gil (2012)
Dimension	Depth	Block	S	S	Sevtsuk (2016)
Dimension	Gross floor area	Block	S	S	Gil (2012)
Dimension	Height	Block	S	S	Hermosilla (2014)
Dimension	Layers (number of floors)	Block	S	S	Gil (2012)
Dimension	Length	Block	S	S	Gil (2012)
Dimension	Longest Diagonal	Block	S	S	Feliciotti (2018)
Dimension	Perimeter	Block	S	S	Gil (2012)
Dimension	Private space area	Block	S	S	Gil (2012)
Dimension	Public space area	Block	S	S	Gil (2012)
Dimension	Volume	Block	S	S	Hermosilla (2014)
Dimension	Width	Block	S	S	Gil (2012)
Dimension	Area	Building	S	S	Colaninno, Cladera and Pfeffer (2011)
Dimension	Bounding box Area	Building	S	S	Hamaina, Leduc and Moreau (2012)
Dimension	Core Area Index	Building	S	S	Colaninno, Cladera and Pfeffer (2011)
Dimension	Floor area	Building	S	S	Schirmer and Axhausen (2015)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Dimension	Height	Building	S	S	Berghauser Pont, M. et al. (2017)
Dimension	Influence zone area	Building	S	S	Schirmer and Axhausen (2015)
Dimension	Length	Building	S	S	Hamaina, Leduc and Moreau (2012)
Dimension	Length fo centrelines	Building	S	S	Schirmer and Axhausen (2015)
Dimension	Number of floors	Building	S	S	Ye and van Nes (2014)
Dimension	Number of stories	Building	S	S	Schirmer and Axhausen (2015)
Dimension	Volume	Building	S	S	Hamaina, Leduc and Moreau (2012)
Dimension	Width	Building	S	S	Hamaina, Leduc and Moreau (2012)
Dimension	Weighted Average Height	Buildings in Block	S	S	Dibble et al. (2017)
Dimension	Weighted Height	Built Fronts	S	S	Dibble et al. (2017)
Dimension	Bounding box Area	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Bounding box Length	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Bounding box Perimeter	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Bounding box Width	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Convex Hull Area	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Convex Hull Perimeter	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Enclosing circle Area	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Enclosing circle Perimeter	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Enclosing circle Radius	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Perimeter	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Width along centerline	Composite	S	S	Schirmer and Axhausen (2015)
Dimension	Convex Hull Area	Courtyard	S	S	Schirmer and Axhausen (2015)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Dimension	Depth	Mean Maximum Gravity block	S	S	Sevtsuk (2016)
Dimension	Widht	Mean Maximum Gravity block	S	S	Sevtsuk (2016)
Dimension	Width	Open space	S	S	Araldi and Fusco (2019)
Dimension	Area	Patch	S	S	Vanderhaegen and Canters (2017)
Dimension	Edge length	Patch	S	S	Vanderhaegen and Canters (2017)
Dimension	Perimeter	Patch	S	S	Vanderhaegen and Canters (2017)
Dimension	Area	Plot	S	S	Dibble et al. (2017)
Dimension	Depth	Plot	S	S	Song and Knaap (2007)
Dimension	Extension on Street Front	Plot	S	S	Dibble et al. (2017)
Dimension	Height	Street	S	S	Sevtsuk (2016)
Dimension	Length	Street	S	S	Dibble et al. (2017)
Dimension	Width	Street	S	S	Dibble et al. (2017)
Dimension	Length	Street front	S	S	Schirmer and Axhausen (2015)
Dimension	Length	Street segment	S	S	Bourdic, Salat and Nowacki (2012)
Dimension	Slope	Surface	S	S	Araldi and Fusco (2019)
Dimension	Area	Urban block related street area (UBRSA)	S	S	Hermosilla (2014)
Dimension	Area	Voronoi cell	S	S	Hamaina, Leduc and Moreau (2012)
Dimension	Median Area	Blocks in buffer area	M	M	Song and Knaap (2007)
Dimension	Median Perimeter	Blocks in neighborhood	M	M	Lowry and Lowry (2014)
Dimension	Mesh size	Grid network	M	M	Siksna (1997)
Dimension	Median Area	Plots in neighborhood	M	M	Lowry and Lowry (2014)
Dimension	Area	Sanctuary Area	M	M	Dibble et al. (2017)
Dimension	Cul-de-sac Length	Street network	M	M	Lowry and Lowry (2014)
Dimension	Length	Street network	M	M	Song and Knaap (2007)
Dimension	Length within radius	Street network	M	M	Krizek (2003)
Dimension	Actual Nonresidential Area	Study area	M	M	Song and Knaap (2004)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Dimension	Zoned nonresidential Area	Study area	M	M	Song and Knaap (2004)
Dimension	Area	Patch	L	L	Pham (2011)
Dimension	Length	Street network	L	L	Khirfan (2011)
Dimension	Area	Study area	L	L	Khirfan (2011)
Dimension	Length	Cul-de-sac	L/M	L/M	Song and Knaap (2003)
Dimension	Effective Mesh size	Plot	L/M	L/M	Hausleitner, B. and Berghauser Pont, M. (2017)
Dimension	Diameter	Street network	L/M	L/M	Boeing (2018b)
Dimension	Length	Street network	L/M	L/M	Boeing (2018a)
Shape	Area perimeter ratio	Block	S	S	Gil (2012)
Shape	Compactness index	Block	S	S	Dibble et al. (2017)
Shape	Contour stretch Length	Block	S	S	Vanderhaegen and Canters (2017)
Shape	Fractal Dimension	Block	S	S	Hermosilla (2014)
Shape	Fragmentation	Block	S	S	Feliciotti (2018)
Shape	Length / width Proportion	Block	S	S	Gil (2012)
Shape	Length of Radial profile radial stretch	Block	S	S	Vanderhaegen and Canters (2017)
Shape	Normalise nr of perimeter alterations	Block	S	S	Vanderhaegen and Canters (2017)
Shape	Normalised nr of radial alterations	Block	S	S	Vanderhaegen and Canters (2017)
Shape	Normalised number of contour alterations	Block	S	S	Vanderhaegen and Canters (2017)
Shape	Percentage of countours	Block	S	S	Vanderhaegen and Canters (2017)
Shape	Percentage of perimeter	Block	S	S	Vanderhaegen and Canters (2017)
Shape	Percentage of radials of radial progfile	Block	S	S	Vanderhaegen and Canters (2017)
Shape	Rectangularity index	Block	S	S	Dibble et al. (2017)
Shape	Shape factor	Block	S	S	Louf and Barthelemy (2014)
Shape	Shape index	Block	S	S	Hermosilla (2014)
Shape	Area/Covex hull ratio	Building	S	S	Steiniger (2008)
Shape	Area/perimeter ratio	Building	S	S	Colaninno, Cladera and Pfeffer (2011)
Shape	Compactness	Building	S	S	Schirmer and Axhausen (2015)
Shape	Corners to centroid distance	Building	S	S	Schirmer and Axhausen (2015)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Shape	Elongation	Building	S	S	Steiniger (2008)
Shape	Form factor	Building	S	S	Bourdic, Salat and Nowacki (2012)
Shape	Generalised width/height	Building	S	S	Hamaina, Leduc and Moreau (2012)
Shape	Lines in skeleton only connected to one side	Building	S	S	Schirmer and Axhausen (2015)
Shape	Number of centreline orientations	Building	S	S	Schirmer and Axhausen (2015)
Shape	Number of centrelines	Building	S	S	Schirmer and Axhausen (2015)
Shape	Number of corners	Building	S	S	Steiniger (2008)
Shape	Shape	Building	S	S	Steiniger (2008)
Shape	Shape index I	Building	S	S	Colaninno, Cladera and Pfeffer (2011)
Shape	Shape index II	Building	S	S	Colaninno, Cladera and Pfeffer (2011)
Shape	Size factor	Building	S	S	Bourdic, Salat and Nowacki (2012)
Shape	Squareness	Building	S	S	Steiniger (2008)
Shape	Surface area/footprint	Building	S	S	Yoshida and Omae (2005)
Shape	Surface area/volume	Building	S	S	Yoshida and Omae (2005)
Shape	Volume/footprint	Building	S	S	Yoshida and Omae (2005)
Shape	Volumetric compactness	Building	S	S	Bourdic, Salat and Nowacki (2012)
Shape	Area/bounding box ratio	Composite	S	S	Schirmer and Axhausen (2015)
Shape	Area/Covex hull ratio	Composite	S	S	Schirmer and Axhausen (2015)
Shape	Area/Enclosing circle ratio	Composite	S	S	Schirmer and Axhausen (2015)
Shape	Courtyard area/area ratio	Composite	S	S	Schirmer and Axhausen (2015)
Shape	Number of wings	Composite	S	S	Schirmer and Axhausen (2015)
Shape	Perimeter/bounding box ratio	Composite	S	S	Schirmer and Axhausen (2015)
Shape	Perimeter/Convex hull perimeter ratio	Composite	S	S	Schirmer and Axhausen (2015)
Shape	Volume to façade ratio	Composite	S	S	Schirmer and Axhausen (2015)
Shape	Width to length ratio of bounding box	Composite	S	S	Schirmer and Axhausen (2015)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Shape	Height/width ratio	open space	S	S	Araldi and Fusco (2019)
Shape	Area weighted fractal dimension	Patch	S	S	Vanderhaegen and Canters (2017)
Shape	Area weighted shape index	Patch	S	S	Vanderhaegen and Canters (2017)
Shape	Edge density	Patch	S	S	Vanderhaegen and Canters (2017)
Shape	Fractal Dimension	Patch	S	S	Vanderhaegen and Canters (2017)
Shape	Perimeter/area ratio	Patch	S	S	Vanderhaegen and Canters (2017)
Shape	Shape index	Patch	S	S	Vanderhaegen and Canters (2017)
Shape	Compactness	Plot	S	S	Bobkova, Marcus and Berghauser Pont (2017)
Shape	Compactness index	Plot	S	S	Dibble et al. (2017)
Shape	Degree of openness	Plot	S	S	Bobkova, Marcus and Berghauser Pont (2017)
Shape	Plot frontage/depth	Plot	S	S	Sevtsuk (2016)
Shape	Rectangularity index	Plot	S	S	Dibble et al. (2017)
Shape	Fractal Dimension	Skyline	S	S	Cooper (2003)
Shape	Acclivity	Street	S	S	Araldi and Fusco (2019)
Shape	Corridor effect	Street	S	S	Araldi and Fusco (2019)
Shape	Height to width ratio	Street	S	S	Oliveira (2013)
Shape	Fractal Dimension	Street edge	S	S	Cooper (2005)
Shape	Linearity/Windingness	Street segment	S	S	Araldi and Fusco (2019)
Shape	Fractal Dimension	Axial map	M	M	Ariza-Villaverde et al. (2013)
Shape	Perimenter Shape index	Neighborhood	M	M	Song and Popkin (2013)
Shape	Perimeter Mean Fractal Dimension Index	Neighborhood	M	M	Song and Popkin (2013)
Shape	Perimeter-area fractal dimension	Neighborhood	M	M	Song and Popkin (2013)
Shape	Edge density	Built-up area	L	L	Seto and Fragkias (2005)
Shape	Fractal based scaling behaviour curve	Built-up area	L	L	Thomas (2010)
Shape	Fractal Dimension	Built-up area	L	L	Gielen (2017)
Shape	Shape complexity	Built-up area	L	L	Seto and Fragkias (2005)
Shape	Shape index	Built-up area	L	L	Gielen (2017)
Shape	Area weighted mean fractal dimension	Patch	L	L	Pham (2011)
Shape	Edge density	Patch	L	L	Pham (2011)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Shape	Fractal Dimension	Urban edge	L	L	Batty and Longley (1987)
Shape	Box-counting fractal dimension	Built-up area	L/M	L/M	Boeing (2018b)
Shape	Hausdorff fractal dimension	Built-up area	L/M	L/M	Boeing (2018b)
Shape	Circuity	Street segment	L/M	L/M	Boeing (2018a)
Distribution	Built Front Ratio	Block	S	S	Dibble et al. (2017)
Distribution	Closeness	Block	S	S	Schirmer and Axhausen (2015)
Distribution	Neighboring blocks	Block	S	S	Hermosilla (2014)
Distribution	Openess	Block	S	S	Hausleitner, B. and Berghauser Pont, M. (2017)
Distribution	Orientation	Block	S	S	Gil (2012)
Distribution	Permeability	Block	S	S	Schirmer and Axhausen (2015)
Distribution	Solar orientation	Block	S	S	Gil (2012)
Distribution	Corner position	Building	S	S	Schirmer and Axhausen (2015)
Distribution	Orientation	Building	S	S	Schirmer and Axhausen (2015)
Distribution	Street Distance	Building	S	S	Schirmer and Axhausen (2015)
Distribution	Alignment	Buildings	S	S	Oliveira (2013)
Distribution	Angle betweeen buildings	Buildings	S	S	Hijazi (2016)
Distribution	Distance betweeen buildings	Buildings	S	S	Hamaina, Leduc and Moreau (2012)
Distribution	Party-walls ratio	Buildings	S	S	Hamaina, Leduc and Moreau (2012)
Distribution	Ground openess	Open space	S	S	Hamaina, Leduc and Moreau (2012)
Distribution	Sky openess	Open space	S	S	Hamaina, Leduc and Moreau (2012)
Distribution	Cul-de-sac presence	Plot	S	S	Song and Knaap (2003)
Distribution	Adjacency	Building	M	M	Vanderhaegen and Canters (2017)
Distribution	Proximity	Building	M	M	Colaninno, Cladera and Pfeffer (2011)
Distribution	Fragmentation degree	Built-up area	L	L	Gielen (2017)
Distribution	Index of concentration	Built-up area	L	L	Gielen (2017)
Distribution	Euclidean nearest neighbour distance	Patches	L	L	Pham (2011)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Distribution	Minimum spanning tree Inter-building distance	Building	L/M	L/M	Caruso, Hilal and Thomas (2017)
Distribution	Clustering	Built-up area	L/M	L/M	Galster (2001)
Distribution	Continuity	Built-up area	L/M	L/M	Galster (2001)
Distribution	Nuclearity	Built-up area	L/M	L/M	Galster (2001)
Distribution	Concentration	Housing units	L/M	L/M	Galster (2001)
Intensity	Building coverage area	Block	S	S	Hermosilla (2014)
Intensity	Covered Area Ratio	Block	S	S	Dibble et al. (2017)
Intensity	Floor Area Ratio	Block	S	S	Dibble et al. (2017)
Intensity	Internal Plot Ratio	Block	S	S	Dibble et al. (2017)
Intensity	Internal Ways Ratio	Block	S	S	Dibble et al. (2017)
Intensity	Normalised built-up volume	Block	S	S	Hermosilla (2014)
Intensity	Normalised number of plots	Block	S	S	Hausleitner, B. and Berghauser Pont, M. (2017)
Intensity	Number of buildings	Block	S	S	Gil (2012)
Intensity	Number of courtyards	Block	S	S	Schirmer and Axhausen (2015)
Intensity	Number of plots	Block	S	S	Dibble et al. (2017)
Intensity	Open Space Ratio	Block	S	S	Dibble et al. (2017)
Intensity	Open space ratio	Block	S	S	Gil (2012)
Intensity	Regular Plot Ratio	Block	S	S	Dibble et al. (2017)
Intensity	Volume/UBRSA area ratio	Block	S	S	Hermosilla (2014)
Intensity	Number of plots	Block frontage	S	S	Sevtsuk (2016)
Intensity	Active Fronts to All Fronts Ratio	Building	S	S	Dibble et al. (2017)
Intensity	Active Fronts to Built Fronts Ratio	Building	S	S	Dibble et al. (2017)
Intensity	Building area to buffer area ratio	Building	S	S	Steiniger (2008)
Intensity	Building area/Influence zone area ratio	Building	S	S	Schirmer and Axhausen (2015)
Intensity	Buildings area to buffer area	Building	S	S	Colaninno, Cladera and Pfeffer (2011)
Intensity	Frequency	Building	S	S	Hallowell (2013)
Intensity	Contiguity	Buildings	S	S	Araldi and Fusco (2019)
Intensity	Covered Area Ratio	Composite	S	S	Schirmer and Axhausen (2015)
Intensity	Number of courtyards	Composite	S	S	Schirmer and Axhausen (2015)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Intensity	Coverage ratio	Patch	S	S	Vanderhaegen and Canters (2017)
Intensity	Density	Patch	S	S	Vanderhaegen and Canters (2017)
Intensity	Floor Space Index	Plot	S	S	Berghauser Pont, M. and Olsson J. (2017)
Intensity	Ground Space Index	Plot	S	S	Berghauser Pont, M. and Olsson J. (2017)
Intensity	Use ratio	Plot	S	S	Dibble et al. (2017)
Intensity	Coverage ratio	Projected area	S	S	Yoshida and Omae (2005)
Intensity	Coverage ratio	Proximity band	S	S	Araldi and Fusco (2019)
Intensity	Frequency of buildings	Proximity band	S	S	Araldi and Fusco (2019)
Intensity	Plot Fragmentation	Proximity band	S	S	Araldi and Fusco (2019)
Intensity	Prevalence of building type	Proximity band	S	S	Araldi and Fusco (2019)
Intensity	Normalised number of plots	Street edge	S	S	Feliciotti (2018)
Intensity	Floor Space Index	Tessellation cell	S	S	Hamaina, Leduc and Moreau (2012)
Intensity	Ground Space Index	Tessellation cell	S	S	Hamaina, Leduc and Moreau (2012)
Intensity	Buildings/UBRSA ratio	Urban block related street area	S	S	Hermosilla (2014)
Intensity	Number of buildings	2 minutes driving distance	S	M	Schirmer and Axhausen (2015)
Intensity	Number of deadends	2 minutes driving distance	S	M	Schirmer and Axhausen (2015)
Intensity	Number of intersections	2 minutes driving distance	S	M	Schirmer and Axhausen (2015)
Intensity	Total floorspace	2 minutes driving distance	S	M	Schirmer and Axhausen (2015)
Intensity	Total length of network	2 minutes driving distance	S	M	Schirmer and Axhausen (2015)
Intensity	Number of plots	Accessible radius	S	M	Bobkova, Marcus and Berghauser Pont (2017)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Intensity	Number of plots	Accessible radius	S	M	Marcus, Berghauser Pont and Bobkova (2017)
Intensity	Spatial attraction of plot	Accessible radius	S	M	Marcus, Berghauser Pont and Bobkova (2017)
Intensity	Total floorspace	Accessible radius	S	M	Marcus, Berghauser Pont and Bobkova (2017)
Intensity	Total open space area	Accessible radius	S	M	Hausleitner, B. and Berghauser Pont, M. (2017)
Intensity	Number of cul-de-sacs	Buffer area	S	M	Song and Popkin (2013)
Intensity	Number of intersections	Buffer area	S	M	Song and Knaap (2007)
Intensity	Number of single-family plots	Buffer area	S	M	Song and Knaap (2007)
Intensity	Proportion of road types	Buffer area	S	M	Song and Popkin (2013)
Intensity	Total open space area	Buffer area	S	M	Song and Knaap (2007)
Intensity	Tree canopy area	Buffer area	S	M	Song and Knaap (2007)
Intensity	Built-up area	Buffer area (100-500m)	S	M	Schirmer and Axhausen (2015)
Intensity	Number of buildings	Buffer area (100-500m)	S	M	Schirmer and Axhausen (2015)
Intensity	Number of deadends	Buffer area (100-500m)	S	M	Schirmer and Axhausen (2015)
Intensity	Number of intersections	Buffer area (100-500m)	S	M	Schirmer and Axhausen (2015)
Intensity	Total floorspace	Buffer area (100-500m)	S	M	Schirmer and Axhausen (2015)
Intensity	Total length of network	Buffer area (100-500m)	S	M	Schirmer and Axhausen (2015)
Intensity	Normalised number of housing units	Neighborhood	M	M	Lowry and Lowry (2014)
Intensity	Number of blocks	Neighborhood	M	M	Song and Knaap (2003)
Intensity	Number of blocks	Neighborhood	M	M	Song and Knaap (2004)
Intensity	Number of intersections	Neighborhood	M	M	Song and Popkin (2013)
Intensity	Normalised floor space area	Sanctuary Area	M	M	Dibble et al. (2017)
Intensity	Normalised length of streets	Sanctuary Area	M	M	Feliciotti (2018)
Intensity	Normalised number of blocks	Sanctuary Area	M	M	Feliciotti (2018)
Intensity	Normalised number of intersections	Sanctuary Area	M	M	Feliciotti (2018)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Intensity	Number of internal streets	Sanctuary Area	M	M	Dibble et al. (2017)
Intensity	Cul-de-sacs to streets ratio	Street network	M	M	Lowry and Lowry (2014)
Intensity	Ingress/ Egress Ratio	Street network	M	M	Dibble et al. (2017)
Intensity	Node degree	Street network	M	M	Feliciotti (2018)
Intensity	Normalised length	Street network	M	M	Dibble et al. (2017)
Intensity	Weighted intersection density	Street network	M	M	Dibble et al. (2017)
Intensity	Strong Grid Pattern Ratio	Street network/Block	M	M	Dibble et al. (2017)
Intensity	Weak Grid Pattern Ratio	Street network/Block	M	M	Dibble et al. (2017)
Intensity	Normalised number of plots	Street	S	M/S	Dibble et al. (2017)
Intensity	Continuity	Built-up area	L	L	Gielen (2017)
Intensity	Discontinuous surface	Built-up area	L	L	Gielen (2017)
Intensity	Housing-normalised open space area	Built-up area	L	L	Gielen (2017)
Intensity	Normalised floor space area	Built-up area	L	L	Gielen (2017)
Intensity	Number of segments	Street network	L	L	Khirfan (2011)
Intensity	Normalised length of streets network	800m-side square	S	L/M	Berghauser Pont, M. and Olsson J. (2017)
Intensity	Normalised length of streets network	800m-side square	S	L/M	Hausleitner, B. and Berghauser Pont, M. (2017)
Intensity	Coefficient of land occupancy	Case study	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Normalised area of pedestrian and bicycle paths	Case study	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Normalised area of roads	Case study	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Normalised built-up area	Case study	L/M	L/M	Ye and van Nes (2014)
Intensity	Normalised floor space area	Case study	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Normalised green space area	Case study	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Normalised number of housing units	Case study	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Normalised number of plots	Case study	L/M	L/M	Bourdic, Salat and Nowacki (2012)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Intensity	Number of Segments	Minimum spanning tree	L/M	L/M	Caruso, Hilal and Thomas (2017)
Intensity	Normalised green space area	Neighborhood	L/M	L/M	Lai et al. (2018)
Intensity	Normalised length of streets	Neighborhood	L/M	L/M	Lai et al. (2018)
Intensity	Normalised number of intersections	Pedestrian network	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Normalised number of intersections	Road network	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Proportion dedicated to public transport	Road network	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Intensity	Average node degree presence	Street network	L/M	L/M	Araldi and Fusco (2019)
Intensity	Housing units-normalised length	Street network	L/M	L/M	Song and Knaap (2003)
Intensity	Node degree	Street network	L/M	L/M	Boeing (2018a)
Intensity	Normalised number of intersections	Street network	L/M	L/M	Lai et al. (2018)
Intensity	Proportion of 3-way intersections	Street network	L/M	L/M	Boeing (2018a)
Intensity	Proportion of 4-way intersections	Street network	L/M	L/M	Boeing (2018a)
Intensity	Proportions of dead-ends	Street network	L/M	L/M	Boeing (2018a)
Intensity	Streets per node	Street network	L/M	L/M	Boeing (2018a)
Intensity	Normalised length of streets network	Urban area	L/M	L/M	Peponis et al. (2007)
Intensity	Normalised number of blocks	Urban area	L/M	L/M	Peponis et al. (2007)
Intensity	Normalised number of housing units	Urban area	L/M	L/M	Galster (2001)
Intensity	Normalised number of intersections	Urban area	L/M	L/M	Peponis et al. (2007)
Intensity	Normalised number of street network nodes	Urban area	L/M	L/M	Boeing (2018b)
Intensity	Normalised number of street segments	Urban area	L/M	L/M	Boeing (2018a)
Intensity	Number of blocks	Urban area	L/M	L/M	Peponis et al. (2007)
Intensity	Number of intersections	Urban area	L/M	L/M	Boeing (2018a)
Intensity	Number of street segments	Urban area	L/M	L/M	Peponis et al. (2007)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Connectivity	Cyclomatic complexity	Pedestrian network	M	M	Bourdic, Salat and Nowacki (2012)
Connectivity	Block accessibility	Street network	S	M	Oliveira (2013)
Connectivity	Accessibility of plots	Street network	S	M	Oliveira (2013)
Connectivity	Accessible plot density	Street network	S	M	Feliciotti (2018)
Connectivity	Connectivity	Street network	M	M	Lowry and Lowry (2014)
Connectivity	External connectivity	Street network	M	M	Song and Knaap (2004)
Connectivity	Internal connectivity	Street network	M	M	Song and Knaap (2004)
Connectivity	Link / Node Ratio	Street network	M	M	Dibble et al. (2017)
Connectivity	Local closeness	Street network	S	M	Feliciotti (2018)
Connectivity	Meshedness	Street network	S	M	Feliciotti (2018)
Connectivity	Redundancy	Street network	S	M	Feliciotti (2018)
Connectivity	Straightness	Street network	S	M	Feliciotti (2018)
Connectivity	Connectivity	Axial map	S	L/M	Hallowell (2013)
Connectivity	Global integration	Axial map	S	L/M	Hallowell (2013)
Connectivity	Local integration	Axial map	S	L/M	Hallowell (2013)
Connectivity	Cyclomatic complexity	Car network	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Connectivity	2 mins driving distance Area	Street network	S	L/M	Schirmer and Axhausen (2015)
Connectivity	Accessibility of buildings	Street network	S	L/M	Schirmer and Axhausen (2015)
Connectivity	Accessibility of floorspace	Street network	S	L/M	Schirmer and Axhausen (2015)
Connectivity	Angular betweenness	Street network	S	L/M	Omer and Kaplan (2018)
Connectivity	Angular betweenness centrality	Street network	S	L/M	Berghauser Pont, M. et al. (2017)
Connectivity	Angular closeness	Street network	S	L/M	Omer and Kaplan (2018)
Connectivity	Average mean maximum gravity	Street network	L/M	L/M	Sevtsuk (2016)
Connectivity	Betweenness centrality	Street network	S	L/M	Porta et al. (2006)
Connectivity	Centrality	Street network	S	L/M	Boeing (2018a)
Connectivity	Closeness centrality	Street network	S	L/M	Porta et al. (2006)
Connectivity	Clustering Coefficient	Street network	L/M	L/M	Boeing (2018a)
Connectivity	Combined betweenness-closeness	Street network	S	L/M	Omer and Kaplan (2018)
Connectivity	Connected Node Ratio	Street network	L/M	L/M	Feliciotti (2018)
Connectivity	Connectivity	Street network	S	L/M	Hillier (various)
Connectivity	Connectivity alpha index	Street network	L/M	L/M	Song and Popkin (2013)
Connectivity	Connectivity beta index	Street network	L/M	L/M	Song and Popkin (2013)
Connectivity	Connectivity cyclomatic index	Street network	L/M	L/M	Song and Popkin (2013)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Connectivity	Connectivity gamma index	Street network	L/M	L/M	Song and Popkin (2013)
Connectivity	Control	Street network	S	L/M	Hillier (various)
Connectivity	Directional distance	Street network	S	L/M	Peponis et al. (2007)
Connectivity	Eigenvector centrality	Street network	S	L/M	Agryzcov et al. (2017)
Connectivity	External connectivity	Street network	L/M	L/M	Song and Knaap (2003)
Connectivity	Gravity	Street network	S	L/M	Sevtsuk and Mekonnen (2012)
Connectivity	Gravity accessibility of Mean Maximum Gravity ...	Street network	S	L/M	Sevtsuk (2016)
Connectivity	Gravity achieved	Street network	S	L/M	Sevtsuk (2016)
Connectivity	Choice	Street network	S	L/M	Hausleitner, B. and Berghauser Pont, M. (2017)
Connectivity	Integration	Street network	S	L/M	Hausleitner, B. and Berghauser Pont, M. (2017)
Connectivity	Internal connectivity	Street network	L/M	L/M	Song and Knaap (2003)
Connectivity	Link / Node Ratio	Street network	L/M	L/M	Dill (2004)
Connectivity	Local betweenness	Street network	S	L/M	Gil (2012)
Connectivity	Local closeness	Street network	S	L/M	Gil (2012)
Connectivity	Node connectivity	Street network	L/M	L/M	Boeing (2018b)
Connectivity	Node/edge connectivity	Street network	L/M	L/M	Boeing (2018b)
Connectivity	PageRank	Street network	L/M	L/M	Boeing (2018a)
Connectivity	Reach	Street network	S	L/M	Sevtsuk and Mekonnen (2012)
Connectivity	Self-loop proportion	Street network	L/M	L/M	Boeing (2018a)
Connectivity	Straightness centrality	Street network	S	L/M	Porta et al. (2006)
Connectivity	Total plots required in block frontage to achi...	Street network	L/M	L/M	Sevtsuk (2016)
Connectivity	Weighted Clustering Coefficient	Street network	L/M	L/M	Boeing (2018a)
Diversity	Block Area Power law distribution	Sanctuary Area	M	M	Feliciotti (2018)
Diversity	Interface type Power law distribution	Sanctuary Area	M	M	Feliciotti (2018)
Diversity	Plot Area Heterogeneity (Gini-Simpson)	Sanctuary Area	M	M	Feliciotti (2018)
Diversity	Plot Area Power law distribution	Sanctuary Area	M	M	Feliciotti (2018)

Continued on next page

Table A4.1: Table of Urban Form Characters. A full overview of measurable urban form characters, showing: category; name/definition of characters according to the Index of Elements approach; urban form character's position according to category and scale.

category	Index	Element	grain	extent	reference
Diversity	Street length power law distribution	Sanctuary Area	M	M	Feliciotti (2018)
Diversity	Intersection proportion	Street network	M	M	Song and Popkin (2013)
Diversity	Dissimilarity index	Tract	S	M	Krizek (2003)
Diversity	Plot area diversity (Simpson's)	Accessible radius	S	L/M	Bobkova, Marcus and Berghauser Pont (2017)
Diversity	Plot Area Diversity	Case study	L/M	L/M	Bourdic, Salat and Nowacki (2012)
Diversity	Block Area Power law distribution	Neighborhood	L/M	L/M	Louf and Barthelemy (2014)
Diversity	Block Shape factor Probaility conditional dist...	Neighborhood	L/M	L/M	Louf and Barthelemy (2014)
Diversity	Diversity index	Street network	L/M	L/M	Agryzkov, Tortosa and Vicent (2018)
Diversity	Scale hierarchy	Street network	L/M	L/M	Bourdic, Salat and Nowacki (2012)

Appendix A6: Supplementary material for chapter 6

A6.1 Spearman's rho rank correlation

Additional figures to figure 6.23 capturing the details of Spearman's rank correlation on subsets of data (see section 6.4.1.2.1).

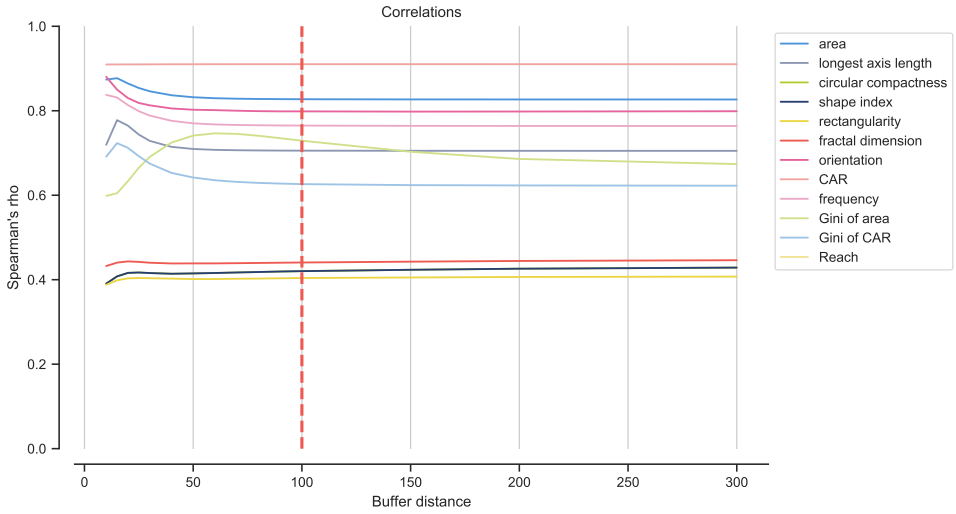


Figure A6.1: Spearman's rho rank correlation between cadastral values and each of the selected buffers of tessellation based on the single-building plots.

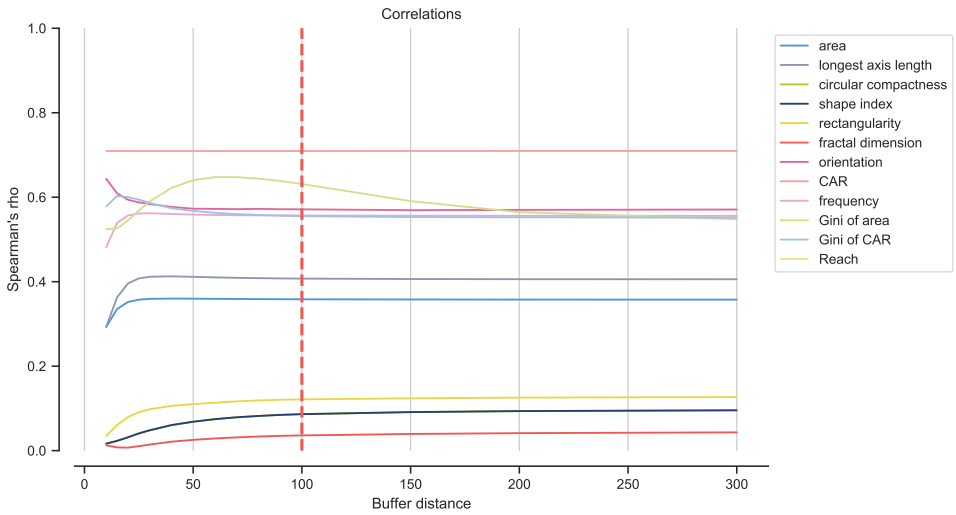


Figure A6.2: Spearman's rho rank correlation between cadastral values and each of the selected buffers of tessellation based on the multi-building plots.

A6.2 NRMSD

Additional figures to figure 6.24 capturing the details of NRMSD on subsets of data (see section 6.4.1.2.2).

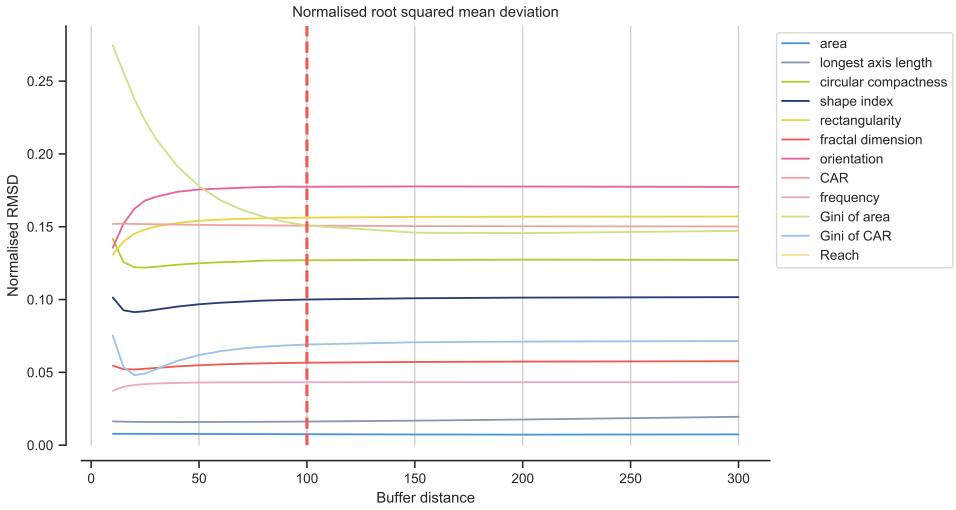


Figure A6.3: NRMSD of cadastral values and each of the selected buffers of tessellation based on the single-building plots.

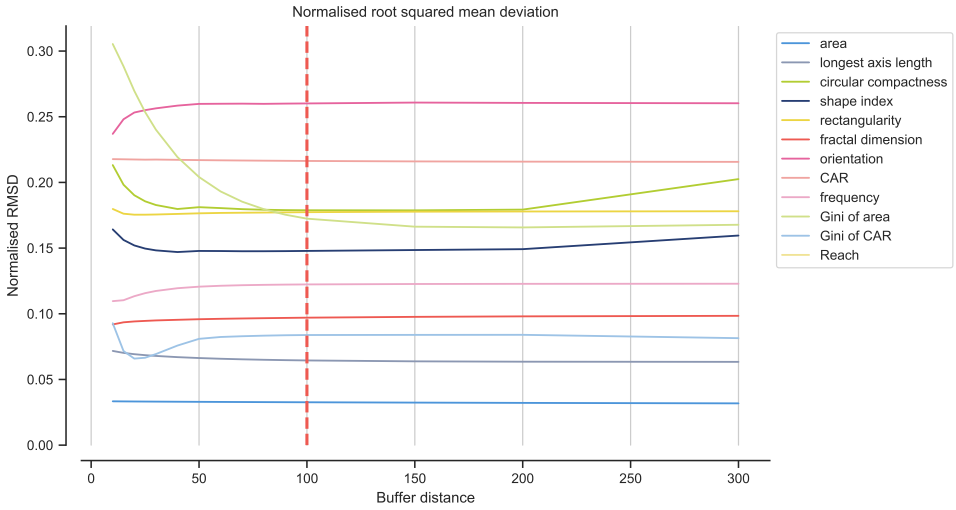


Figure A6.4: NRMSD of cadastral values and each of the selected buffers of tessellation based on the multi-building plots.

A6.3 LISA accuracy

Additional figures to figure 6.25 capturing the details of LISA accuracy on subsets of data (see section 6.4.1.2.3).

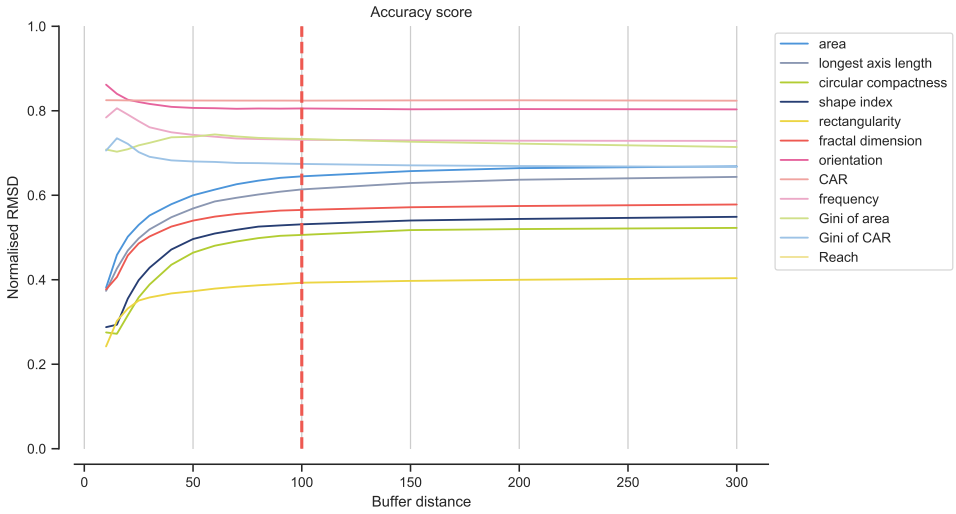


Figure A6.5: LISA accuracy of cadastral values and each of the selected buffers of tessellation based on the single-building plots.

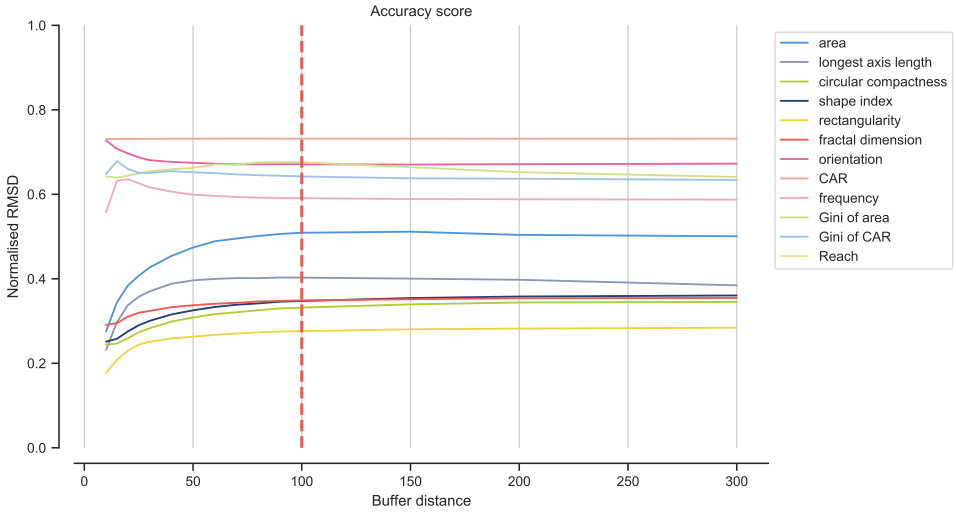


Figure A6.6: LISA accuracy of cadastral values and each of the selected buffers of tessellation based on the multi-building plots.

Appendix A7: Supplementary material for chapter 7

7.1 Selection of primary characters

Supplementary material for section 7.1.2.1.1. Table A7.1 contains initial selection of applicable characters and reasoning behind its selection or exclusion.

Table A7.1: Initial selection of applicable characters as an extraction from the Table of Urban Form Characters (Table A4.1.) Selection reflects steps defined in section 7.1.2.1.1. Result indicates whether tested character is included in the final set of primary characters or not, alternatively specifies the reason. Characters using different data than specified in the relational model are excluded a priori.

id	index	element	extent	category	result
sdbAre	area	building	S	dimension	included
sdbFlA	floor area	building	S	dimension	collinear
sdbHei	height	building	S	dimension	included
sdbVol	volume	building	S	dimension	included
sdbPer	perimeter	building	S	dimension	included
sdbCoA	courtyard area	building	S	dimension	included
sdbBRA	bounding rectangle area	building	S	dimension	collinear
sdbBRW	bounding rectangle width	building	S	dimension	collinear
sbdBRL	bounding rectangle length	building	S	dimension	collinear

Continued on next page

Table A7.1: Initial selection of applicable characters as an extraction from the Table of Urban Form Characters (Table A4.1.) Selection reflects steps defined in section 7.1.2.1.1. Result indicates whether tested character is included in the final set of primary characters or not, alternatively specifies the reason. Characters using different data than specified in the relational model are excluded a priori.

id	index	element	extent	category	result
sdbBRP	bounding rectangle perimeter	building	S	dimension	collinear
sdbECR	enclosing circle radius	building	S	dimension	collinear
sdbCHA	convex hull area	building	S	dimension	collinear
sdbCHP	convex hull perimeter	building	S	dimension	collinear
ssbFoF	form factor	building	S	shape	included
ssbFra	fractal dimension	building	S	shape	collinear
ssbVFR	volume/façade ratio	building	S	shape	included
ssbCCo	circular compactness	building	S	shape	included
ssbSCo	square compactness	building	S	shape	collinear
ssbCon	convexeity	building	S	shape	collinear
ssbCor	corners	building	S	shape	included
ssbShI	shape index	building	S	shape	collinear
ssbSqu	squareness	building	S	shape	included
ssbERI	equivalent rectangular index	building	S	shape	included
ssbElo	elongation	building	S	shape	included
ssbCCD	centroid - corners distance deviation	building	S	shape	included
ssbCCM	centroid - corners mean distance	building	S	shape	included
stbOri	solar orientation	building	S	distribution	included
stbSAl	street alignment	building	S	distribution	included
stbCeA	cell alignment	building	S	distribution	included
sdcLAL	longest axis length	tessellation cell	S	dimension	included
sdcAre	area	tessellation cell	S	dimension	included

Continued on next page

Table A7.1: Initial selection of applicable characters as an extraction from the Table of Urban Form Characters (Table A4.1.) Selection reflects steps defined in section 7.1.2.1.1. Result indicates whether tested character is included in the final set of primary characters or not, alternatively specifies the reason. Characters using different data than specified in the relational model are excluded a priori.

id	index	element	extent	category	result
sdcBRA	bounding rectangle area	tessellation cell	S	dimension	collinear
sdcBRW	bounding rectangle width	tessellation cell	S	dimension	collinear
sdcBRL	bounding rectangle length	tessellation cell	S	dimension	collinear
sdcBRP	bounding rectangle perimeter	tessellation cell	S	dimension	collinear
sdcECR	enclosing circle radius	tessellation cell	S	dimension	collinear
sdcCHA	convex hull area	tessellation cell	S	dimension	collinear
sdcCHP	convex hull perimeter	tessellation cell	S	dimension	collinear
sscCCo	circular compactness	tessellation cell	S	shape	included
sscSCo	square compactness	tessellation cell	S	shape	collinear
sscElo	elongation	tessellation cell	S	shape	collinear
sscFra	fractal dimension	tessellation cell	S	shape	collinear
sscCon	convexeity	tessellation cell	S	shape	collinear
sscShI	shape index	tessellation cell	S	shape	collinear
sscERI	equivalent rectangular index	tessellation cell	S	shape	included
sscElo	elongation	tessellation cell	S	shape	collinear
stcOri	solar orientation	tessellation cell	S	distribution	included
stcSAl	street alignment	tessellation cell	S	distribution	included
sicCAR	coverage area ratio	tessellation cell	S	intensity	included
sicFAR	floor area ratio	tessellation cell	S	intensity	included
sdsLen	length	street segment	S	dimension	included
sdsSPW	street profile width	street	S	dimension	included
sdsSPH	street profile height	street	S	dimension	included

Continued on next page

Table A7.1: Initial selection of applicable characters as an extraction from the Table of Urban Form Characters (Table A4.1.) Selection reflects steps defined in section 7.1.2.1.1. Result indicates whether tested character is included in the final set of primary characters or not, alternatively specifies the reason. Characters using different data than specified in the relational model are excluded a priori.

id	index	element	extent	category	result
sdsSPR	street profile height/width ratio	street segment	S	shape	included
sdsSPO	street openness	street	S	distribution	included
sdsSWD	setback deviation	street	S	diversity	included
sdsSHD	height deviation	street	S	diversity	included
sssLin	linearity	street segment	S	shape	included
sdsAre	area covered	tessellation cell	S	dimension	included
sisBpS	buildings per segment	building	S	intensity	no meaning
sisBpM	buildings per meter	building	S	intensity	included
sddAre	area covered	tessellation cell	S	dimension	included
sddBpN	buildings per node	building	S	intensity	no meaning
mtbSWR	shared walls ratio	building	M	distribution	included
mtbAli	building alignment	building	M	distribution	included
mtbNDi	mean neighbour distance	building	M	distribution	included
mtcNei	neigbours	tessellation cell	M	distribution	no meaning
mtcWNe	neigbours per m	tessellation cell	M	distribution	included
mdcAre	area covered	tessellation cell	M	dimension	included
mtsMDO	mean deviation of orientation	street segment	M	distribution	no meaning
misRea	reached cells	tessellation cell	M	intensity	included
mdsLen	length	street segment	M	dimension	collinear
mdsAre	area covered	tessellation cell	M	dimension	included
mtdDeg	node degree	street node	M	distribution	included
mtdMDi	mean distance to nodes	street node	M	dimension	included
midRea	reached cells	tessellation cell	M	intensity	included

Continued on next page

Table A7.1: Initial selection of applicable characters as an extraction from the Table of Urban Form Characters (Table A4.1.) Selection reflects steps defined in section 7.1.2.1.1. Result indicates whether tested character is included in the final set of primary characters or not, alternatively specifies the reason. Characters using different data than specified in the relational model are excluded a priori.

id	index	element	extent	category	result
midAre	area covered	tessellation cell	M	dimension	included
libNCo	number of courtyards	building	L	intensity	included
ldbPWL	perimeter wall length	building	L	dimension	included
ltbIBD	mean inter-building distance	building	L	distribution	included
ltcBuA	building adjacency	building	L	distribution	included
licGDe	gross density	tessellation cell	L	intensity	included
ltcWRB	weighted reached blocks	tessellation cell	L	intensity	included
ldkAre	block area	block	L	dimension	included
ldkPer	block perimeter	block	L	dimension	included
ldkBRA	bounding rectangle area	block	L	dimension	collinear
ldkBRW	bounding rectangle width	block	L	dimension	collinear
ldkBRL	bounding rectangle length	block	L	dimension	collinear
ldkBRP	bounding rectangle perimeter	block	L	dimension	collinear
ldkECR	enclosing circle radius	block	L	dimension	collinear
ldkCHA	convex hull area	block	L	dimension	collinear
ldkCHP	convex hull perimeter	block	L	dimension	collinear
lskElo	block elongation	block	L	shape	collinear
lskFra	block fractal dimension	block	L	shape	collinear

Continued on next page

Table A7.1: Initial selection of applicable characters as an extraction from the Table of Urban Form Characters (Table A4.1.) Selection reflects steps defined in section 7.1.2.1.1. Result indicates whether tested character is included in the final set of primary characters or not, alternatively specifies the reason. Characters using different data than specified in the relational model are excluded a priori.

id	index	element	extent	category	result
lskCCo	block circular compactness	block	L	shape	included
lskSCo	blocks square compactness	block	L	shape	collinear
lskCon	block convexeity	block	L	shape	collinear
lskShI	block shape index	block	L	shape	collinear
lskERI	block equivalent rectangular index	block	L	shape	included
lskCWA	Compactness-weighted axis	block	L	shape	included
ltkOri	block solar orientation	block	L	distribution	included
ltkNei	block neighbours	block	L	distribution	no meaning
ltkWBN	weighted block neighbours	block	L	distribution	included
likBpB	buildings per block	block	L	intensity	no meaning
likWBB	weighted buildings per block	block	L	intensity	included
lcdMes	meshedness	street network	L	connectivity	included
ldsMSL	mean segment length	street network	L	dimension	included
ldsTSL	total segment length	street network	L	dimension	collinear
ldsCDL	cul-de-sac length	street network	L	dimension	included
ldsRea	reached cells	tessellation cell	L	dimension	included
lddNDe	node density	street network	L	intensity	included
lddRea	reached cells	tessellation cell	L	dimension	included
lddARe	area covered	tessellation cell	L	dimension	included
lcnMND	mean node degree	street network	L	connectivity	collinear

Continued on next page

Table A7.1: Initial selection of applicable characters as an extraction from the Table of Urban Form Characters (Table A4.1.) Selection reflects steps defined in section 7.1.2.1.1. Result indicates whether tested character is included in the final set of primary characters or not, alternatively specifies the reason. Characters using different data than specified in the relational model are excluded a priori.

id	index	element	extent	category	result
linPDE	proportion of dead-ends	street network	L	connectivity	included
linP3W	proportion of 3-way intersections	street network	L	connectivity	included
linP4W	proportion of 4-way intersections	street network	L	connectivity	included
linWID	weighted intersection density	street network	L	intensity	included
licSpA	spatial attraction	tessellation cell	L	intensity	scale
lcnSLP	self-loop proportion	street network	L	connectivity	collinear
lcnNeC	network clustering	street network	L	connectivity	collinear
lncWNC	weighted network clustering	street network	L	connectivity	collinear
lcnEdC	edge connectivity	street network	L	connectivity	collinear
lcnNoC	node connectivity	street network	L	connectivity	collinear
lcnCyC	cyclomatic complexity	street network	L	connectivity	collinear
lcnENR	edge / node ratio	street network	L	connectivity	collinear
lcnExC	external connectivity	street network	L	connectivity	no data
lcnGaI	gamma index	street network	L	connectivity	collinear
lcnDiD	directional distance	street network	L	connectivity	ineffective
lcnGCC	global clustering coefficient	street network	L	connectivity	meaningless
lcnRed	redundancy	street network	L	connectivity	ineffective
lcnClo	local closeness centrality	street network	L	connectivity	included
xcnSCl	square clustering	street network	XL	connectivity	included

The following figures measure Spearman correlation between tested characters within each group of elements. It illustrates the reasoning behind exclusion of characters due to collinearity. From a group of collinear characters, only one is included in the final set. The selection is driven by variance, interpretability and literature.

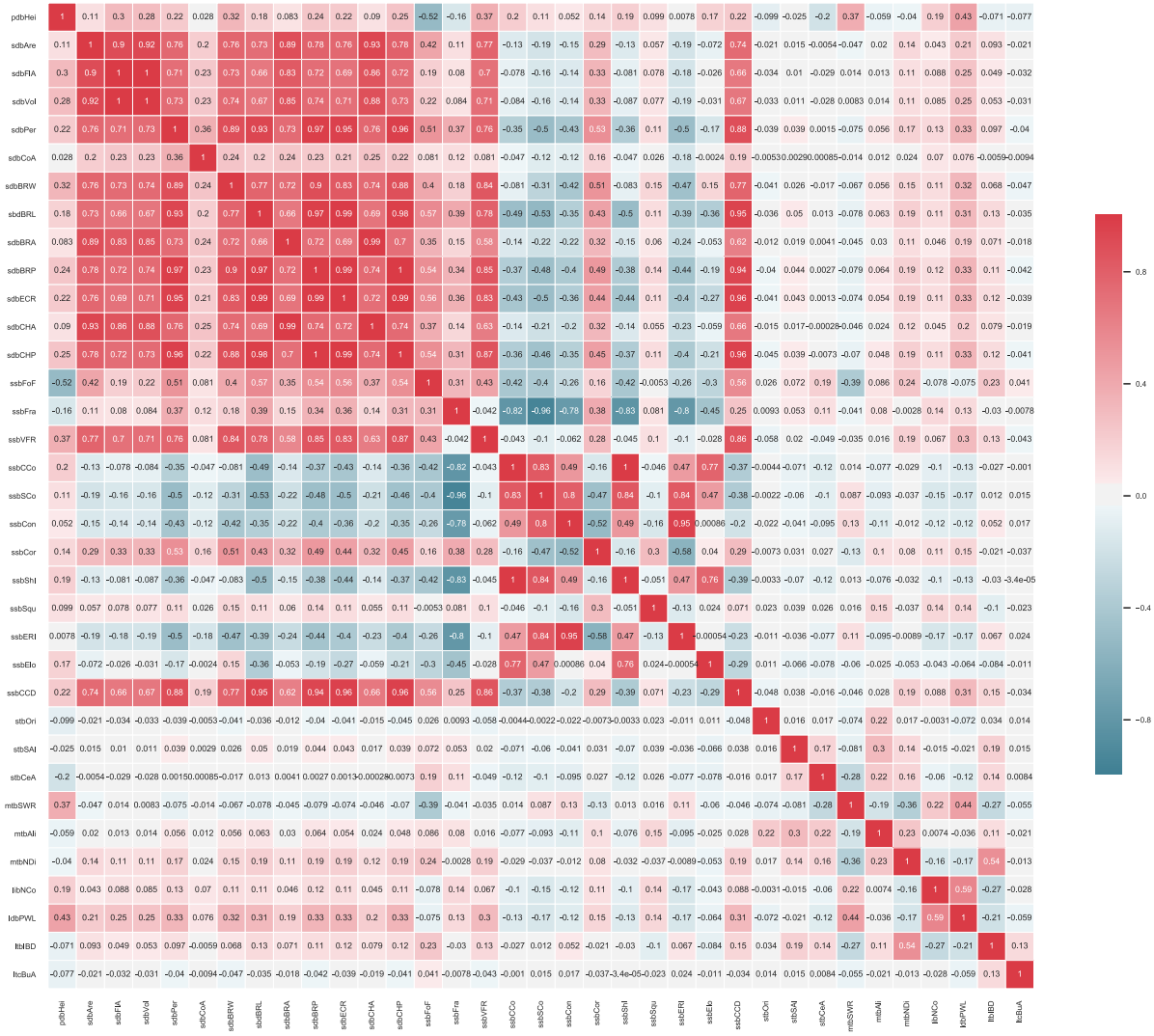


Figure A7.1: Correlation matrix of Spearman's rho values capturing the statistical relationship between morphometric values of tested characters based on buildings.

Figure A7.2: Correlation matrix of Spearman's rho values capturing the statistical relationship between morphometric values of tested characters based on tessellation cells.

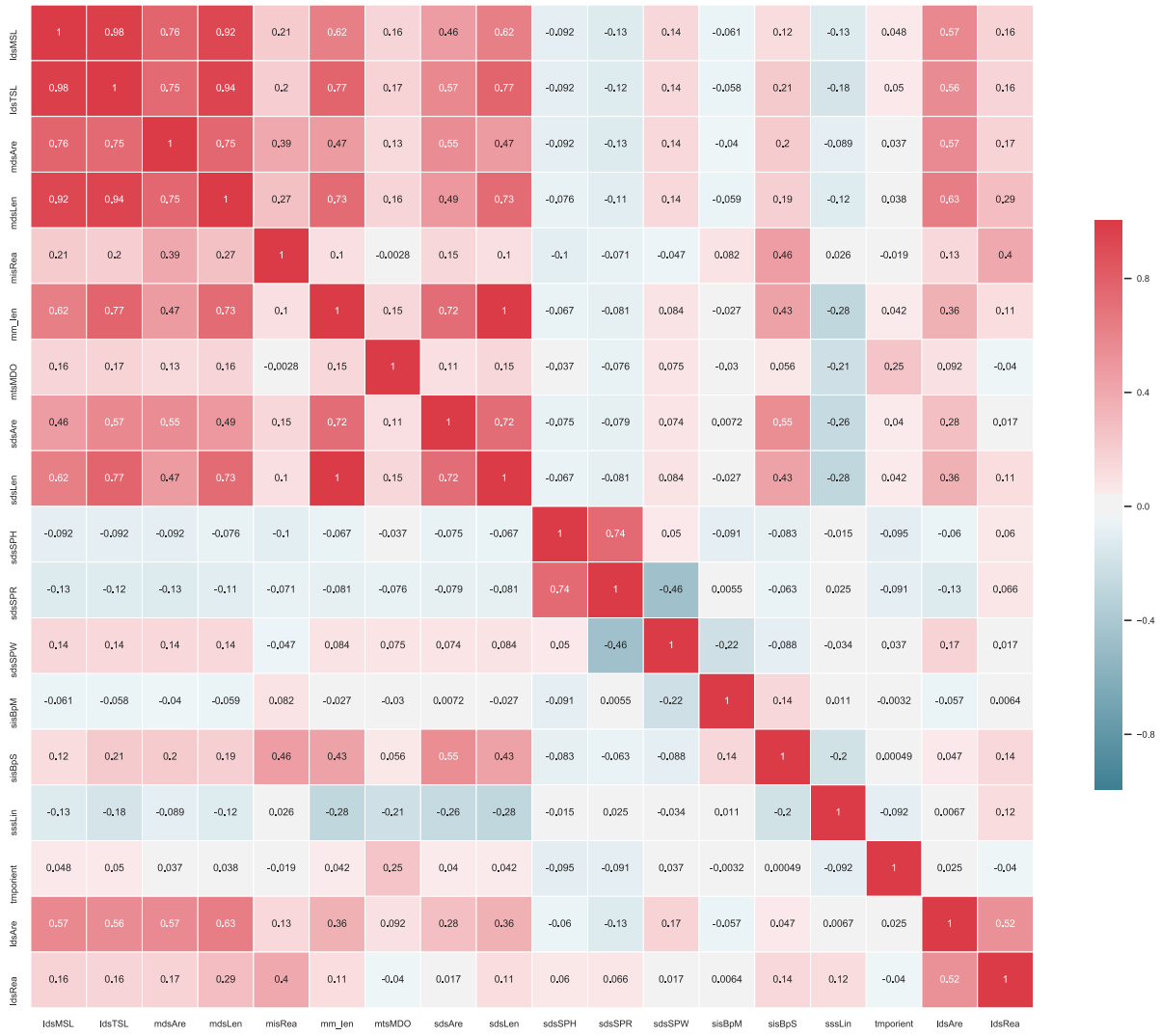


Figure A7.3: Correlation matrix of Spearman's rho values capturing the statistical relationship between morphometric values of tested characters based on streets.

Figure A7.4: Correlation matrix of Spearman's rho values capturing the statistical relationship between morphometric values of tested characters based on blocks.

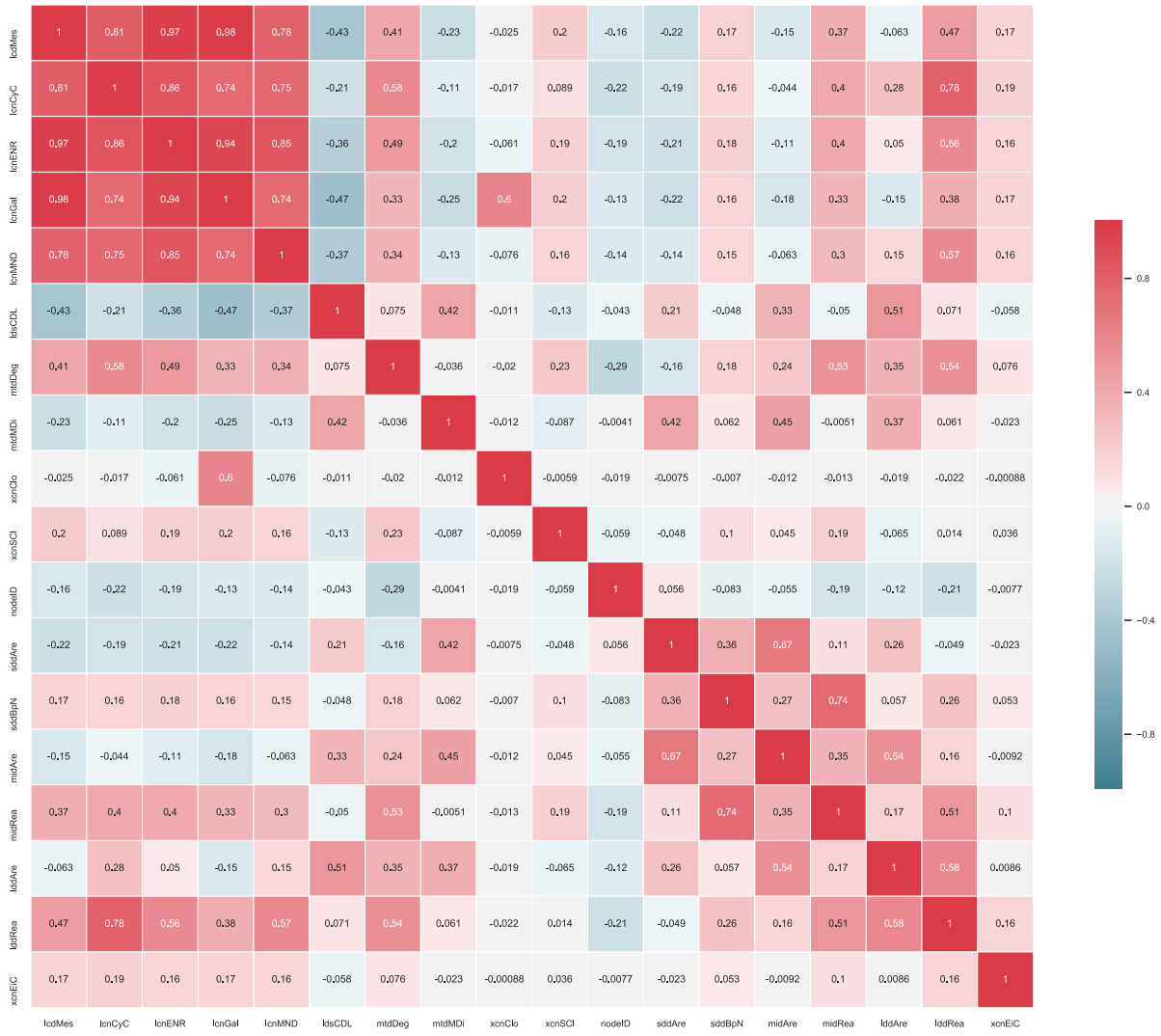


Figure A7.5: Correlation matrix of Spearman's rho values capturing the statistical relationship between morphometric values of tested characters based on network nodes.

7.2 Classification of primary characters

Classification of primary characters into categories. See section 7.1.2.1.2 for details.

Table A7.2: Classification of primary characters. Grain is S for all characters.

notation	index	element	category	extent
a_{blg}	area	building	dimension	S
h_{blg}	height	building	dimension	S
v_{blg}	volume	building	dimension	S
p_{blg}	perimeter	building	dimension	S
a_{blgc}	courtyard area	building	dimension	S
FoF_{blg}	form factor	building	shape	S
VFR_{blg}	volume to façade ratio	building	shape	S
CCo_{blg}	circular compactness	building	shape	S
Cor_{blg}	corners	building	shape	S
Squ_{blg}	squareness	building	shape	S
ERI_{blg}	equivalent rectangular index	building	shape	S
Elo_{blg}	elongation	building	shape	S
CCD_{blg}	centroid - corner distance deviation	building	shape	S
CCM_{blg}	centroid - corner mean distance	building	shape	S
Ori_{blg}	solar orientation	building	distribution	S
SAl_{blg}	street alignment	building	distribution	S
CAl_{blg}	cell alignment	building	distribution	S
LAL_{cell}	longest axis length	tessellation cell	dimension	S
a_{cell}	area	tessellation cell	dimension	S
CCo_{cell}	circular compactness	tessellation cell	shape	S
ERI_{cell}	equivalent rectangular index	tessellation cell	shape	S
Ori_{cell}	solar orientation	tessellation cell	distribution	S
SAl_{cell}	street alignment	tessellation cell	distribution	S

notation	index	element	category	extent
CAR_{cell}	coverage area ratio	tessellation cell	intensity	S
FAR_{cell}	floor area ratio	tessellation cell	intensity	S
l_{edg}	length	street segment	dimension	S
w_{sp}	width	street profile	dimension	S
h_{sp}	height	street profile	dimension	S
HWR_{sp}	height to width ratio	street profile	shape	S
Ope_{sp}	openness	street profile	distribution	S
$wDev_{sp}	width deviation	street profile	diversity	S
$hDev_{sp}$	height deviation	street profile	diversity	S
Lin_{edg}	linearity	street segment	shape	S
a_{edg}	area covered	street segment	dimension	S
BpM_{edg}	buildings per meter	street segment	intensity	S
a_{node}	area covered	street node	dimension	S
SWR_{blg}	shared walls ratio	adjacent buildings	distribution	S
Ali_{blg}	alignment	neighbouring buildings	distribution	S
NDi_{blg}	mean distance	neighbouring buildings	distribution	S
WNe_{cell}	weighted neighbours	tessellation cell	distribution	S
a_{cell_n}	area covered	neighbouring cells	dimension	S
RC_{edg_n}	reached cells	neighbouring segments	intensity	S
a_{edg_n}	reached area	neighbouring segments	dimension	S
deg_{node}	degree	street node	distribution	S
MDi_{node}	mean distance to neighbouring nodes	street node	dimension	S
RC_{node_n}	reached cells	neighbouring nodes	intensity	S
a_{node_n}	reached area	neighbouring nodes	dimension	S
$NCo_{blg_{adj}}$	number of courtyards	adjacent buildings	intensity	S
$p_{blg_{adj}}$	perimeter wall length	adjacent buildings	dimension	S

notation	index	element	category	extent
IBD_{blg}	mean inter-building distance	neighbouring buildings	distribution	S
BuA_{blg}	building adjacency	neighbouring buildings	distribution	S
$GFAR_{cell}$	gross floor area ratio	neighbouring tessellation cells	intensity	S
WRB_{cell}	weighted reached blocks	neighbouring tessellation cells	intensity	S
a_{blk}	area	block	dimension	S
p_{blk}	perimeter	block	dimension	S
CCo_{blk}	circular compactness	block	shape	S
ERI_{blk}	equivalent rectangular index	block	shape	S
CWA_{blk}	compactness-weighted axis	block	shape	S
Ori_{blk}	solar orientation	block	distribution	S
wN_{blk}	weighted neighbours	block	distribution	S
wC_{blk}	weighted cells	block	intensity	S
Mes_{node}	local meshedness	street network	connectivity	M
MSL_{edg}	mean segment length	street network	dimension	S
CDL_{node}	cul-de-sac length	street network	dimension	S
RC_{edg}	reached cells	street network	dimension	S
D_{node}	node density	street network	intensity	M
$RC_{node_{net}}$	reached cells	street network	dimension	S
$a_{node_{net}}$	reached area	street network	dimension	S
pCD_{node}	proportion of cul-de-sacs	street network	connectivity	M
$p3W_{node}$	proportion of 3-way intersections	street network	connectivity	M
$p4W_{node}$	proportion of 4-way intersections	street network	connectivity	M
wD_{node}	weighted node density	street network	intensity	M
lCC_{node}	local closeness centrality	street network	connectivity	M

notation	index	element	category	extent
sCl_{node}	square clustering	street network	connectivity	L

References for primary characters. Characters without a reference are newly introduced or adapted to the point where it would not be correct to refer to its original implementation. See section 7.1.2.1.2 for details.

Table A7.3: Reference table for primary characters. Contains references to existing literature and to identifier of each character used within computational Jupyter notebooks.

notation	reference	id
a_{blg}	(Hallowell and Baran, 2013)	sdbAre
h_{blg}	(Schirmer and Axhausen, 2015)	sdbHei
v_{blg}	(Yoshida and Omae, 2005)	sdbVol
p_{blg}	(Vanderhaegen and Canters, 2017)	sdbPer
a_{blg_c}	(Schirmer and Axhausen, 2015)	sdbCoA
FoF_{blg}	(Bourdic *et al.*, 2012)	ssbFoF
VFR_{blg}	(Yoshida and Omae, 2005)	ssbVFR
CCo_{blg}	(Dibble *et al.*, 2017)	ssbCCo
Cor_{blg}	(Steiniger *et al.*, 2008)	ssbCor
Squ_{blg}	(Steiniger *et al.*, 2008)	ssbSqu
ERI_{blg}	(Basaraner and Cetinkaya, 2017)	ssbERI
Elo_{blg}	(Steiniger *et al.*, 2008)	ssbElo
CCD_{blg}		ssbCCD
CCM_{blg}	(Schirmer and Axhausen, 2015)	ssbCCM
Ori_{blg}	(Schirmer and Axhausen, 2015)	stbOri
SAl_{blg}	(Schirmer and Axhausen, 2015)	stbSAl
CAl_{blg}		stbCeA
LAL_{cell}		sdcLAL
a_{cell}	(Hamaina *et al.*, 2012)	sdcAre
CCo_{cell}		sscCCo
ERI_{cell}		sscERI
Ori_{cell}		stcOri
SAl_{cell}		stcSAl
CAR_{cell}	(Hamaina *et al.*, 2013)	sicCAR
FAR_{cell}	(Hamaina *et al.*, 2013)	sicFAR
l_{edg}	(Gil *et al.*, 2012)	sdsLen
w_{sp}	(Araldi and Fusco, 2019)	sdsSPW

notation	reference	id
h_{sp}	(Araldi and Fusco, 2019)	sdsSPH
HWR_{sp}	(Araldi and Fusco, 2019)	sdsSPR
Ope_{sp}	(Araldi and Fusco, 2019)	sdsSPO
$wDev_{sp} $	(Araldi and Fusco, 2019)	sdsSWD
$hDev_{sp}$	(Araldi and Fusco, 2019)	sdsSHD
Lin_{edg}	(Araldi and Fusco, 2019)	sssLin
a_{edg}		sdsAre
BpM_{edg}		sisBpM
a_{node}		sddAre
SWR_{blg}	(Hamaina *et al.*, 2012)	mtbSWR
Ali_{blg}	(Hijazi *et al.*, 2016)	mtbAli
NDi_{blg}	(Hijazi *et al.*, 2016)	mtbNDi
WNe_{cell}		mtcWNe
a_{cell_n}		mdcAre
RC_{edg_n}		misRea
a_{edg_n}		mdsAre
deg_{node}	(Boeing, 2018)	mtdDeg
MDi_{node}		mtdMDi
RC_{node_n}		midRea
a_{node_n}		midAre
$NCo_{blg_{adj}}$	(Schirmer and Axhausen, 2015)	libNCo
$p_{blg_{adj}}$		ldbPWL
IBD_{blg}	(Caruso *et al.*, 2017)	ltbIBD
BuA_{blg}	(Vanderhaegen and Canters, 2017)	ltcBuA
$GFAR_{cell}$	(Dibble *et al.*, 2017)	licGDe
WRB_{cell}		ltcWRB
a_{blk}	(Dibble *et al.*, 2017)	ldkAre
p_{blk}	(Gil *et al.*, 2012)	ldkPer
CCo_{blk}	(Schirmer and Axhausen, 2015)	lskCCo
ERI_{blk}	(Basaraner and Cetinkaya, 2017)	lskERI
CWA_{blk}	(Feliciotti, 2018)	lskCWA
Ori_{blk}	(Gil *et al.*, 2012)	ltkOri

notation	reference	id
wN_{blk}		ltkWNB
wC_{blk}		likWBB
Mes_{node}	(Feliciotti, 2018)	lcdMes
MSL_{edg}		ldsMSL
CDL_{node}		ldsCDL
RC_{edg}		ldsRea
D_{node}		lddNDe
$RC_{node_{net}}$		lddRea
$a_{node_{net}}$		lddARe
pCD_{node}	(Lowry and Lowry, 2014)	linPDE
$p3W_{node}$	(Boeing, 2018)	linP3W
$p4W_{node}$	(Boeing, 2018)	linP4W
wD_{node}	(Dibble *et al.*, 2017)	linWID
lCC_{node}	(Porta *et al.*, 2006)	lcnClo
sCl_{node}		xcnSCl

7.3 Sectional diagram analysis

Sectional diagrams show the distribution of measured values along the longitudinal section through the whole case study. Diagrams were generated every 1km and individually assessed to understand the effect of a different number of steps on the distribution. The overall aim is to use that number of steps which illustrate the tendency within the local area but in too smooth to disable identification of boundaries between areas of different characters. Based on the visual assessment of sectional diagrams, three topological steps are the closest option to the goal mentioned above. Figures A7.6 and A7.7 below illustrate one such diagram and its detail.

Note: Please refer to the PDF version of the thesis for a better clarity. Sectional diagrams are not optimised for print.

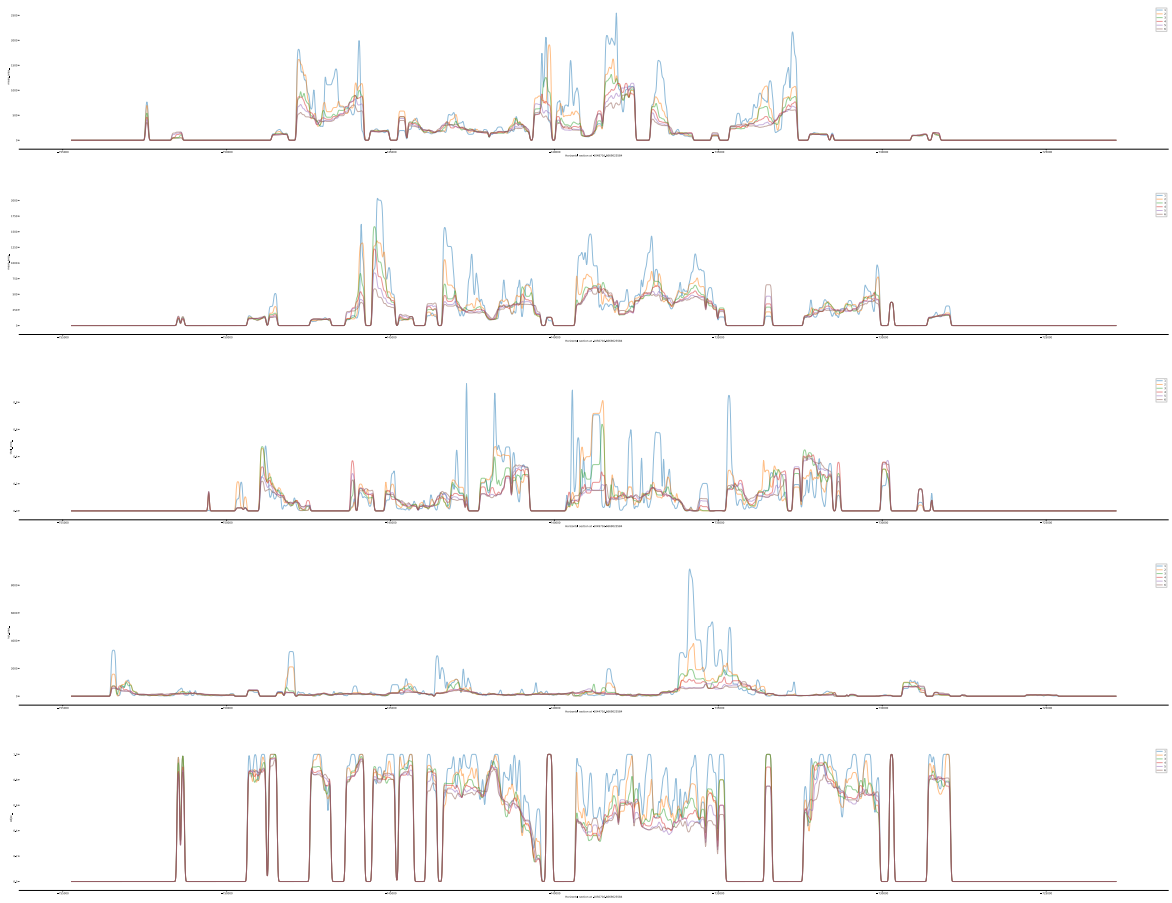
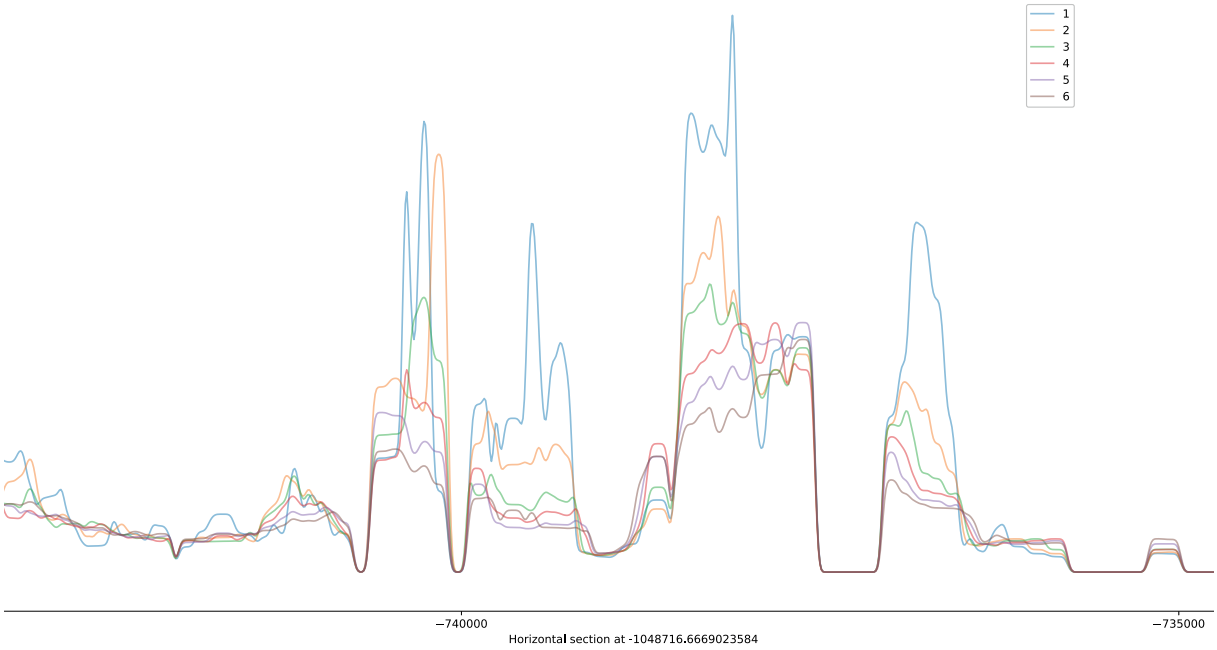


Figure A7.6: Full sectional diagrams illustrating spatial distribution of selected values (IQM of area, ID Theil index of area, IQR of area, adjacency).



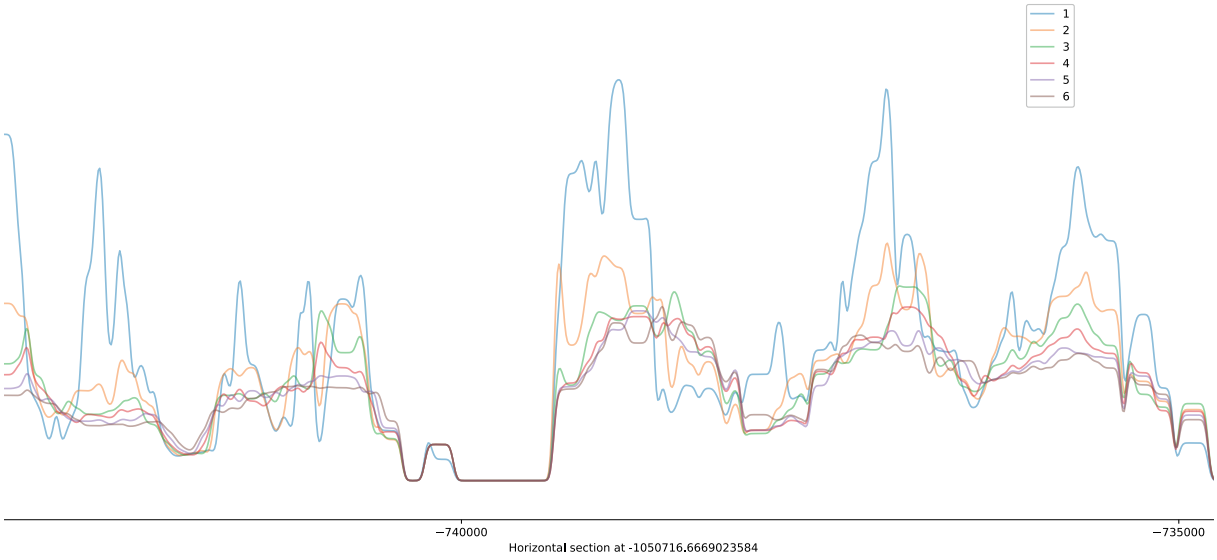


Figure A7.7: Details of two sectional diagrams illustrating spatial distribution of selected values (IQM of area, ID Theil index of area).

7.4 Analysis of local central tendency characters

This appendix presents results of tests of different of measuring central tendency to better understand the differences between them. The aim is to select one (or more) to be used within the cluster analysis.

Tested characters:

- building area
- tessellation area
- height
- building circular compactness
- building solar orientation
- cell circular compactness
- CAR
- shared walls ratio

Tested options:

- mean
- interdecile mean
- interquartile mean
- median

Tested on 3 topological steps.

The aim is to identify such a method which is not prone to outlier effect but at the same time reflect the best true nature of each context.

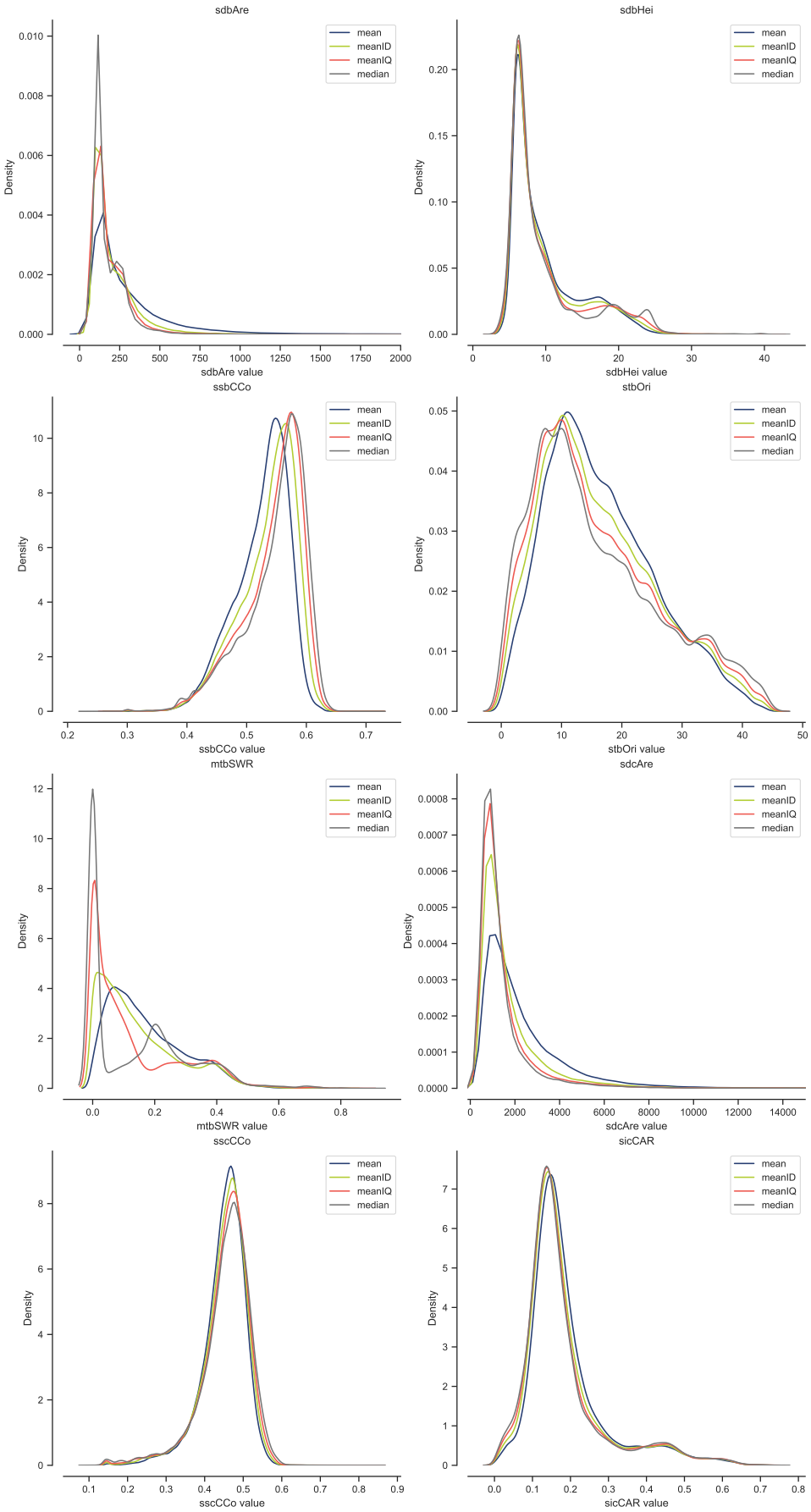


Figure A7.8: Distributions of values measured using each tested option to capture local central tendency.

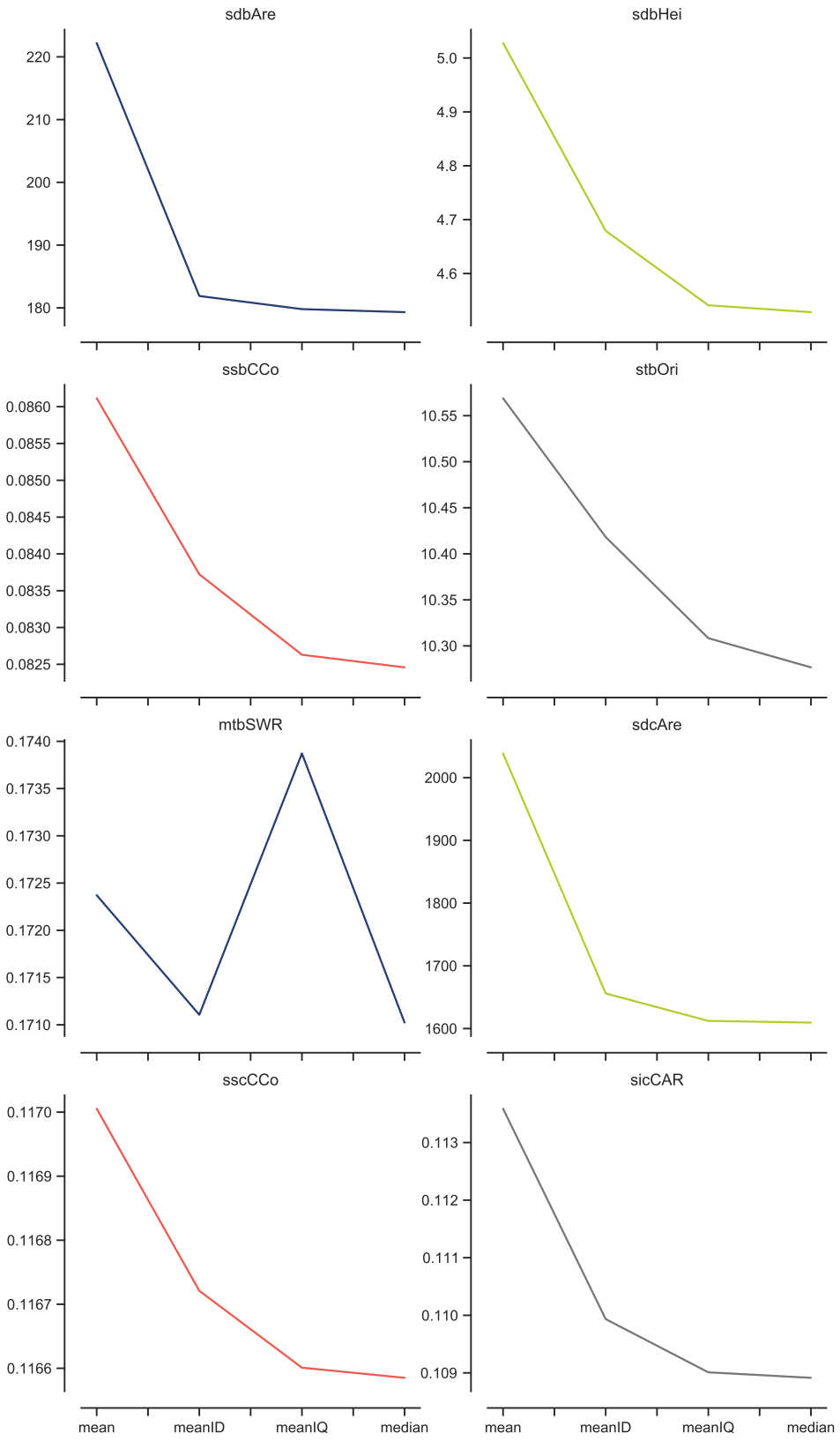


Figure A7.9: Mean deviations of measured primary characters compared to the central tendency value. The lower the deviation is, the better reflection of central tendecy the method offers. It is clear that mean deviates the most, while IQM and median are very similar.

7.5 Comparison of characters capturing properties of distributions

Supplementary data for section 7.1.2.2.2.4. Figures below show spatial distribution (figures A7.10 - A7.17)and correlation (figures A7.18, A7.21) of selected characters capturing properties of distributions.

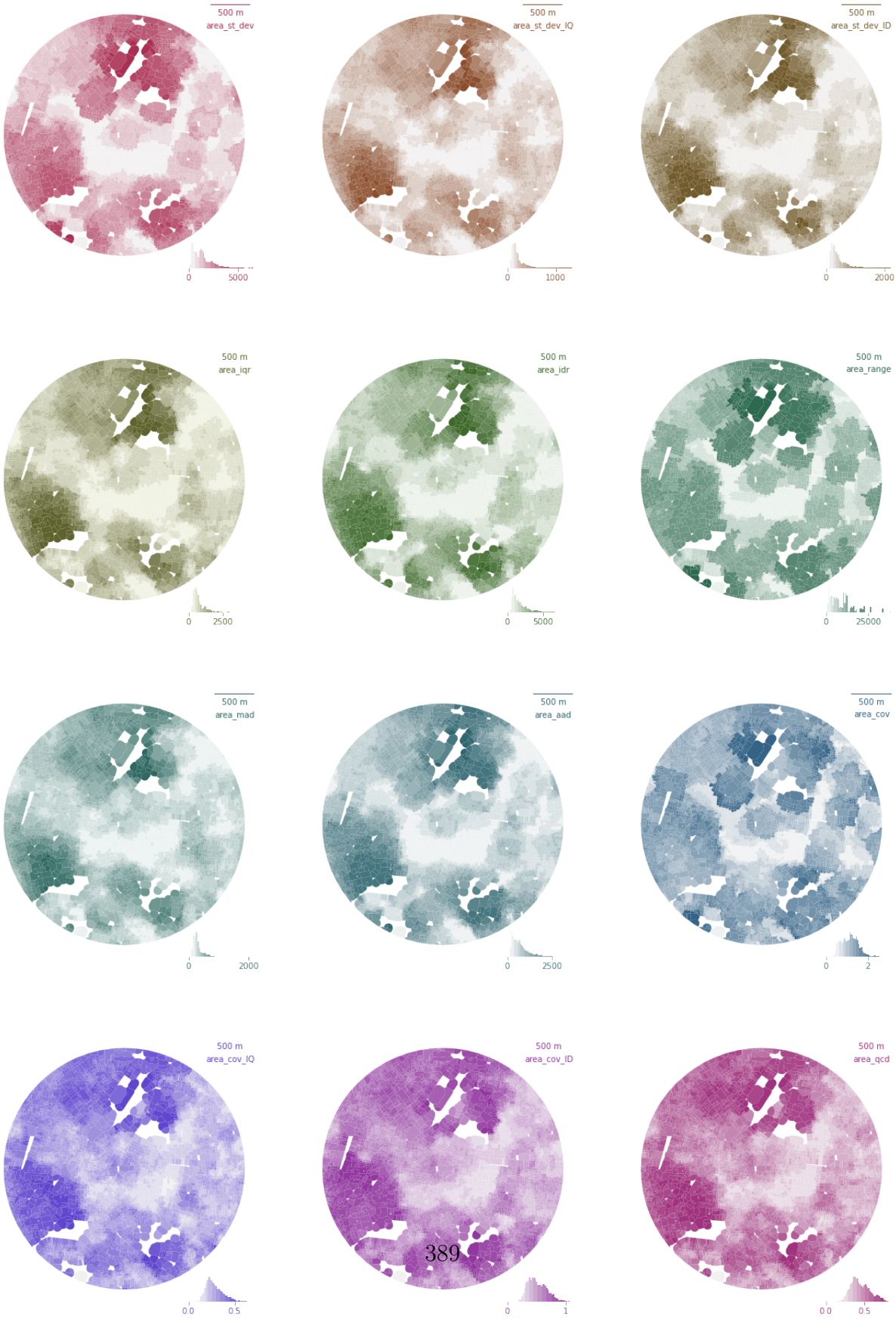


Figure A7.10: Spatial distribution of characters capturing properties of distributions tested on area of a building.

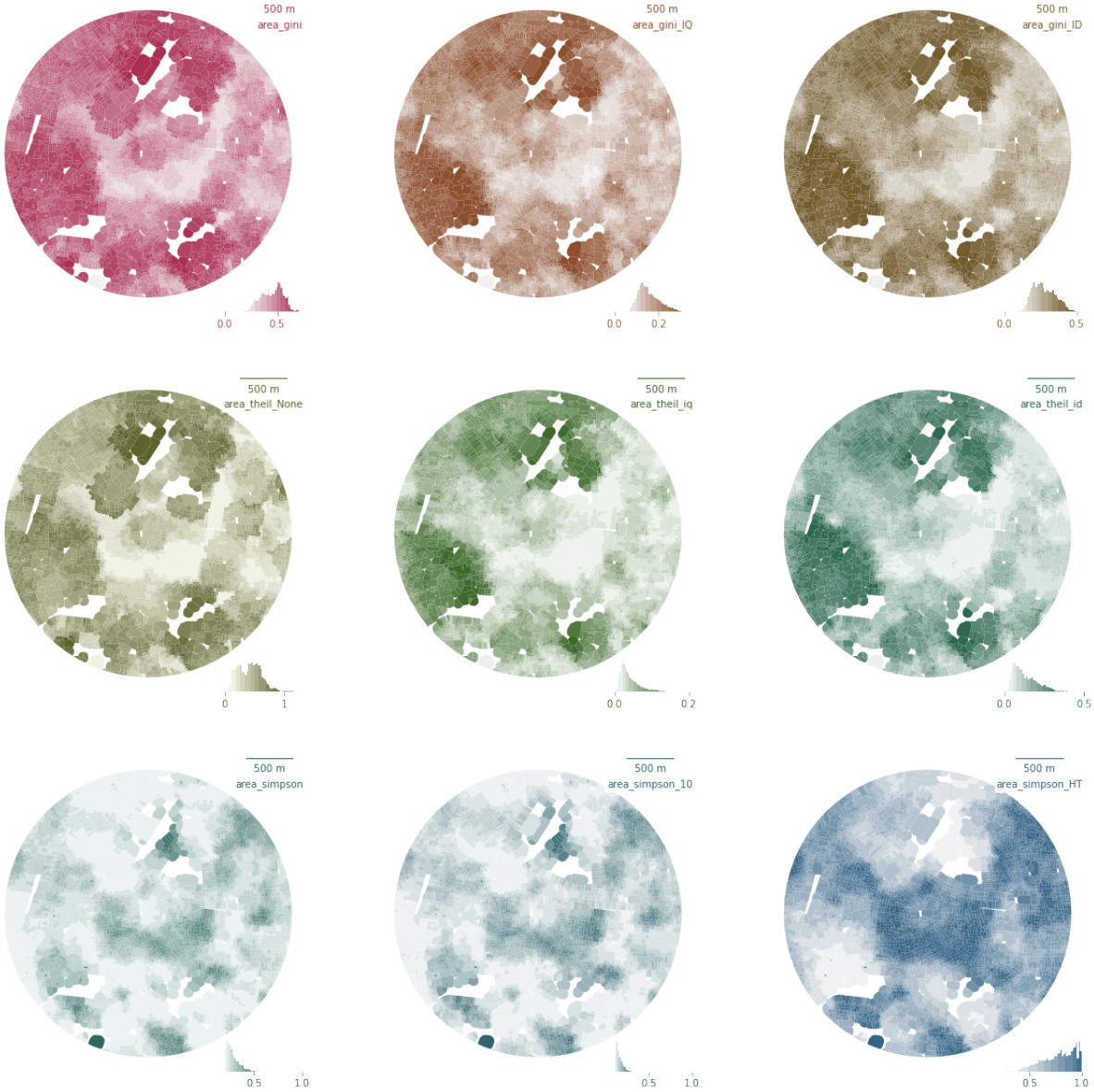


Figure A7.11: Spatial distribution of characters capturing properties of distributions tested on area of a building (cont.).

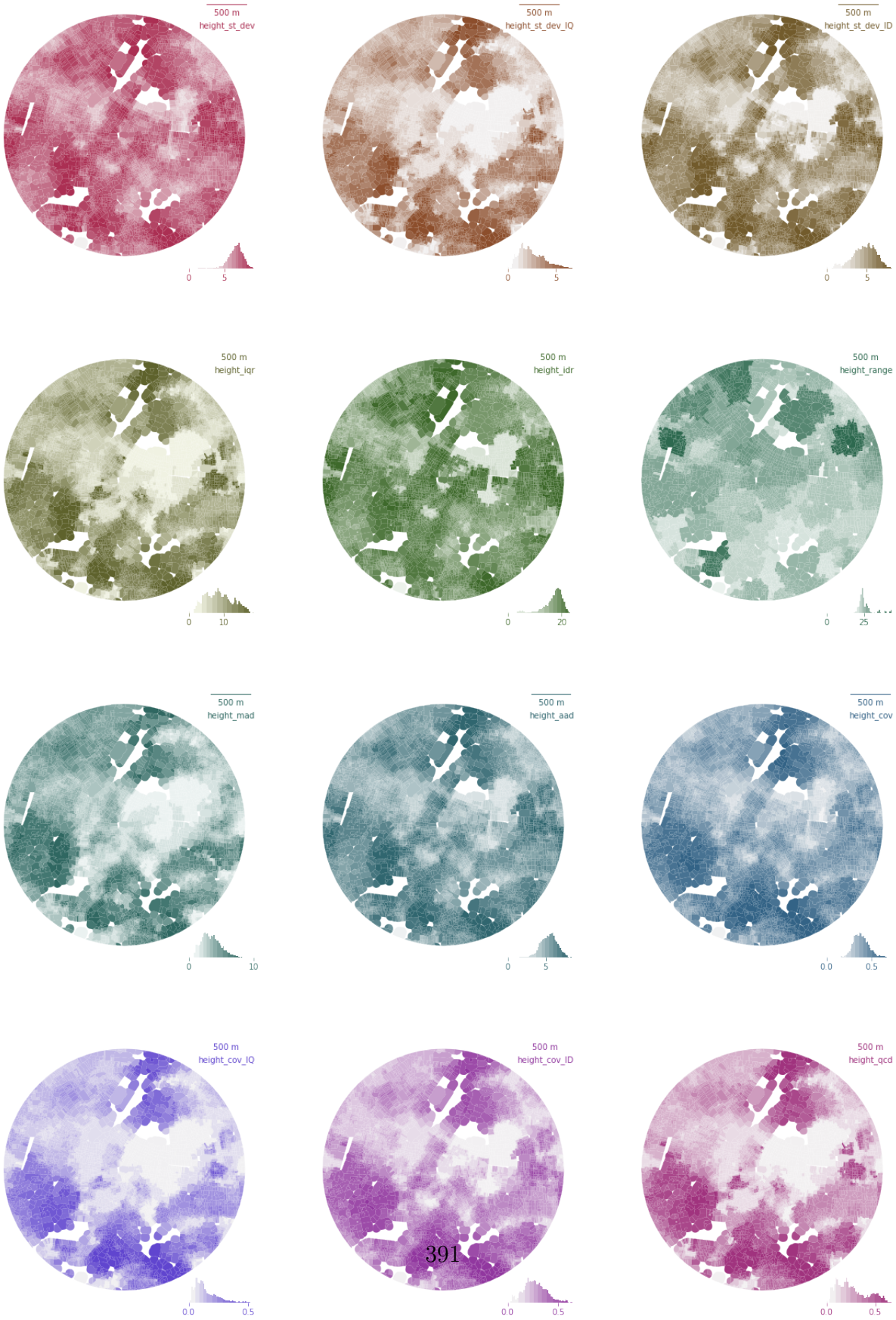


Figure A7.12: Spatial distribution of characters capturing properties of distributions tested on height of a building.

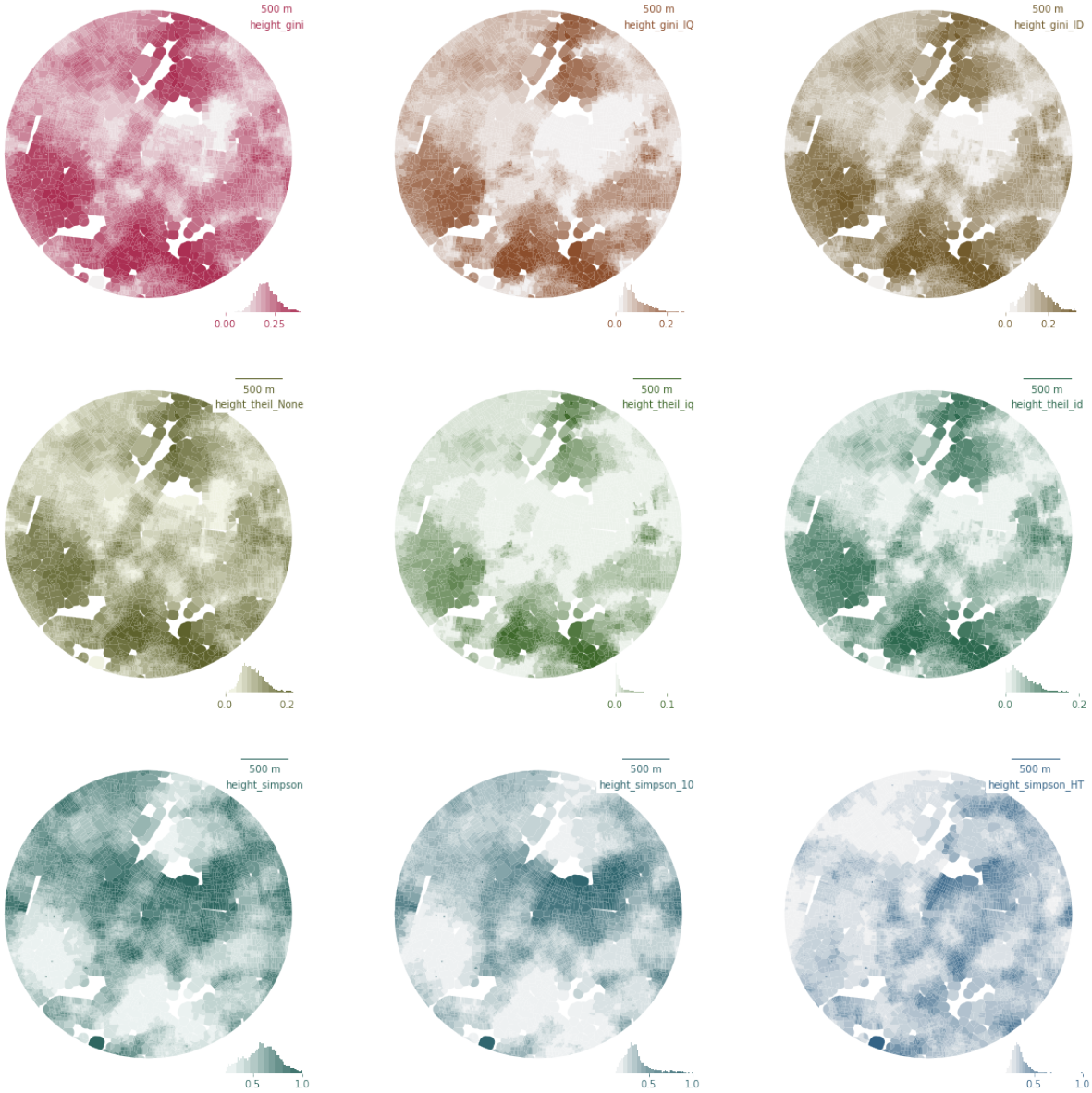


Figure A7.13: Spatial distribution of characters capturing properties of distributions tested on height of a building (cont.).

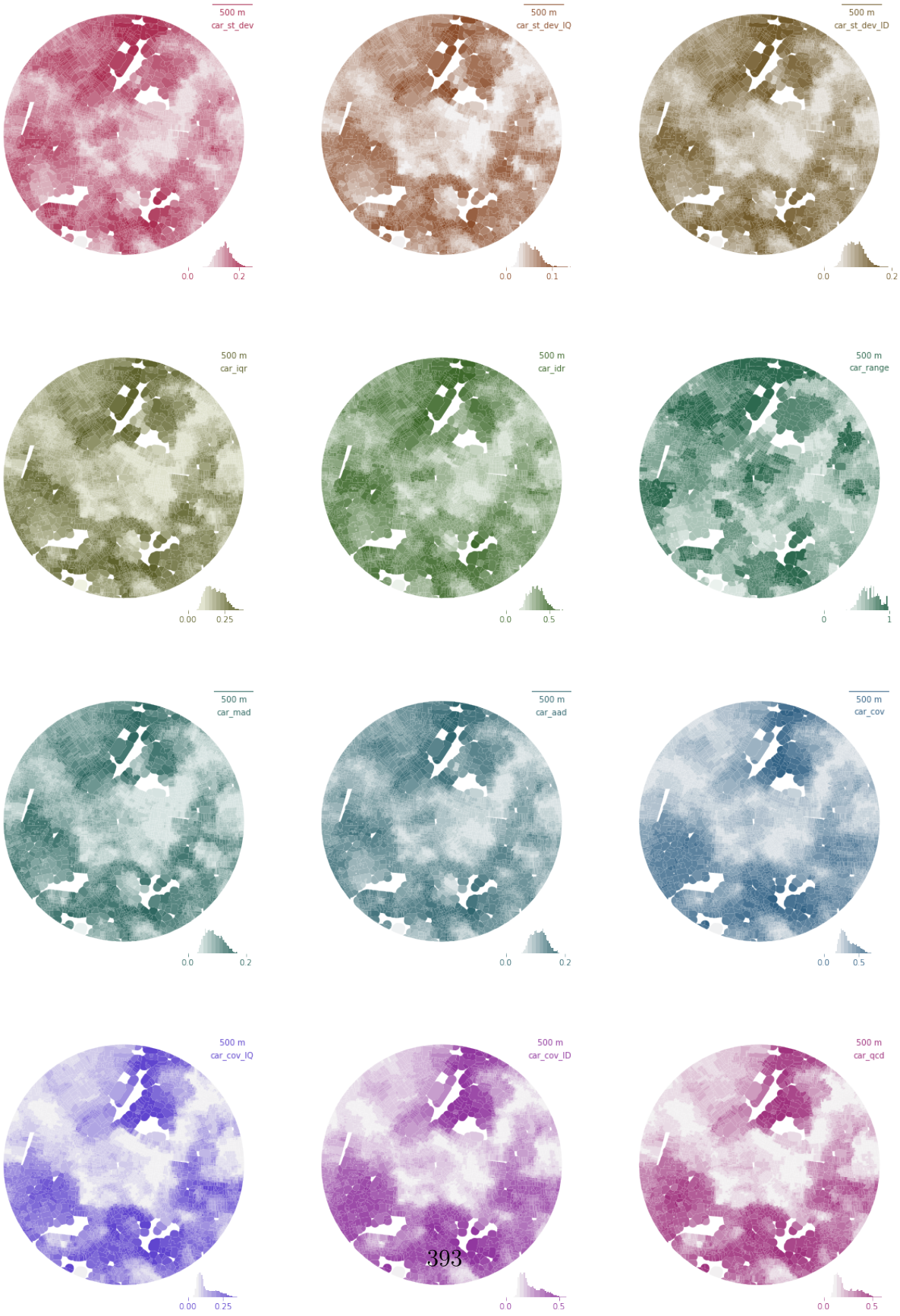


Figure A7.14: Spatial distribution of characters capturing properties of distributions tested on coverage area ratio of tessellation cell.

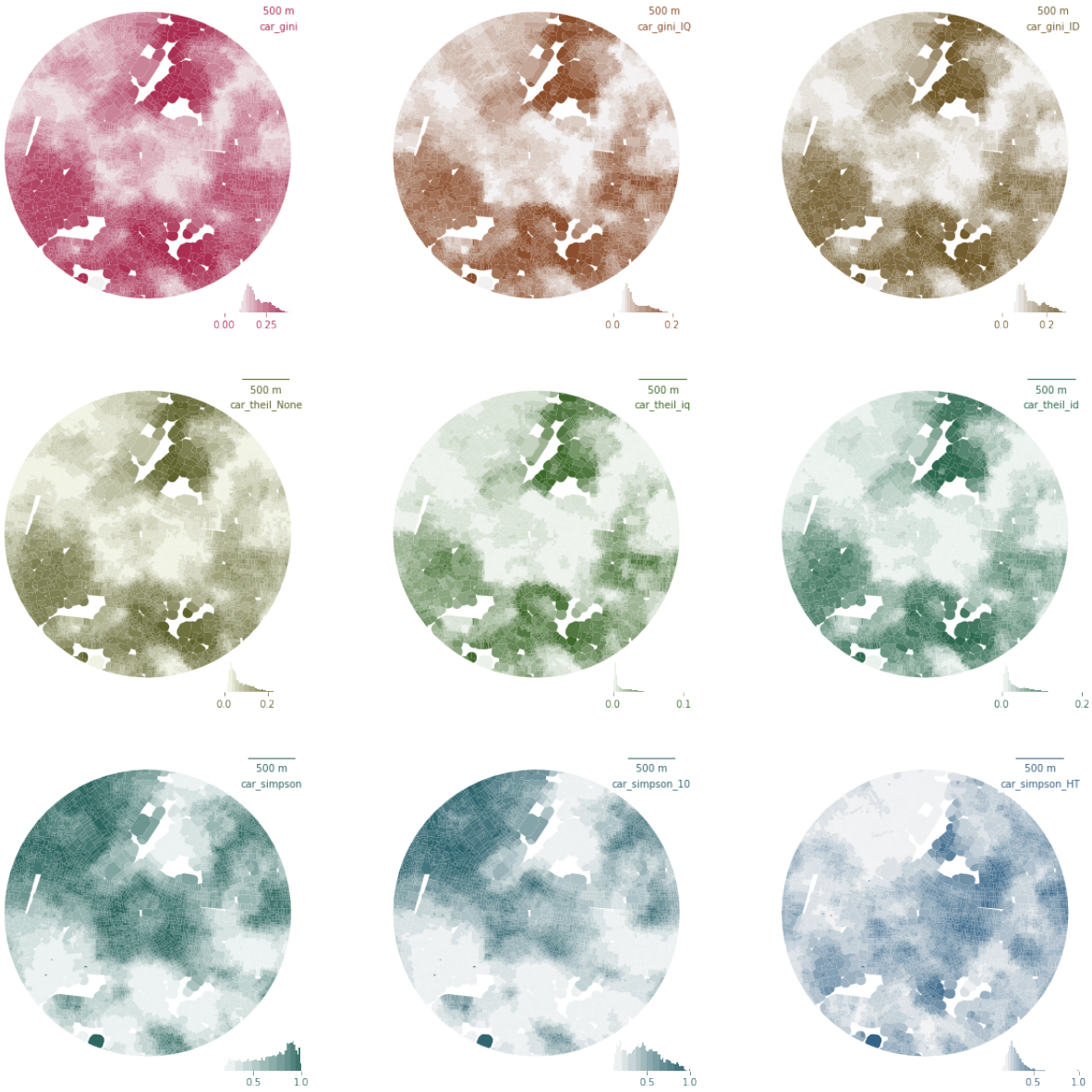


Figure A7.15: Spatial distribution of characters capturing properties of distributions tested on coverage area ratio of tessellation cell (cont.).

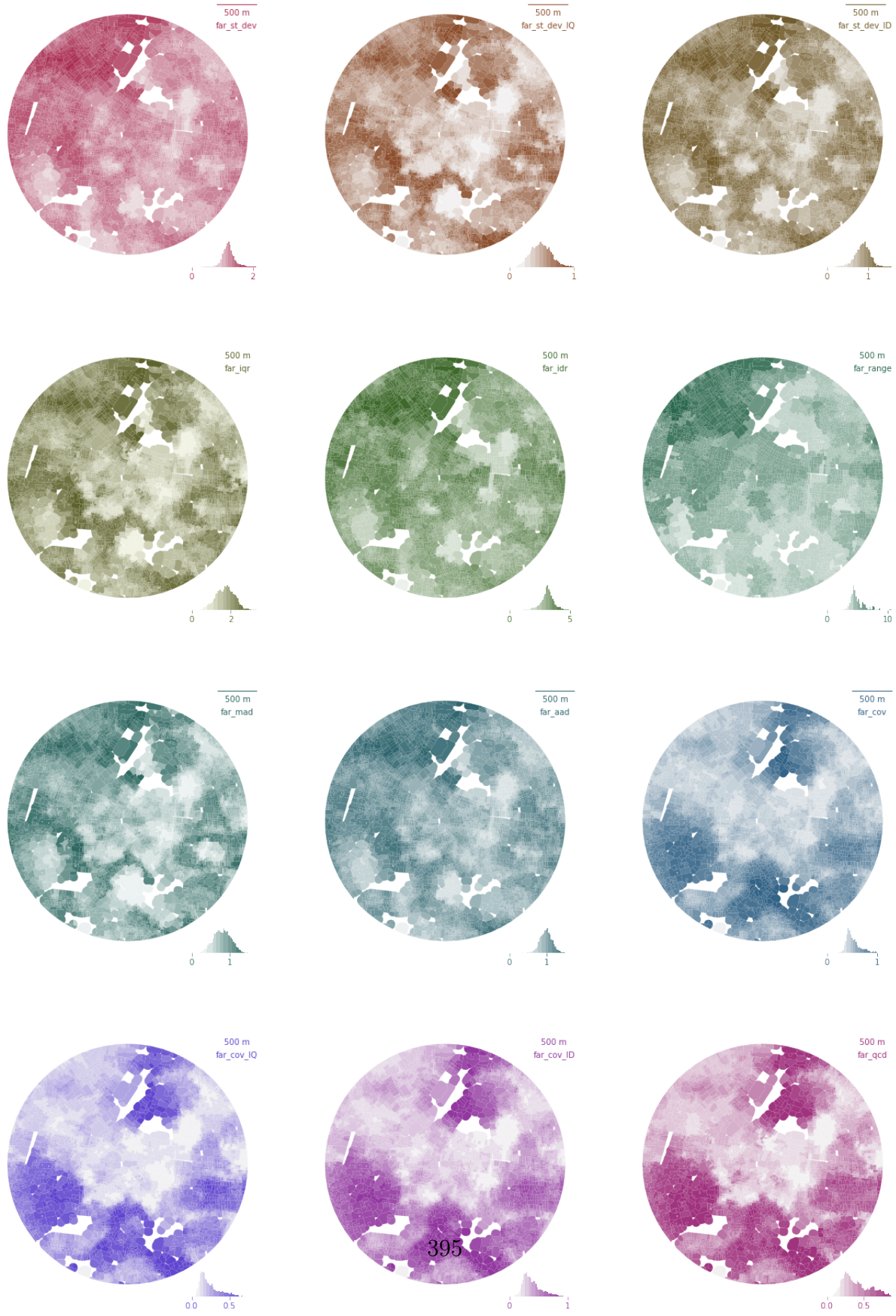


Figure A7.16: Spatial distribution of characters capturing properties of distributions tested on floor area ratio of tessellation cell.

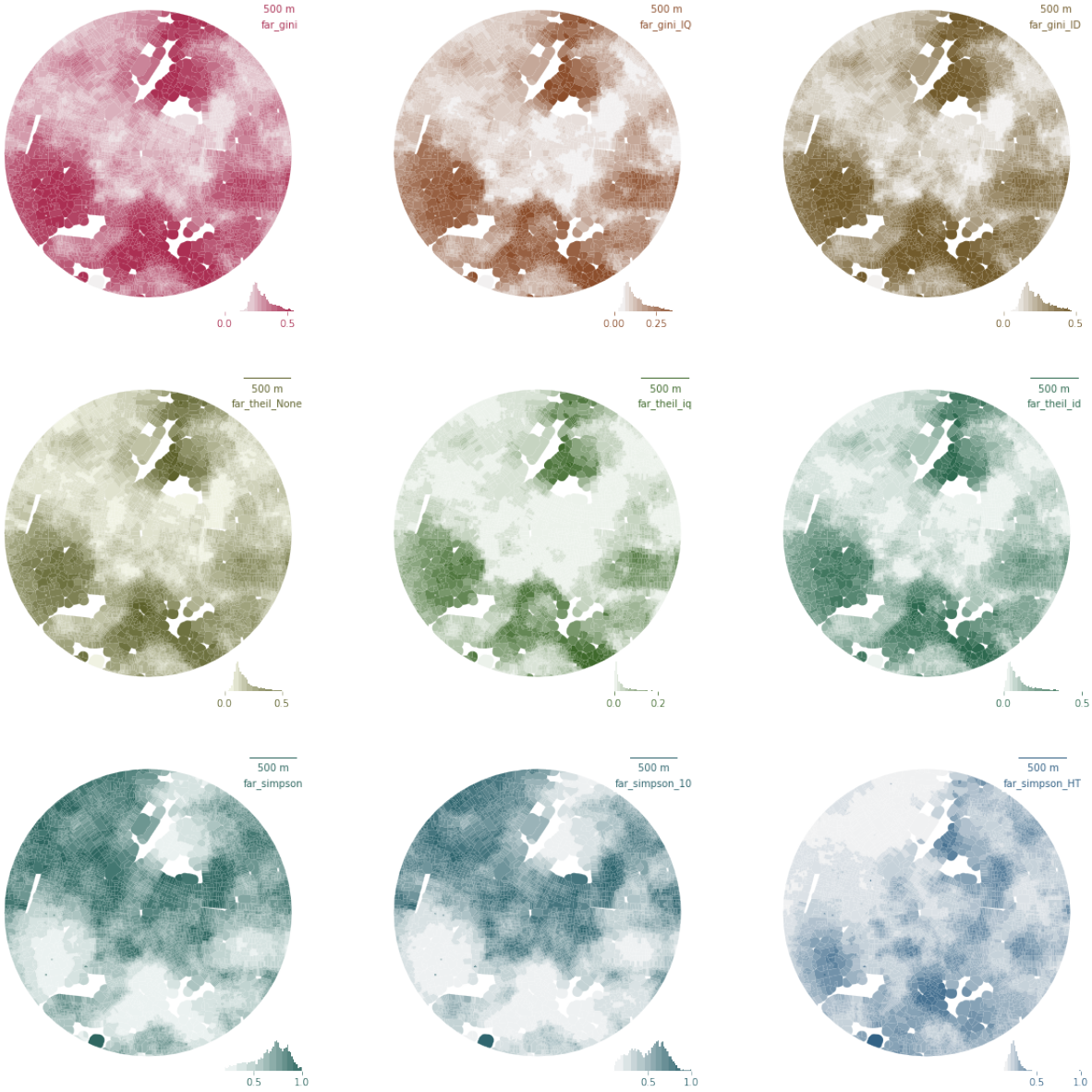


Figure A7.17: Spatial distribution of characters capturing properties of distributions tested on floor area ratio of tessellation cell (cont.).

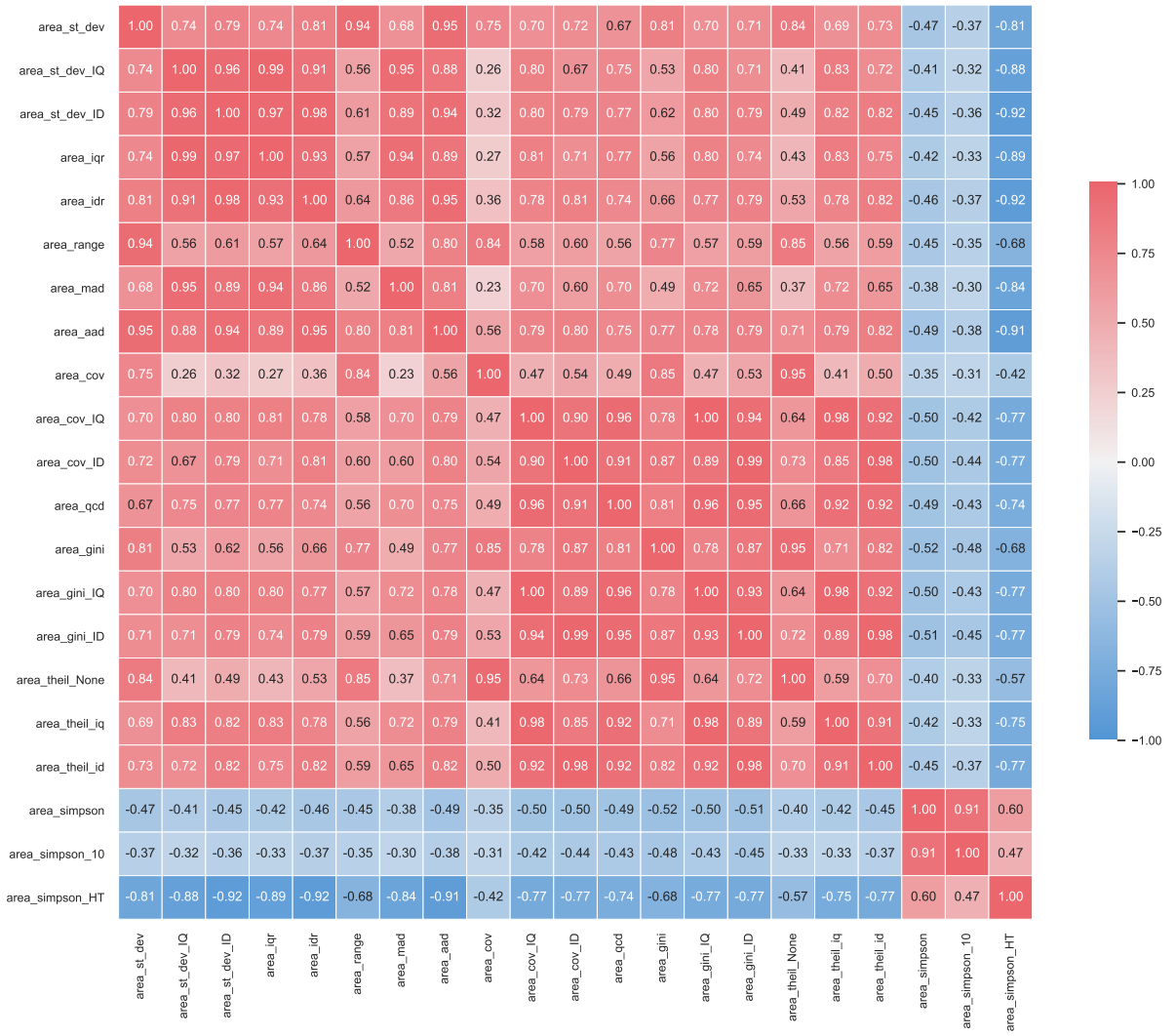


Figure A7.18: Correlation of characters capturing properties of distributions tested on area of a building.

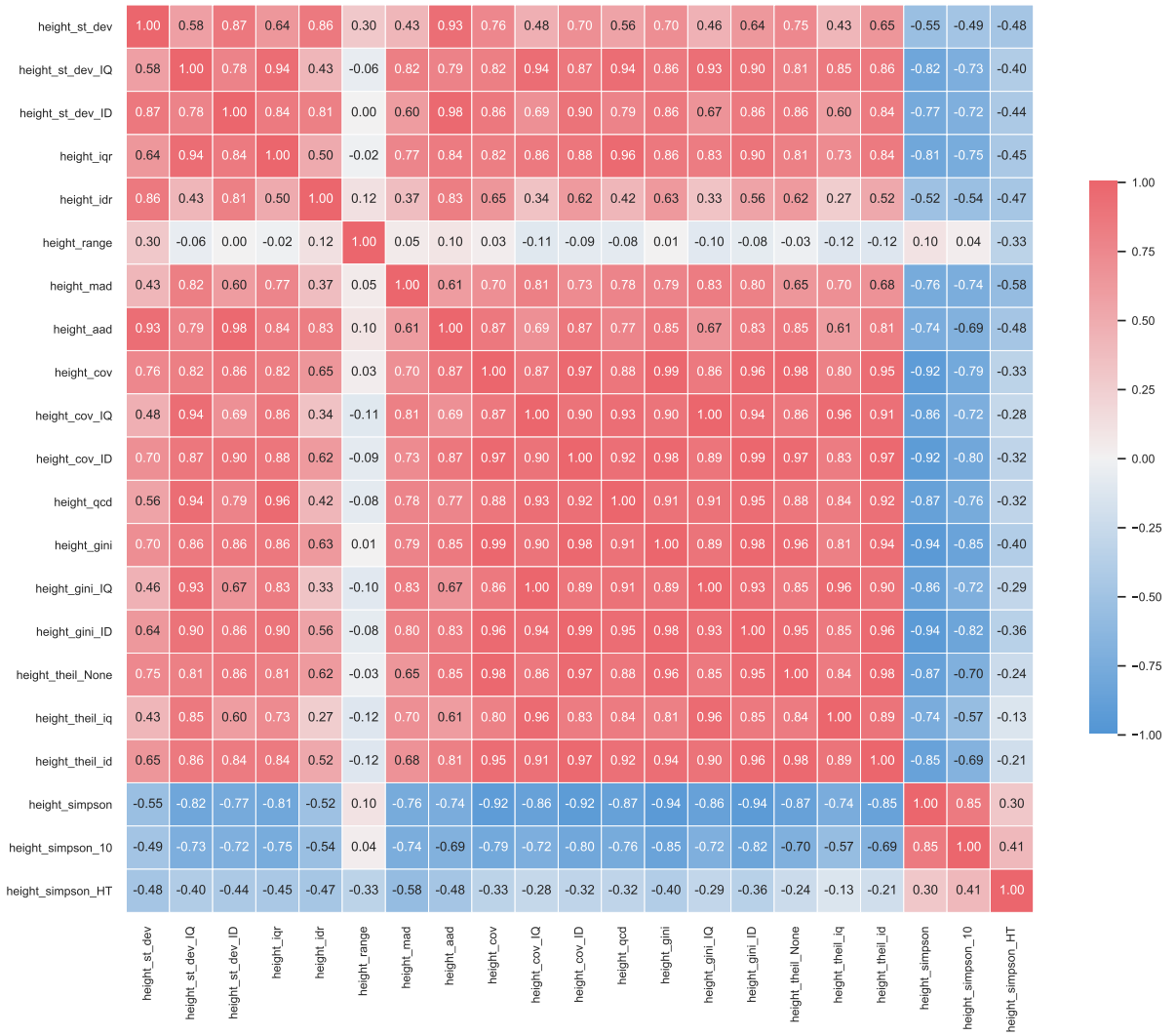


Figure A7.19: Correlation of characters capturing properties of distributions tested on height of a building.

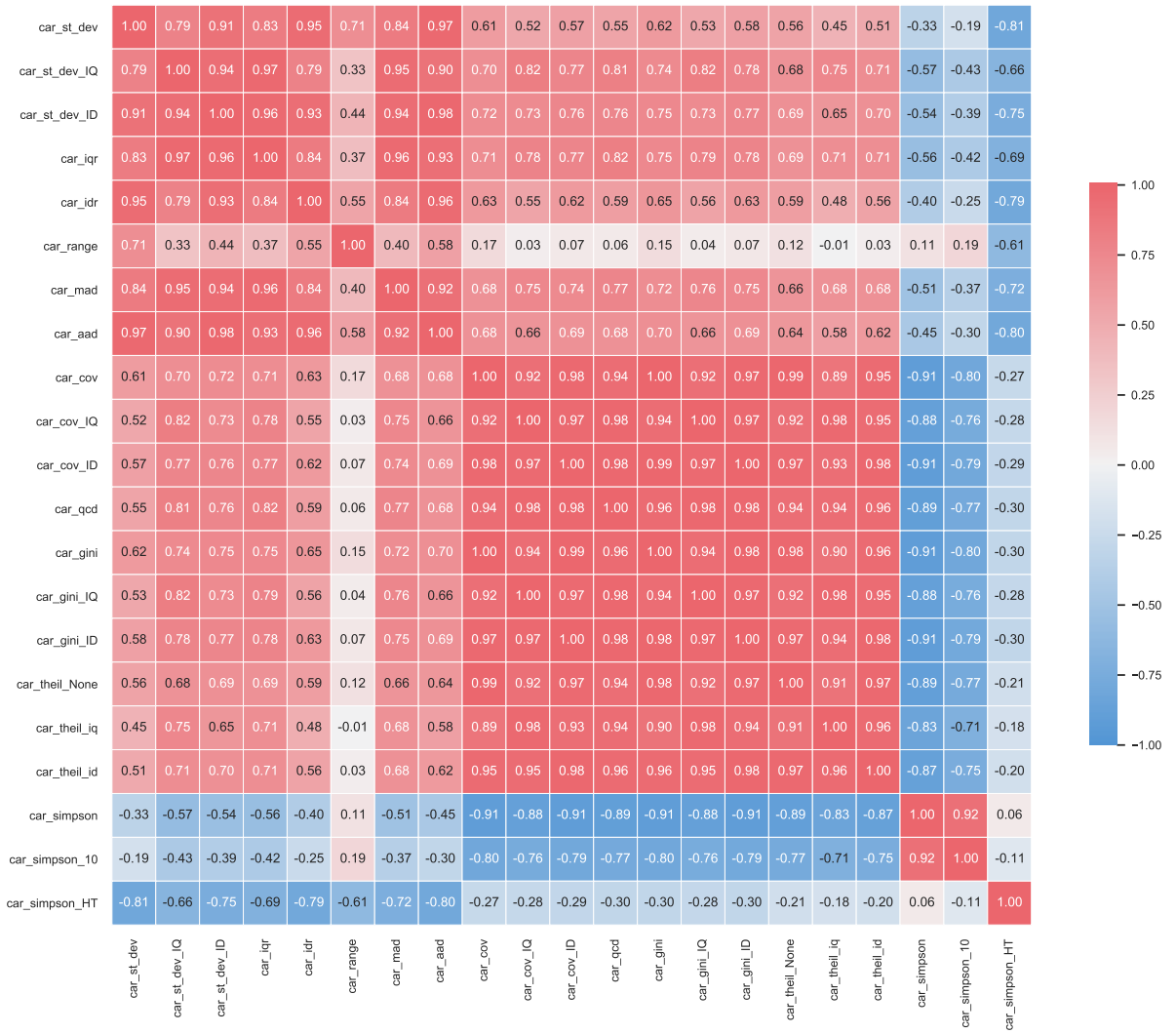


Figure A7.20: Correlation of characters capturing properties of distributions tested on coverage area ratio of tessellation cell.

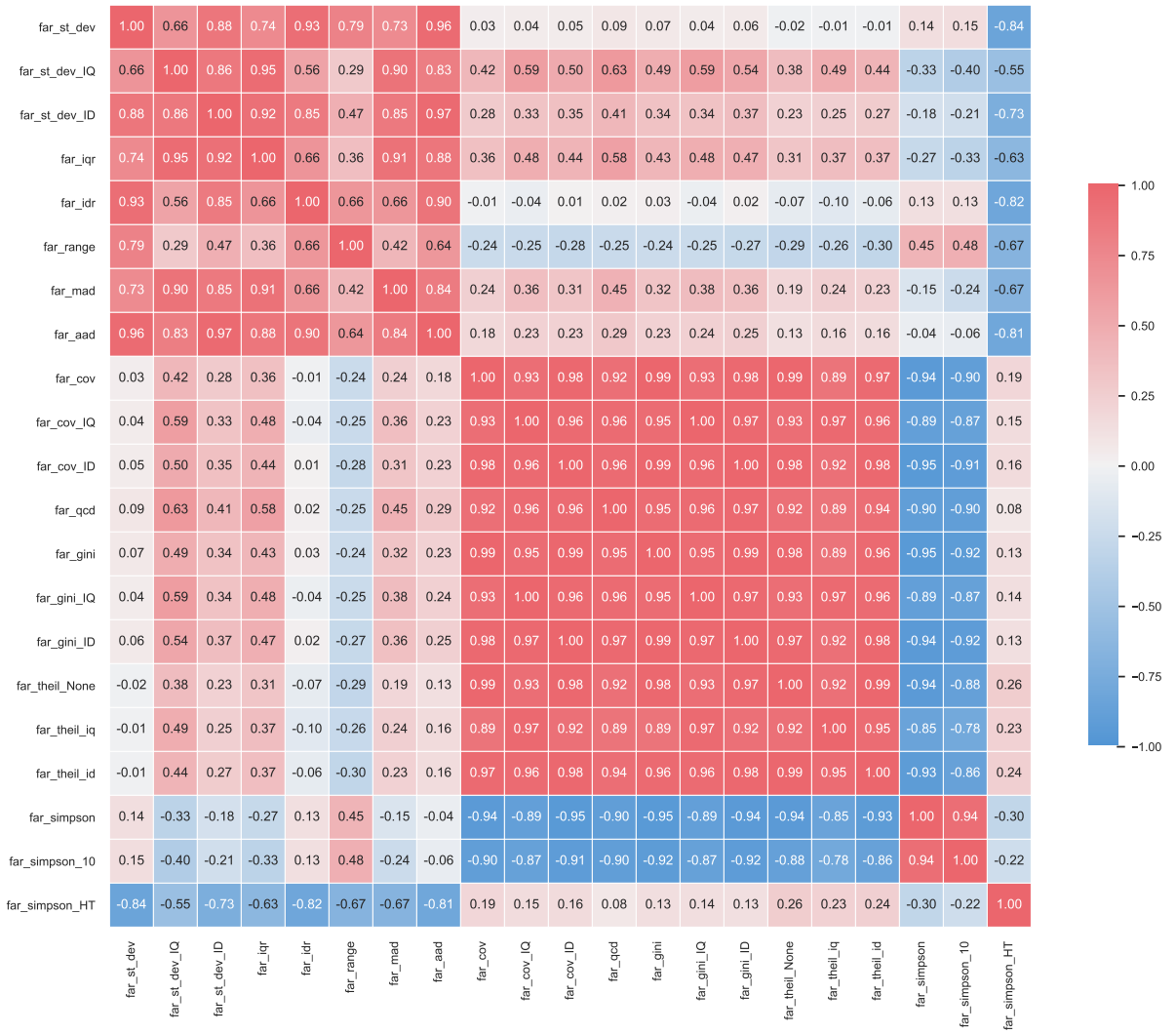


Figure A7.21: Correlation of characters capturing properties of distributions tested on floor area ratio of tessellation cell.

7.6 Spatial autocorrelation of morphometric characters

Results of assessment of spatial autocorrelation of morphometric characters. Values which show significant autocorrelation (Moran's I) tend to capture contiguous patterns.

See sections 7.1.2.1.1 and 7.1.3 for details.

Primary characters

Table A7.4: Global Moran's I spatial autocorrelation of primary characters. Key to character IDs is available in table A7.3.

	I	p_sim	p_norm
lcdMes	0.85362	0.001	0
lcnClo	0.862475	0.001	0
ldbPWL	0.592621	0.001	0
lddARe	0.798202	0.001	0
lddNDe	0.577277	0.001	0
lddRea	0.736014	0.001	0
ldkAre	0.731736	0.001	0
ldkPer	0.676249	0.001	0
ldsCDL	0.709263	0.001	0
ldsMSL	0.815984	0.001	0
ldsRea	0.810434	0.001	0
libNCo	0.710798	0.001	0
licGDe	0.938796	0.001	0
likWBB	0.655763	0.001	0
linP3W	0.852988	0.001	0
linP4W	0.911839	0.001	0
linPDE	0.815483	0.001	0
linWID	0.629364	0.001	0
lskCCo	0.533229	0.001	0
lskCWA	0.65397	0.001	0
lskERI	0.562295	0.001	0
ltbIBD	0.819981	0.001	0
ltcBuA	0.918511	0.001	0
ltcWRB	0.786888	0.001	0
ltkOri	0.607896	0.001	0
ltkWNB	0.729433	0.001	0
mdcAre	0.509591	0.001	0
mdsAre	0.780909	0.001	0
midAre	0.793405	0.001	0
midRea	0.664584	0.001	0

	I	p_sim	p_norm
misRea	0.642062	0.001	0
mtbAli	0.302355	0.001	0
mtbNDi	0.319068	0.001	0
mtbSWR	0.387207	0.001	0
mtcWNe	0.43183	0.001	0
mtdDeg	0.338198	0.001	0
mtdMDi	0.699161	0.001	0
sdbAre	0.0909169	0.001	0
sdbCoA	0.0101648	0.001	0
sdbHei	0.508558	0.001	0
sdbPer	0.171366	0.001	0
sdbVol	0.101862	0.001	0
sdcAre	0.244891	0.001	0
sdcLAL	0.377011	0.001	0
sddAre	0.762546	0.001	0
sdsAre	0.736197	0.001	0
sdsLen	0.63461	0.001	0
sdsSHD	0.45782	0.001	0
sdsSPH	0.738876	0.001	0
sdsSPO	0.542333	0.001	0
sdsSPR	0.700646	0.001	0
sdsSPW	0.395235	0.001	0
sdsSWD	0.357013	0.001	0
sicCAR	0.510195	0.001	0
sicFAR	0.60611	0.001	0
sisBpM	0.153048	0.001	0
ssbCCD	0.0972815	0.001	0
ssbCCM	0.246668	0.001	0
ssbCCo	0.159506	0.001	0
ssbCor	0.101771	0.001	0
ssbERI	0.0928927	0.001	0
ssbElo	0.171279	0.001	0

	I	p_sim	p_norm
ssbFoF	0.267186	0.001	0
ssbSqu	0.136881	0.001	0
ssbVFR	0.251518	0.001	0
sscCCo	0.152741	0.001	0
sscERI	0.0677793	0.001	0
sssLin	0.412019	0.001	0
stbCeA	0.0993836	0.001	0
stbOri	0.540843	0.001	0
stbSAl	0.285133	0.001	0
stcOri	0.291539	0.001	0
stcSAl	0.147622	0.001	0
xcnSCl	0.466779	0.001	0

Contextual characters

Table A7.5: Global Moran's I spatial autocorrelation of contextual characters. Key to character IDs is available in table A7.3.

	I	p_sim	p_norm
lcdMes_meanIQ3	0.955598	0.001	0
lcdMes_rangeIQ3	0.50685	0.001	0
lcdMes_simpson	0.689865	0.001	0
lcdMes_theilID3	0.681572	0.001	0
lcnClo_meanIQ3	0.959075	0.001	0
lcnClo_rangeIQ3	0.536619	0.001	0
lcnClo_simpson	0.726854	0.001	0
lcnClo_theilID3	0.658432	0.001	0
ldbPWL_meanIQ3	0.940908	0.001	0
ldbPWL_rangeIQ3	0.772736	0.001	0
ldbPWL_simpson	0.877407	0.001	0
ldbPWL_theilID3	0.652227	0.001	0
lddARe_meanIQ3	0.950351	0.001	0
lddARe_rangeIQ3	0.630399	0.001	0

	I	p_sim	p_norm
lddARe_simpson	0.814212	0.001	0
lddARe_theilID3	0.62819	0.001	0
lddNDe_meanIQ3	0.907499	0.001	0
lddNDe_rangeIQ3	0.620502	0.001	0
lddNDe_simpson	0.810196	0.001	0
lddNDe_theilID3	0.675864	0.001	0
lddRea_meanIQ3	0.931665	0.001	0
lddRea_rangeIQ3	0.543221	0.001	0
lddRea_simpson	0.674778	0.001	0
lddRea_theilID3	0.66969	0.001	0
ldkAre_meanIQ3	0.928541	0.001	0
ldkAre_rangeIQ3	0.570235	0.001	0
ldkAre_simpson	0.788347	0.001	0
ldkAre_theilID3	0.522912	0.001	0
ldkPer_meanIQ3	0.918906	0.001	0
ldkPer_rangeIQ3	0.585438	0.001	0
ldkPer_simpson	0.777687	0.001	0
ldkPer_theilID3	0.597642	0.001	0
ldsCDL_meanIQ3	0.921323	0.001	0
ldsCDL_rangeIQ3	0.60509	0.001	0
ldsCDL_simpson	0.810959	0.001	0
ldsCDL_theilID3	0.534048	0.001	0
ldsMSL_meanIQ3	0.944622	0.001	0
ldsMSL_rangeIQ3	0.583908	0.001	0
ldsMSL_simpson	0.799475	0.001	0
ldsMSL_theilID3	0.615352	0.001	0
ldsRea_meanIQ3	0.95118	0.001	0
ldsRea_rangeIQ3	0.600647	0.001	0
ldsRea_simpson	0.80886	0.001	0
ldsRea_theilID3	0.618526	0.001	0
libNCo_meanIQ3	0.949998	0.001	0
libNCo_rangeIQ3	0.841513	0.001	0

	I	p_sim	p_norm
libNCo_simpson	0.909637	0.001	0
libNCo_theilID3	0.561885	0.001	0
licGDe_meanIQ3	0.976722	0.001	0
licGDe_rangeIQ3	0.758523	0.001	0
licGDe_simpson	0.845604	0.001	0
licGDe_theilID3	0.60388	0.001	0
likWBB_meanIQ3	0.904411	0.001	0
likWBB_rangeIQ3	0.59347	0.001	0
likWBB_simpson	0.803015	0.001	0
likWBB_theilID3	0.599414	0.001	0
linP3W_meanIQ3	0.962415	0.001	0
linP3W_rangeIQ3	0.513581	0.001	0
linP3W_simpson	0.690559	0.001	0
linP3W_theilID3	0.56403	0.001	0
linP4W_meanIQ3	0.976016	0.001	0
linP4W_rangeIQ3	0.508686	0.001	0
linP4W_simpson	0.71637	0.001	0
linP4W_theilID3	0.646389	0.001	0
linPDE_meanIQ3	0.95383	0.001	0
linPDE_rangeIQ3	0.551748	0.001	0
linPDE_simpson	0.827951	0.001	0
linPDE_theilID3	0.627833	0.001	0
linWID_meanIQ3	0.922797	0.001	0
linWID_rangeIQ3	0.602441	0.001	0
linWID_simpson	0.815416	0.001	0
linWID_theilID3	0.595879	0.001	0
lskCCo_meanIQ3	0.882728	0.001	0
lskCCo_rangeIQ3	0.518586	0.001	0
lskCCo_simpson	0.682315	0.001	0
lskCCo_theilID3	0.714485	0.001	0
lskCWA_meanIQ3	0.913762	0.001	0
lskCWA_rangeIQ3	0.593642	0.001	0

	I	p_sim	p_norm
lskCWA_simpson	0.790063	0.001	0
lskCWA_theilID3	0.568035	0.001	0
lskERI_meanIQ3	0.897361	0.001	0
lskERI_rangeIQ3	0.566972	0.001	0
lskERI_simpson	0.671685	0.001	0
lskERI_theilID3	0.785261	0.001	0
ltbIBD_meanIQ3	0.916813	0.001	0
ltbIBD_rangeIQ3	0.641772	0.001	0
ltbIBD_simpson	0.760663	0.001	0
ltbIBD_theilID3	0.619924	0.001	0
ltcBuA_meanIQ3	0.961734	0.001	0
ltcBuA_rangeIQ3	0.631009	0.001	0
ltcBuA_simpson	0.79074	0.001	0
ltcBuA_theilID3	0.725754	0.001	0
ltcWRB_meanIQ3	0.915843	0.001	0
ltcWRB_rangeIQ3	0.737241	0.001	0
ltcWRB_simpson	0.850155	0.001	0
ltcWRB_theilID3	0.609031	0.001	0
ltkOri_meanIQ3	0.904165	0.001	0
ltkOri_rangeIQ3	0.536899	0.001	0
ltkOri_simpson	0.697909	0.001	0
ltkOri_theilID3	0.68647	0.001	0
ltkWNB_meanIQ3	0.928892	0.001	0
ltkWNB_rangeIQ3	0.54861	0.001	0
ltkWNB_simpson	0.74367	0.001	0
ltkWNB_theilID3	0.60766	0.001	0
mdcAre_meanIQ3	0.861546	0.001	0
mdcAre_rangeIQ3	0.751638	0.001	0
mdcAre_simpson	0.8328	0.001	0
mdcAre_theilID3	0.574389	0.001	0
mdsAre_meanIQ3	0.944663	0.001	0
mdsAre_rangeIQ3	0.617205	0.001	0

	I	p_sim	p_norm
mdsAre_simpson	0.807483	0.001	0
mdsAre_theilID3	0.57988	0.001	0
midAre_meanIQ3	0.947303	0.001	0
midAre_rangeIQ3	0.620141	0.001	0
midAre_simpson	0.810922	0.001	0
midAre_theilID3	0.586441	0.001	0
midRea_meanIQ3	0.904878	0.001	0
midRea_rangeIQ3	0.507217	0.001	0
midRea_simpson	0.766685	0.001	0
midRea_theilID3	0.610119	0.001	0
misRea_meanIQ3	0.903812	0.001	0
misRea_rangeIQ3	0.534213	0.001	0
misRea_simpson	0.77369	0.001	0
misRea_theilID3	0.604976	0.001	0
mtbAli_meanIQ3	0.845607	0.001	0
mtbAli_rangeIQ3	0.741485	0.001	0
mtbAli_simpson	0.83494	0.001	0
mtbAli_theilID3	0.676688	0.001	0
mtbNDi_meanIQ3	0.855789	0.001	0
mtbNDi_rangeIQ3	0.755206	0.001	0
mtbNDi_simpson	0.846039	0.001	0
mtbNDi_theilID3	0.729665	0.001	0
mtbSWR_meanIQ3	0.920196	0.001	0
mtbSWR_rangeIQ3	0.744766	0.001	0
mtbSWR_simpson	0.839345	0.001	0
mtbSWR_theilID3	0.642402	0.001	0
mtcWNe_meanIQ3	0.869615	0.001	0
mtcWNe_rangeIQ3	0.654455	0.001	0
mtcWNe_simpson	0.75302	0.001	0
mtcWNe_theilID3	0.682041	0.001	0
mtdDeg_meanIQ3	0.835449	0.001	0
mtdDeg_rangeIQ3	0.519432	0.001	0

	I	p_sim	p_norm
mtdDeg_simpson	0.705833	0.001	0
mtdDeg_theilID3	0.723812	0.001	0
mtdMDi_meanIQ3	0.925984	0.001	0
mtdMDi_rangeIQ3	0.593588	0.001	0
mtdMDi_simpson	0.806522	0.001	0
mtdMDi_theilID3	0.588363	0.001	0
sdbAre_meanIQ3	0.820124	0.001	0
sdbAre_rangeIQ3	0.756028	0.001	0
sdbAre_simpson	0.900207	0.001	0
sdbAre_theilID3	0.743363	0.001	0
sdbCoA_meanIQ3	0.676399	0.001	0
sdbCoA_rangeIQ3	0.472574	0.001	0
sdbCoA_simpson	0.902308	0.001	0
sdbCoA_theilID3	0.74238	0.001	0
sdbHei_meanIQ3	0.941584	0.001	0
sdbHei_rangeIQ3	0.78105	0.001	0
sdbHei_simpson	0.92606	0.001	0
sdbHei_theilID3	0.756276	0.001	0
sdbPer_meanIQ3	0.880991	0.001	0
sdbPer_rangeIQ3	0.809868	0.001	0
sdbPer_simpson	0.881562	0.001	0
sdbPer_theilID3	0.770573	0.001	0
sdbVol_meanIQ3	0.840984	0.001	0
sdbVol_rangeIQ3	0.764246	0.001	0
sdbVol_simpson	0.904012	0.001	0
sdbVol_theilID3	0.708305	0.001	0
sdcAre_meanIQ3	0.846201	0.001	0
sdcAre_rangeIQ3	0.788472	0.001	0
sdcAre_simpson	0.833662	0.001	0
sdcAre_theilID3	0.64362	0.001	0
sdcLAL_meanIQ3	0.858866	0.001	0
sdcLAL_rangeIQ3	0.721431	0.001	0

	I	p_sim	p_norm
sdcLAL_simpson	0.823481	0.001	0
sdcLAL_theilID3	0.657668	0.001	0
sddAre_meanIQ3	0.937527	0.001	0
sddAre_rangeIQ3	0.611057	0.001	0
sddAre_simpson	0.806723	0.001	0
sddAre_theilID3	0.548801	0.001	0
sdsAre_meanIQ3	0.925027	0.001	0
sdsAre_rangeIQ3	0.606761	0.001	0
sdsAre_simpson	0.805935	0.001	0
sdsAre_theilID3	0.524675	0.001	0
sdsLen_meanIQ3	0.91244	0.001	0
sdsLen_rangeIQ3	0.598489	0.001	0
sdsLen_simpson	0.796079	0.001	0
sdsLen_theilID3	0.576659	0.001	0
sdsSHD_meanIQ3	0.891916	0.001	0
sdsSHD_rangeIQ3	0.718592	0.001	0
sdsSHD_simpson	0.876636	0.001	0
sdsSHD_theilID3	0.68296	0.001	0
sdsSPH_meanIQ3	0.954994	0.001	0
sdsSPH_rangeIQ3	0.685786	0.001	0
sdsSPH_simpson	0.885641	0.001	0
sdsSPH_theilID3	0.659245	0.001	0
sdsSPO_meanIQ3	0.910225	0.001	0
sdsSPO_rangeIQ3	0.525057	0.001	0
sdsSPO_simpson	0.715839	0.001	0
sdsSPO_theilID3	0.773536	0.001	0
sdsSPR_meanIQ3	0.964001	0.001	0
sdsSPR_rangeIQ3	0.725299	0.001	0
sdsSPR_simpson	0.889633	0.001	0
sdsSPR_theilID3	0.64308	0.001	0
sdsSPW_meanIQ3	0.879784	0.001	0
sdsSPW_rangeIQ3	0.540892	0.001	0

	I	p_sim	p_norm
sdsSPW_simpson	0.695483	0.001	0
sdsSPW_theilID3	0.683547	0.001	0
sdsSWD_meanIQ3	0.859928	0.001	0
sdsSWD_rangeIQ3	0.544153	0.001	0
sdsSWD_simpson	0.706063	0.001	0
sdsSWD_theilID3	0.683033	0.001	0
sicCAR_meanIQ3	0.942188	0.001	0
sicCAR_rangeIQ3	0.775051	0.001	0
sicCAR_simpson	0.877473	0.001	0
sicCAR_theilID3	0.771826	0.001	0
sicFAR_meanIQ3	0.964149	0.001	0
sicFAR_rangeIQ3	0.889602	0.001	0
sicFAR_simpson	0.934126	0.001	0
sicFAR_theilID3	0.74142	0.001	0
sisBpM_meanIQ3	0.834023	0.001	0
sisBpM_rangeIQ3	0.520404	0.001	0
sisBpM_simpson	0.774814	0.001	0
sisBpM_theilID3	0.593089	0.001	0
ssbCCD_meanIQ3	0.849419	0.001	0
ssbCCD_rangeIQ3	0.780761	0.001	0
ssbCCD_simpson	0.84992	0.001	0
ssbCCD_theilID3	0.813876	0.001	0
ssbCCM_meanIQ3	0.889831	0.001	0
ssbCCM_rangeIQ3	0.819487	0.001	0
ssbCCM_simpson	0.881239	0.001	0
ssbCCM_theilID3	0.78196	0.001	0
ssbCCo_meanIQ3	0.864255	0.001	0
ssbCCo_rangeIQ3	0.749751	0.001	0
ssbCCo_simpson	0.788952	0.001	0
ssbCCo_theilID3	0.790146	0.001	0
ssbCor_meanIQ3	0.841669	0.001	0
ssbCor_rangeIQ3	0.697948	0.001	0

	I	p_sim	p_norm
ssbCor_simpson	0.826181	0.001	0
ssbCor_theilID3	0.688075	0.001	0
ssbERI_meanIQ3	0.856649	0.001	0
ssbERI_rangeIQ3	0.808194	0.001	0
ssbERI_simpson	0.812285	0.001	0
ssbERI_theilID3	0.848991	0.001	0
ssbElo_meanIQ3	0.864312	0.001	0
ssbElo_rangeIQ3	0.698684	0.001	0
ssbElo_simpson	0.765896	0.001	0
ssbElo_theilID3	0.791906	0.001	0
ssbFoF_meanIQ3	0.893053	0.001	0
ssbFoF_rangeIQ3	0.777009	0.001	0
ssbFoF_simpson	0.857507	0.001	0
ssbFoF_theilID3	0.805237	0.001	0
ssbSqu_meanIQ3	0.89365	0.001	0
ssbSqu_rangeIQ3	0.827903	0.001	0
ssbSqu_simpson	0.873586	0.001	0
ssbSqu_theilID3	0.659983	0.001	0
ssbVFR_meanIQ3	0.892769	0.001	0
ssbVFR_rangeIQ3	0.777149	0.001	0
ssbVFR_simpson	0.876502	0.001	0
ssbVFR_theilID3	0.772907	0.001	0
sscCCo_meanIQ3	0.837064	0.001	0
sscCCo_rangeIQ3	0.633735	0.001	0
sscCCo_simpson	0.730701	0.001	0
sscCCo_theilID3	0.781972	0.001	0
sscERI_meanIQ3	0.814612	0.001	0
sscERI_rangeIQ3	0.691816	0.001	0
sscERI_simpson	0.720924	0.001	0
sscERI_theilID3	0.75613	0.001	0
sssLin_meanIQ3	0.834239	0.001	0
sssLin_rangeIQ3	0.626923	0.001	0

	I	p_sim	p_norm
sssLin_simpson	0.815024	0.001	0
sssLin_theilID3	0.731798	0.001	0
stbCeA_meanIQ3	0.833449	0.001	0
stbCeA_rangeIQ3	0.75756	0.001	0
stbCeA_simpson	0.838166	0.001	0
stbCeA_theilID3	0.784623	0.001	0
stbOri_meanIQ3	0.909016	0.001	0
stbOri_rangeIQ3	0.639928	0.001	0
stbOri_simpson	0.797732	0.001	0
stbOri_theilID3	0.752463	0.001	0
stbSAl_meanIQ3	0.844173	0.001	0
stbSAl_rangeIQ3	0.694922	0.001	0
stbSAl_simpson	0.827305	0.001	0
stbSAl_theilID3	0.61161	0.001	0
stcOri_meanIQ3	0.890039	0.001	0
stcOri_rangeIQ3	0.696291	0.001	0
stcOri_simpson	0.795035	0.001	0
stcOri_theilID3	0.805094	0.001	0
stcSAl_meanIQ3	0.832939	0.001	0
stcSAl_rangeIQ3	0.719319	0.001	0
stcSAl_simpson	0.821816	0.001	0
stcSAl_theilID3	0.739995	0.001	0
xcnSCl_meanIQ3	0.838289	0.001	0
xcnSCl_rangeIQ3	0.592369	0.001	0
xcnSCl_simpson	0.780986	0.001	0
xcnSCl_theilID3	0.422105	0.001	0

7.7 Statistical overview of contextual characters results

Interquartile mean

Table A7.6: Overview of the contextual morphometric values of interquartile mean for the whole case study. Key to character IDs is available in table A7.3.

	mean	std	min	25%	50%	75%	max
stcOri	18	6.8	0.12	13	18	23	42
sdcLAL	69	25	25	51	64	82	310
sdcAre	2400	2100	180	1100	1700	2900	50000
sscCCo	0.45	0.055	0.14	0.42	0.46	0.49	0.74
sscERI	0.97	0.018	0.86	0.96	0.97	0.98	1.1
stcSAl	9.3	3.7	0.12	6.7	9	12	36
sicCAR	0.19	0.1	0.0022	0.13	0.16	0.21	0.73
sicFAR	0.66	0.7	0.0022	0.25	0.39	0.73	4.4
mtcWNe	0.045	0.013	0.0016	0.036	0.045	0.054	0.15
mdcAre	18000	14000	1600	8700	14000	22000	370000
licGDe	0.57	0.64	0.0022	0.2	0.36	0.65	4.1
ltcWRB	8.8e-05	5.6e-05	2.7e-06	4.6e-05	7.8e-05	0.00012	0.00048
sdbHei	9.9	4.7	3	6.3	8.1	12	37
sdbAre	310	330	33	130	200	360	10000
sdbVol	3800	4600	130	910	2000	5100	150000
sdbPer	68	26	24	49	60	79	460
sdbCoA	3.3	15	0	0	0	0	340
ssbFoF	1.4	0.3	0.73	1.2	1.4	1.6	4.8
ssbVFR	3.1	0.9	1.4	2.5	2.9	3.7	19
ssbCCo	0.53	0.044	0.27	0.5	0.53	0.56	0.72
ssbCor	9.2	2.7	4	7.5	8.6	10	54
ssbSqu	5.4	3.5	0.016	2.9	4.5	7.1	39
ssbERI	0.93	0.03	0.7	0.92	0.94	0.95	1.1
ssbElo	0.7	0.085	0.23	0.65	0.72	0.76	0.96
ssbCCM	9.8	3.5	3.8	7.3	8.9	11	59
ssbCCD	1.7	0.81	0.00036	1.1	1.5	2	16

	mean	std	min	25%	50%	75%	max
stbOri	16	8.8	0.026	9.8	15	22	44
stbSAl	6.8	4.5	0.038	3.3	6.1	9.3	39
stbCeA	7	3	0.058	4.8	6.8	9	27
mtbSWR	0.17	0.12	0	0.077	0.14	0.25	0.75
mtbAli	4.9	2.7	0.005	2.8	4.7	6.7	38
mtbNDi	26	10	0	19	25	32	120
libNCo	0.59	2.7	0	0	0	0.049	51
ldbPWL	180	190	24	69	110	210	2100
ltbIBD	27	9.7	0	21	26	33	120
ltcBuA	0.65	0.22	0.083	0.5	0.69	0.82	1
mtdDeg	3.1	0.44	1	2.9	3.1	3.4	4.8
lcdMes	0.15	0.054	-0.23	0.11	0.15	0.19	0.32
linP3W	0.64	0.1	0	0.57	0.64	0.71	0.93
linP4W	0.23	0.11	0	0.15	0.22	0.3	0.72
linPDE	0.13	0.077	0	0.072	0.11	0.17	1
lcnClo	5.3e-06	2.3e-06	6.8e-08	3.6e-06	5.1e-06	6.7e-06	1.7e-05
ldsCDL	280	310	0	78	190	370	3600
xcnSCl	0.056	0.055	0	0.012	0.046	0.083	0.75
mtdMDi	170	120	36	110	140	190	3300
lddNDe	0.013	0.004	0.0028	0.01	0.012	0.014	0.063
linWID	0.025	0.0079	0	0.02	0.024	0.029	0.11
lddRea	190	71	2	140	190	230	630
lddARe	370000	270000	39000	230000	300000	410000	3.8e+06
sddAre	30000	39000	3000	12000	19000	32000	660000
midRea	52	22	2	38	49	62	270
midAre	97000	96000	12000	51000	70000	110000	1.2e+06
sdsLen	230	200	36	140	180	260	3300
sdsSPW	29	5	11	26	29	32	50
sdsSPH	10	5.1	0	6.6	8.5	13	38
sdsSPR	0.41	0.27	0	0.24	0.31	0.47	2
sdsSPO	0.58	0.14	0.027	0.5	0.58	0.66	1
sdsSWD	3.6	1.2	0	2.8	3.6	4.4	9.8

	mean	std	min	25%	50%	75%	max
sdsSHD	2.3	1.5	0	1.2	1.8	2.9	14
sssLin	0.95	0.077	0	0.93	0.97	0.99	1
sdsAre	31000	46000	2100	10000	17000	32000	740000
sisBpM	0.074	0.031	0.0013	0.056	0.07	0.086	0.79
misRea	44	19	2	32	40	52	230
mdsAre	86000	92000	8600	41000	59000	95000	1.1e+06
ldsMSL	150	67	57	110	140	170	1200
ldsRea	350000	270000	39000	220000	280000	390000	3.8e+06
ldkAre	120000	200000	3200	26000	56000	130000	2e+06
ldkPer	1500	1400	250	720	1100	1800	13000
lskCCo	0.43	0.09	0.13	0.37	0.43	0.5	0.98
lskERI	0.86	0.096	0.36	0.81	0.88	0.93	1.1
lskCWA	370	370	0.43	140	240	450	3100
ltkOri	18	9.4	0.034	10	17	25	45
ltkWNB	0.0074	0.0035	0	0.0047	0.0071	0.0097	0.025
likWBB	0.00088	0.00051	3e-05	0.00051	0.0008	0.0012	0.004

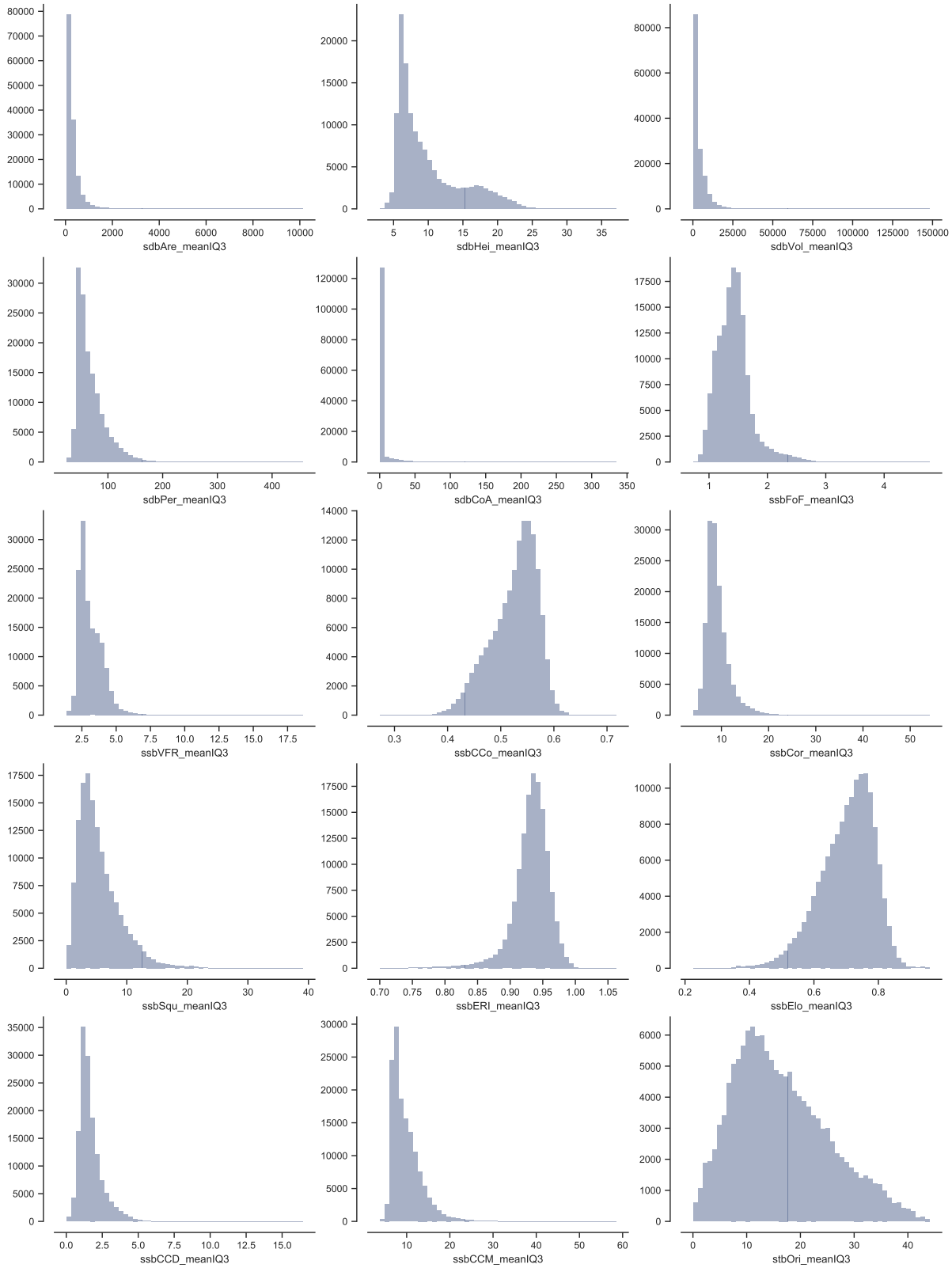


Figure A7.22: Histograms of interquartile mean for characters 1-15 are showing the variety of distributions within the measured contextual data.

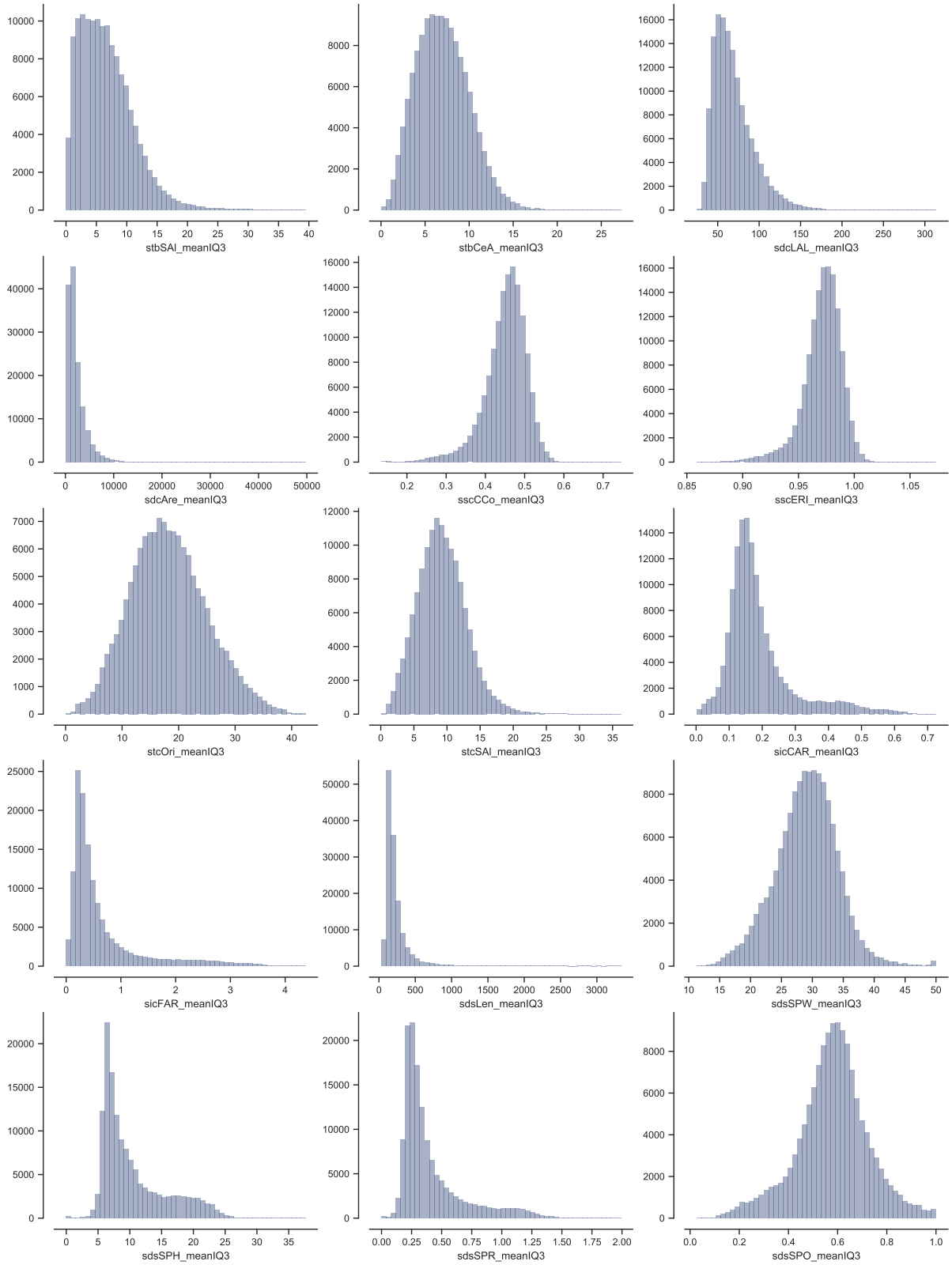


Figure A7.23: Histograms of interquartile mean for characters 16-30 are showing the variety of distributions within the measured contextual data.

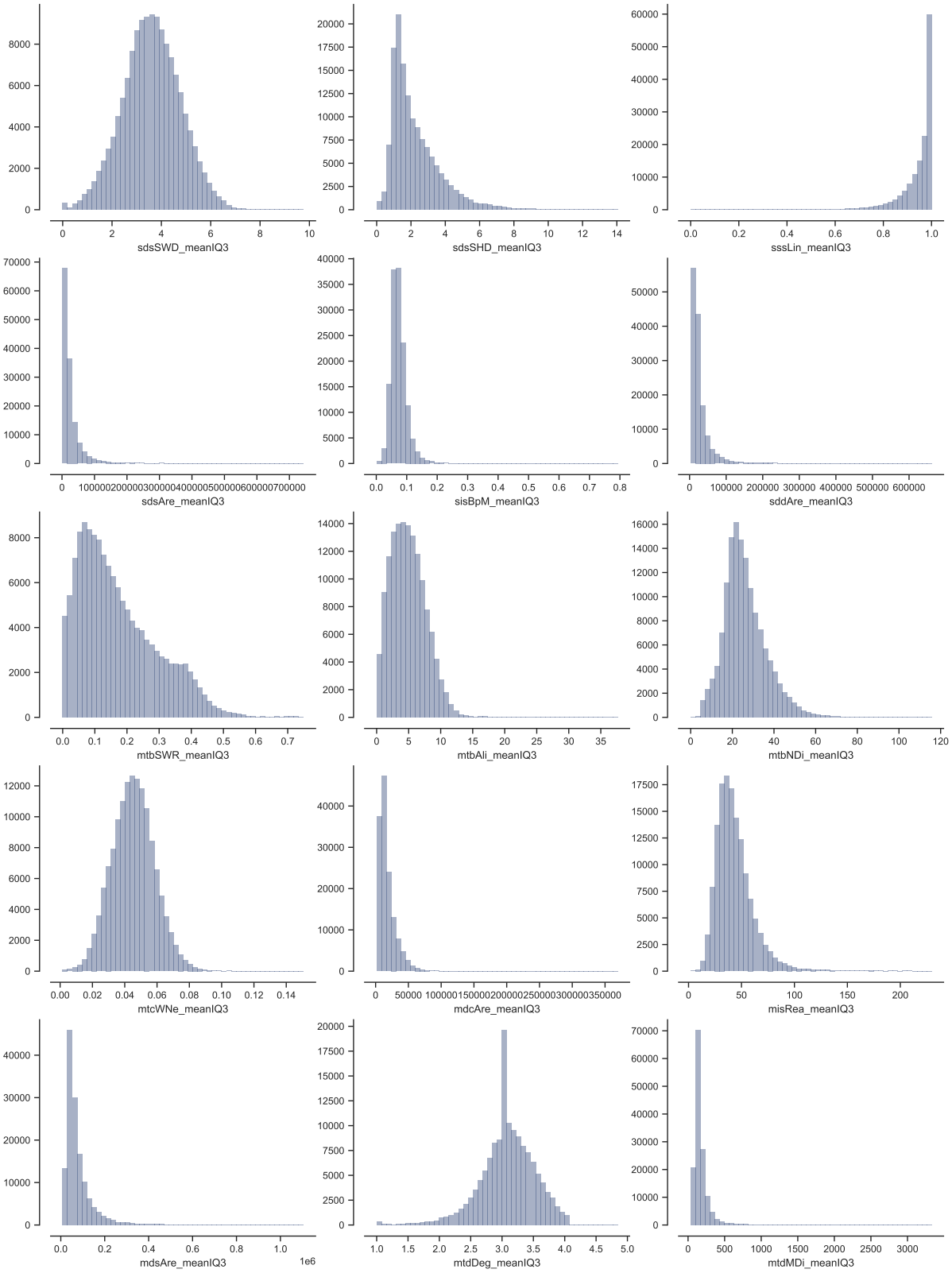


Figure A7.24: Histograms of interquartile mean for characters 31-45 are showing the variety of distributions within the measured contextual data.

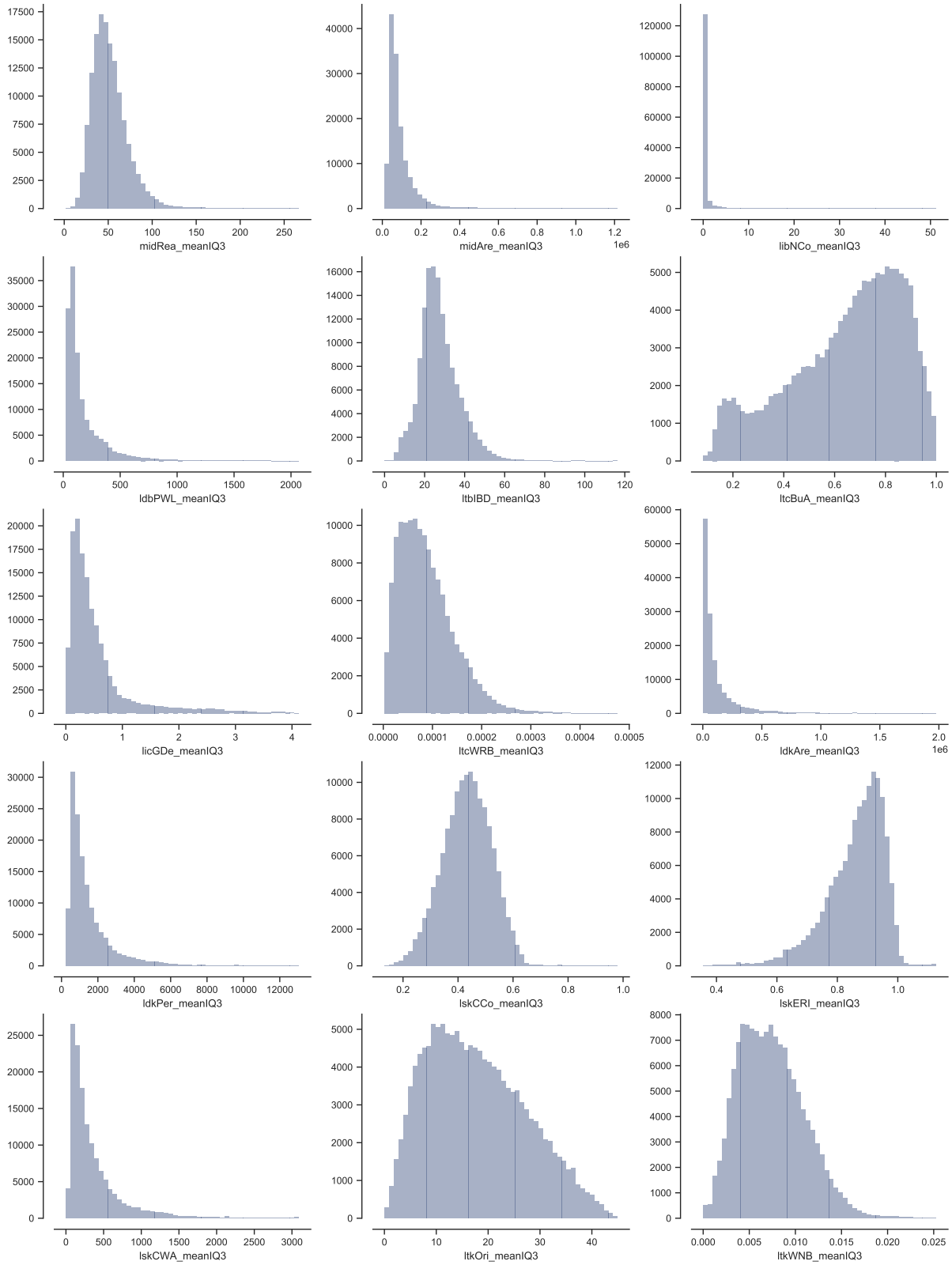


Figure A7.25: Histograms of interquartile mean for characters 46-60 are showing the variety of distributions within the measured contextual data.

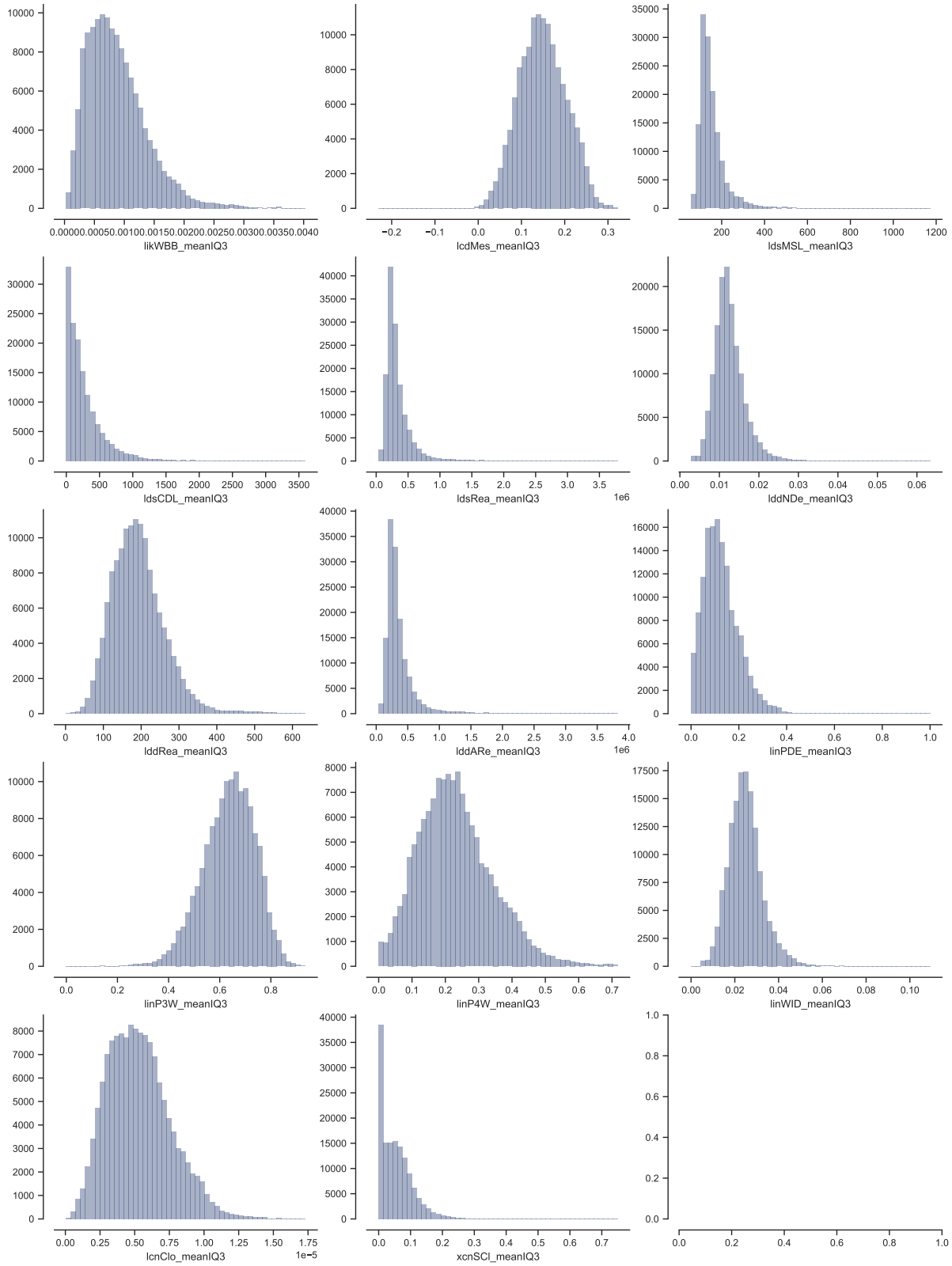


Figure A7.26: Histograms of interquartile mean for characters 61-74 are showing the variety of distributions within the measured contextual data.

Interquartile range

Table A7.7: Overview of the contextual morphometric values of interquartile range for the whole case study. Key to character IDs is available in table A7.3.

	mean	std	min	25%	50%	75%	max
stcOri	14	7.4	0.013	8.8	14	19	45
sdcLAL	36	23	0.045	17	30	51	160
sdcAre	1800	2000	3.7	590	1100	2200	47000
sscCCo	0.19	0.052	0.00011	0.15	0.18	0.22	0.5
sscERI	0.069	0.022	0.00024	0.054	0.067	0.081	0.27
stcSAl	11	5.4	0.013	7.2	10	14	41
sicCAR	0.13	0.062	2e-05	0.083	0.11	0.15	0.62
sicFAR	0.56	0.56	2e-05	0.19	0.32	0.73	3.8
mtcWNe	0.021	0.0089	7.8e-07	0.015	0.02	0.025	0.15
mdcAre	13000	13000	0	4500	9300	18000	290000
licGDe	0.19	0.22	0	0.062	0.12	0.24	2.4
ltcWRB	4e-05	3.1e-05	0	1.8e-05	3.3e-05	5.4e-05	0.00038
sdbHei	5	4.9	0	2	3.1	6.1	42
sdbAre	170	250	0.097	57	96	180	17000
sdbVol	2400	3300	2.1	480	1100	3500	170000
sdbPer	28	26	0.018	13	20	32	420
sdbCoA	0.096	2	0	0	0	0	160
ssbFoF	0.47	0.29	0.00044	0.28	0.39	0.58	4.2
ssbVFR	1.1	0.84	0.0029	0.58	0.86	1.4	21
ssbCCo	0.13	0.054	0.00027	0.086	0.12	0.16	0.37
ssbCor	5.6	2.9	0	4	5.2	6.8	76
ssbSqu	5.7	6	0.001	1.1	2.4	9.9	45
ssbERI	0.087	0.046	1.6e-05	0.059	0.079	0.1	0.44
ssbElo	0.26	0.09	0.00025	0.2	0.26	0.32	0.69
ssbCCM	3.7	3.7	0.0029	1.6	2.5	4.2	57
ssbCCD	1.7	1.1	0.00014	1	1.5	2.1	22
stbOri	9.8	8.8	7.3e-09	2.3	7.6	15	45
stbSAl	7.9	6.7	2.5e-08	2.5	6.4	11	42

	mean	std	min	25%	50%	75%	max
stbCeA	8.8	4.7	0.015	5.3	8.5	12	43
mtbSWR	0.2	0.13	0	0.12	0.21	0.28	0.9
mtbAli	4.7	2.8	0	2.6	4.6	6.5	24
mtbNDi	15	9.4	0	8.6	13	20	82
libNCo	0.62	3.2	0	0	0	0	52
ldbPWL	130	220	0	28	62	140	3200
ltbIBD	6.6	4.2	0	3.5	5.7	8.7	49
ltcBuA	0.1	0.061	0	0.057	0.085	0.13	0.62
mtdDeg	0.6	0.73	0	0	0	1	4
lcdMes	0.029	0.024	0	0.013	0.024	0.038	0.47
linP3W	0.051	0.046	0	0.024	0.043	0.067	0.8
linP4W	0.044	0.037	0	0.02	0.036	0.059	0.46
linPDE	0.039	0.045	0	0.016	0.03	0.05	0.97
lcnClo	1.2e-06	9.1e-07	0	5.5e-07	1e-06	1.7e-06	7e-06
ldsCDL	200	290	0	11	110	260	4100
xcnSCl	0.056	0.074	0	0	0.04	0.08	1
mtdMDi	74	98	0	21	43	87	1500
lddNDe	0.0028	0.0038	0	0.00098	0.0019	0.0035	0.096
linWID	0.0055	0.0065	0	0.002	0.0039	0.0069	0.15
lddRea	56	44	0	25	46	76	480
lddARe	140000	180000	0	42000	89000	170000	3.7e+06
sddAre	20000	31000	0	4400	9700	22000	590000
midRea	22	17	0	10	18	29	220
midAre	49000	66000	0	14000	28000	56000	900000
sdsLen	140	190	0	44	85	160	2900
sdsSPW	8.9	5.3	0	4.8	8.8	13	36
sdsSPH	2.9	3.5	0	0.77	1.6	3.7	38
sdsSPR	0.18	0.17	0	0.063	0.12	0.23	2.2
sdsSPO	0.19	0.11	0	0.11	0.18	0.26	0.91
sdsSWD	2.3	1.3	0	1.3	2.2	3.1	8.7
sdsSHD	1.7	2	0	0.53	1	2.1	20
sssLin	0.073	0.13	0	0	0.011	0.09	1

	mean	std	min	25%	50%	75%	max
sdsAre	23000	40000	0	4900	11000	25000	710000
sisBpM	0.039	0.029	0	0.021	0.034	0.051	0.92
misRea	19	16	0	9	16	25	180
mdsAre	46000	67000	0	12000	25000	53000	970000
ldsMSL	31	43	0	8	17	36	1100
ldsRea	130000	170000	0	40000	80000	160000	2.9e+06
ldkAre	98000	200000	0	5300	21000	96000	2e+06
ldkPer	940	1500	0	98	340	1100	13000
lskCCo	0.12	0.095	0	0.042	0.1	0.18	0.56
lskERI	0.11	0.11	0	0.031	0.08	0.16	0.65
lskCWA	270	410	0	30	100	340	3000
ltkOri	9.9	9.5	0	2.3	6.9	15	45
ltkWNB	0.0028	0.0024	0	0.00085	0.0023	0.0041	0.019
likWBB	0.00049	0.00048	0	0.00014	0.00037	0.0007	0.0053

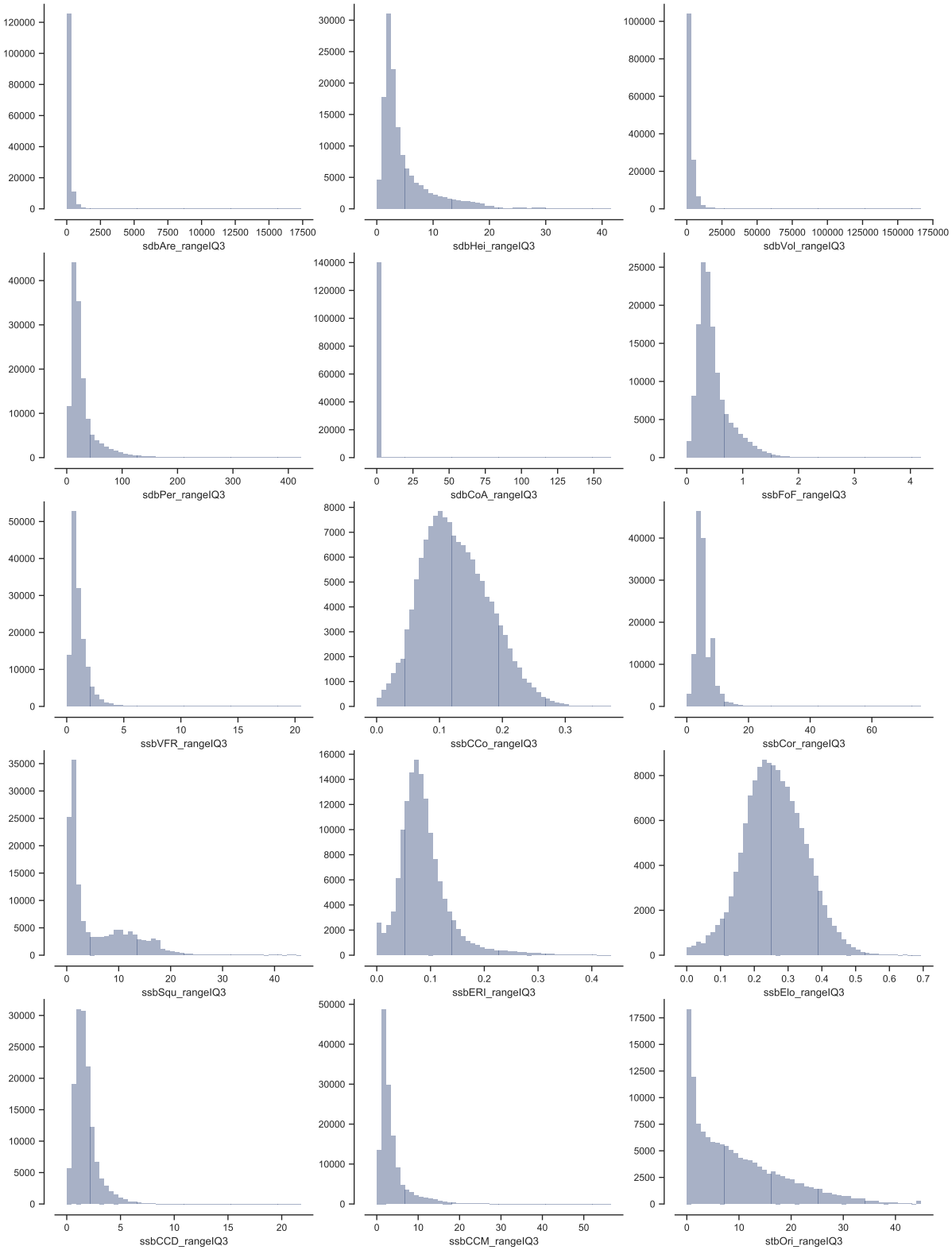


Figure A7.27: Histograms of interquartile range for characters 1-15 are showing the variety of distributions within the measured contextual data.

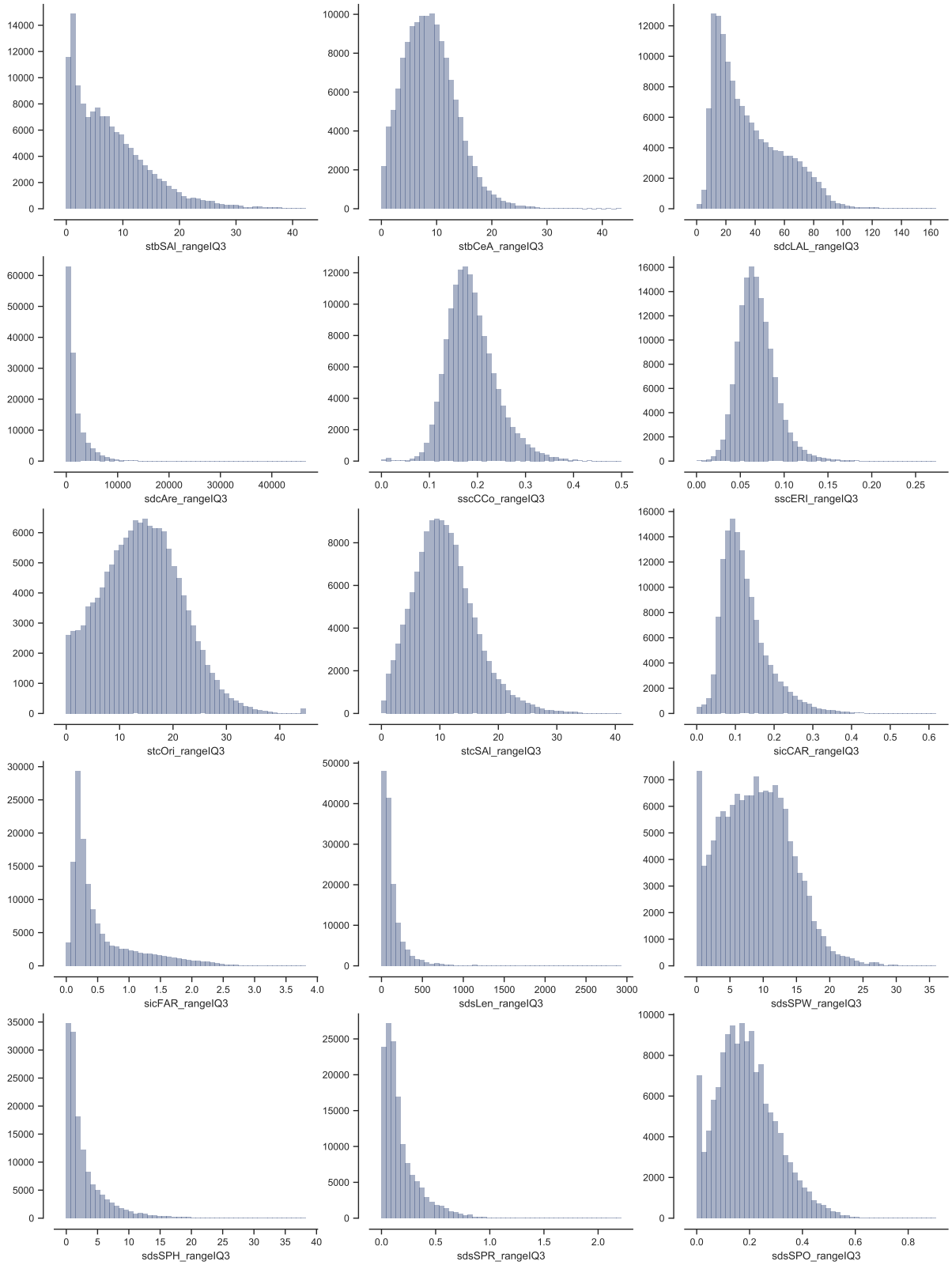


Figure A7.28: Histograms of interquartile range for characters 16-30 are showing the variety of distributions within the measured contextual data.

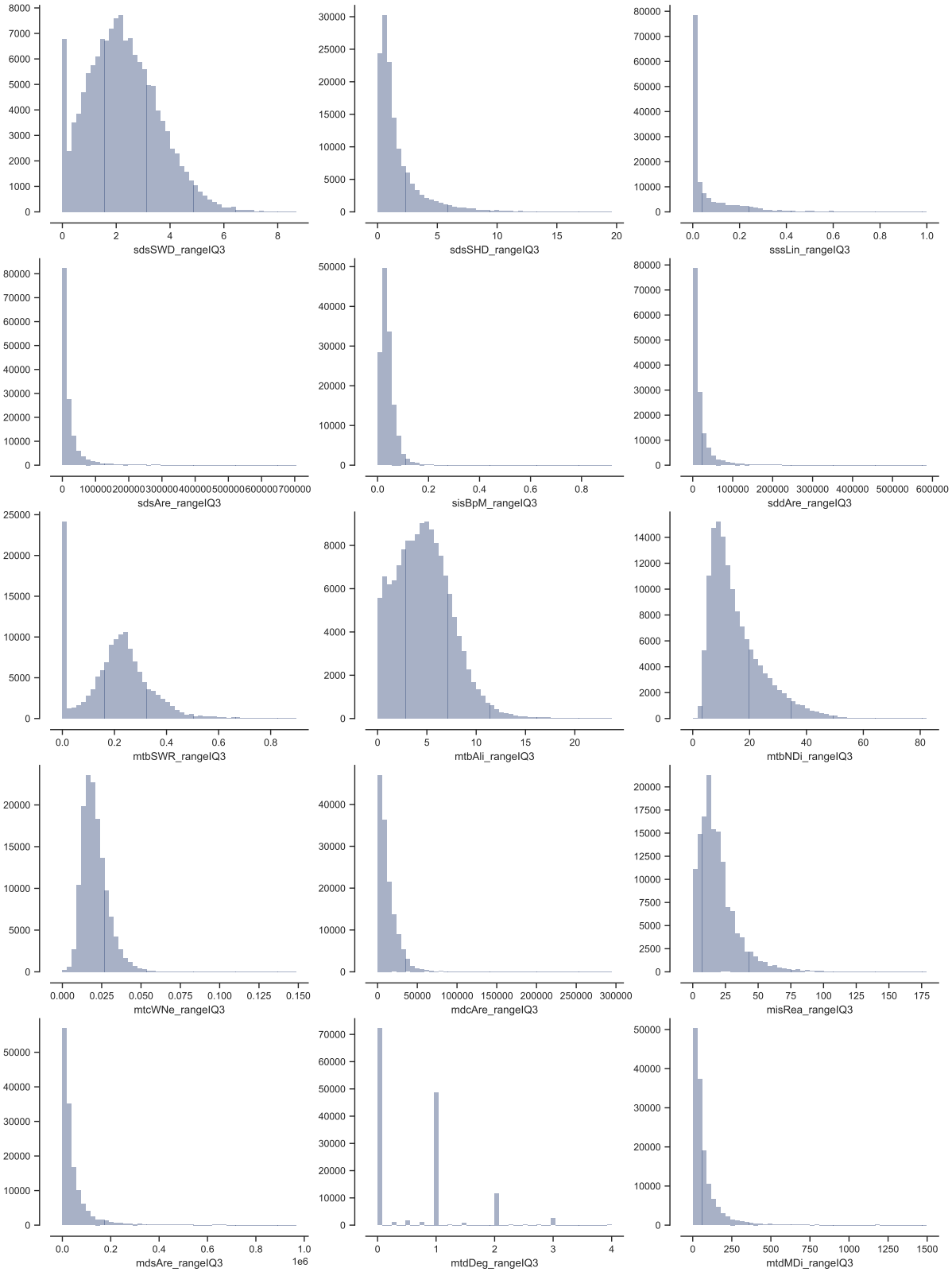


Figure A7.29: Histograms of interquartile range for characters 31-45 are showing the variety of distributions within the measured contextual data.

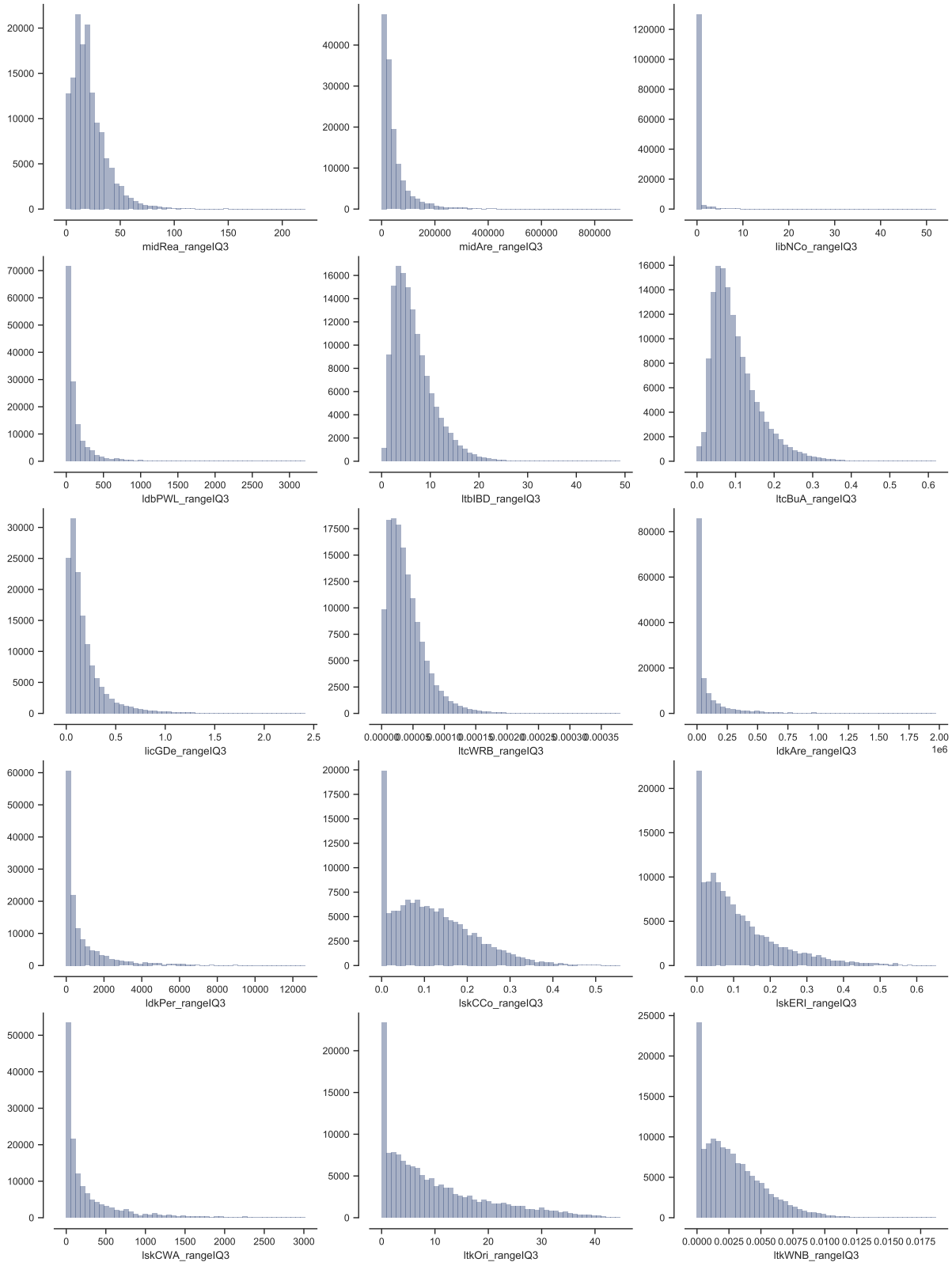


Figure A7.30: Histograms of interquartile range for characters 46-60 are showing the variety of distributions within the measured contextual data.

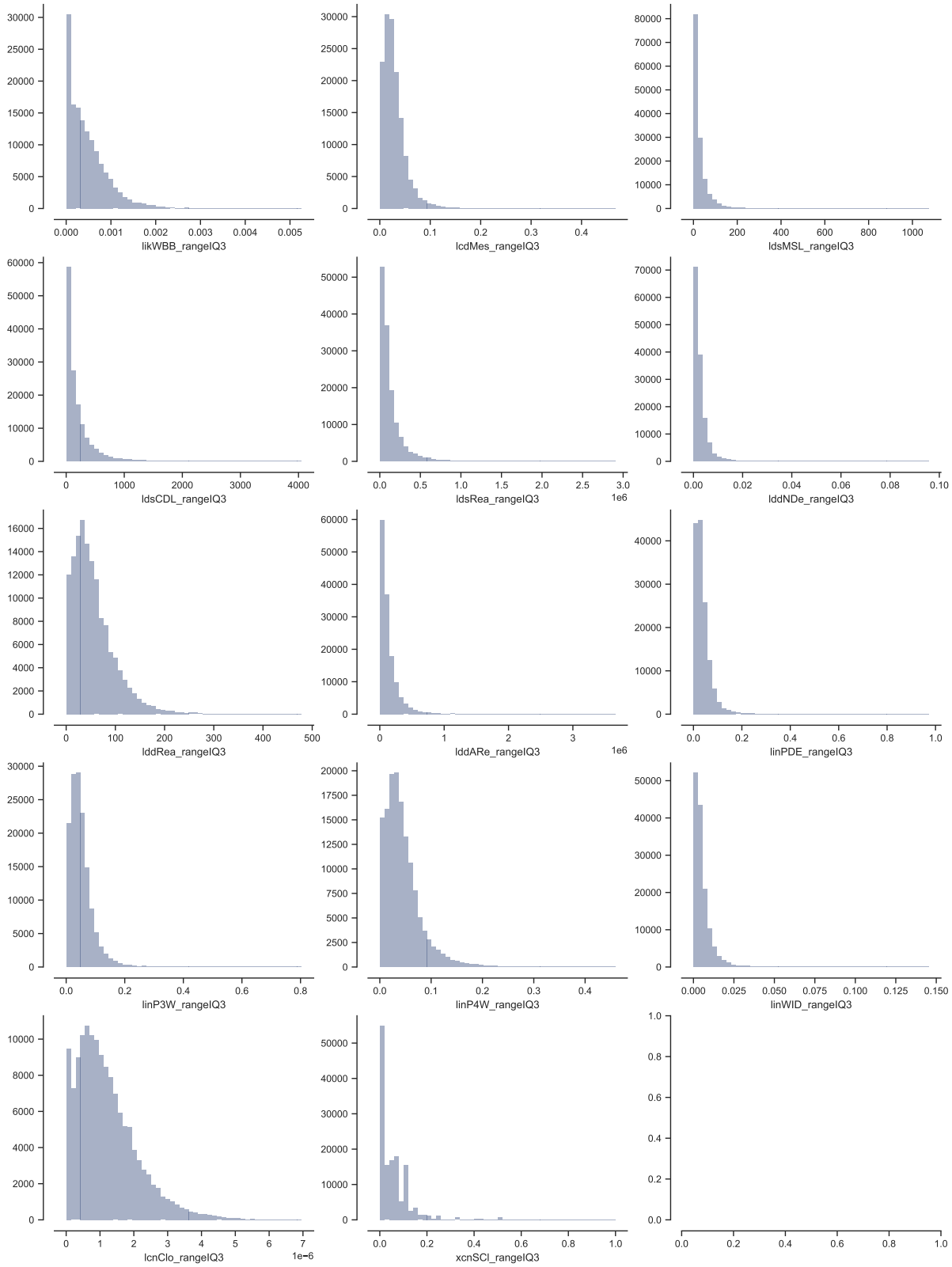


Figure A7.31: Histograms of interquartile range for characters 61-74 are showing the variety of distributions within the measured contextual data.

Interdecile Theil index

Table A7.8: Overview of the contextual morphometric values of interdecile Theil index for the whole case study. Key to character IDs is available in table A7.3.

	mean	std	min	25%	50%	75%	max
stcOri	0.14	0.13	8.3e-08	0.059	0.11	0.19	2.2
sdcLAL	0.046	0.033	2e-08	0.02	0.039	0.066	0.27
sdcAre	0.18	0.12	1.7e-07	0.085	0.15	0.24	1.6
sscCCo	0.027	0.022	1.3e-08	0.014	0.021	0.032	0.41
sscERI	0.00078	0.0005	2.5e-08	0.00047	0.00067	0.00096	0.023
stcSAl	0.32	0.16	1e-06	0.21	0.29	0.4	2.4
sicCAR	0.093	0.072	1e-06	0.043	0.073	0.12	0.78
sicFAR	0.16	0.11	1e-06	0.084	0.14	0.22	1.2
mtcWNe	0.037	0.029	1.2e-07	0.017	0.029	0.049	0.39
mdcAre	0.11	0.082	0	0.048	0.084	0.14	1
licGDe	0.026	0.032	-1.1e-16	0.0073	0.017	0.034	0.75
ltcWRB	0.038	0.036	-2.2e-16	0.014	0.026	0.049	0.47
sdbHei	0.044	0.046	0	0.015	0.027	0.056	0.51
sdbAre	0.12	0.12	2.5e-07	0.04	0.073	0.15	1.5
sdbVol	0.18	0.17	2.5e-07	0.064	0.13	0.25	2.1
sdbPer	0.04	0.042	1.6e-08	0.014	0.024	0.05	0.47
sdbCoA	0.041	0.34	-1.1e-16	0	0	0	4.7
ssbFoF	0.023	0.025	2.8e-08	0.0075	0.013	0.029	0.34
ssbVFR	0.024	0.025	2.2e-07	0.0093	0.016	0.03	0.39
ssbCCo	0.01	0.0082	1.3e-07	0.004	0.0077	0.014	0.097
ssbCor	0.087	0.045	-1.1e-16	0.06	0.079	0.1	0.8
ssbSqu	0.54	0.27	0.00012	0.33	0.53	0.71	2.2
ssbERI	0.0017	0.0021	4.2e-11	0.00063	0.0011	0.0019	0.032
ssbElo	0.022	0.016	5.9e-07	0.01	0.018	0.03	0.19
ssbCCM	0.031	0.036	9.4e-09	0.0087	0.017	0.038	0.41
ssbCCD	0.38	0.28	3.4e-07	0.19	0.3	0.48	2.7
stbOri	0.13	0.18	4.4e-09	0.017	0.075	0.17	3.6
stbSAl	0.4	0.24	6.1e-07	0.23	0.35	0.52	3

	mean	std	min	25%	50%	75%	max
stbCeA	0.48	0.22	6.3e-07	0.33	0.45	0.6	2.4
mtbSWR	0.88	0.78	-1.1e-16	0.3	0.69	1.3	4.3
mtbAli	0.22	0.17	0	0.11	0.18	0.29	2.3
mtbNDi	0.065	0.051	0	0.031	0.052	0.085	0.67
libNCo	0.14	0.43	-1.1e-16	0	0	0	4.6
ldbPWL	0.11	0.11	-2.2e-16	0.03	0.07	0.15	1.2
ltbIBD	0.01	0.012	-2.2e-16	0.0031	0.0066	0.013	0.28
ltcBuA	0.0073	0.011	0	0.001	0.0029	0.0085	0.15
mtdDeg	0.029	0.04	-1.1e-16	0.005	0.0095	0.049	0.27
lcdMes	0.024	0.096	-5.6e-16	0.0021	0.0056	0.015	3.3
linP3W	0.0038	0.044	-6.7e-16	0.00042	0.00095	0.002	3.9
linP4W	0.038	0.16	-6.7e-16	0.0022	0.0058	0.016	3.9
linPDE	0.073	0.22	-4.4e-16	0.0065	0.017	0.046	3.9
lcnClo	0.02	0.041	-6.7e-16	0.0031	0.0088	0.023	1.6
ldsCDL	0.38	0.51	-6.7e-16	0.038	0.17	0.51	4.3
xcnSCl	0.54	0.65	-4.4e-16	0.008	0.31	0.84	4.7
mtdMDi	0.045	0.06	-5.6e-16	0.0099	0.025	0.057	0.96
lddNDe	0.017	0.044	-6.7e-16	0.0022	0.0058	0.015	1.5
linWID	0.017	0.056	-5.6e-16	0.0024	0.0061	0.015	3.9
lddRea	0.029	0.039	-1.1e-16	0.0052	0.014	0.037	0.94
lddARe	0.043	0.06	-5.6e-16	0.0079	0.022	0.053	0.97
sddAre	0.11	0.11	-6.7e-16	0.036	0.078	0.15	1.2
midRea	0.046	0.047	-1.1e-16	0.015	0.032	0.063	0.71
midAre	0.067	0.077	-6.7e-16	0.018	0.042	0.088	1.1
sdsLen	0.083	0.09	-6.7e-16	0.027	0.056	0.11	1.1
sdsSPW	0.022	0.019	-5.6e-16	0.0087	0.018	0.03	0.25
sdsSPH	0.03	0.082	-5.6e-16	0.0026	0.0076	0.025	3
sdsSPR	0.049	0.084	-6.7e-16	0.013	0.028	0.054	3.7
sdsSPO	0.033	0.049	-6.7e-16	0.0092	0.02	0.039	1.5
sdsSWD	0.13	0.15	-6.7e-16	0.034	0.084	0.18	3.7
sdsSHD	0.16	0.19	-5.6e-16	0.038	0.091	0.21	3.7
sssLin	0.009	0.044	-5.6e-16	0	0.00033	0.0043	2

	mean	std	min	25%	50%	75%	max
sdsAre	0.15	0.14	-6.7e-16	0.059	0.11	0.2	1.5
sisBpM	0.061	0.056	-6.7e-16	0.025	0.047	0.08	1.1
misRea	0.049	0.049	-1.1e-16	0.016	0.034	0.067	0.6
mdsAre	0.079	0.089	-5.6e-16	0.021	0.049	0.1	1
ldsMSL	0.012	0.025	-5.6e-16	0.0012	0.0039	0.011	0.35
ldsRea	0.042	0.061	-5.6e-16	0.0072	0.021	0.051	0.96
ldkAre	0.22	0.25	-6.7e-16	0.04	0.13	0.32	1.9
ldkPer	0.1	0.13	-6.7e-16	0.013	0.053	0.15	1
lskCCo	0.023	0.028	-6.7e-16	0.0048	0.014	0.032	0.31
lskERI	0.0063	0.011	-6.7e-16	0.00058	0.0022	0.0072	0.12
lskCWA	0.16	0.18	-6.7e-16	0.034	0.1	0.23	1.4
ltkOri	0.14	0.18	-6.7e-16	0.018	0.078	0.2	2.6
ltkWNB	0.041	0.047	-6.7e-16	0.0092	0.026	0.058	0.54
likWBB	0.081	0.082	-6.7e-16	0.019	0.057	0.12	0.84

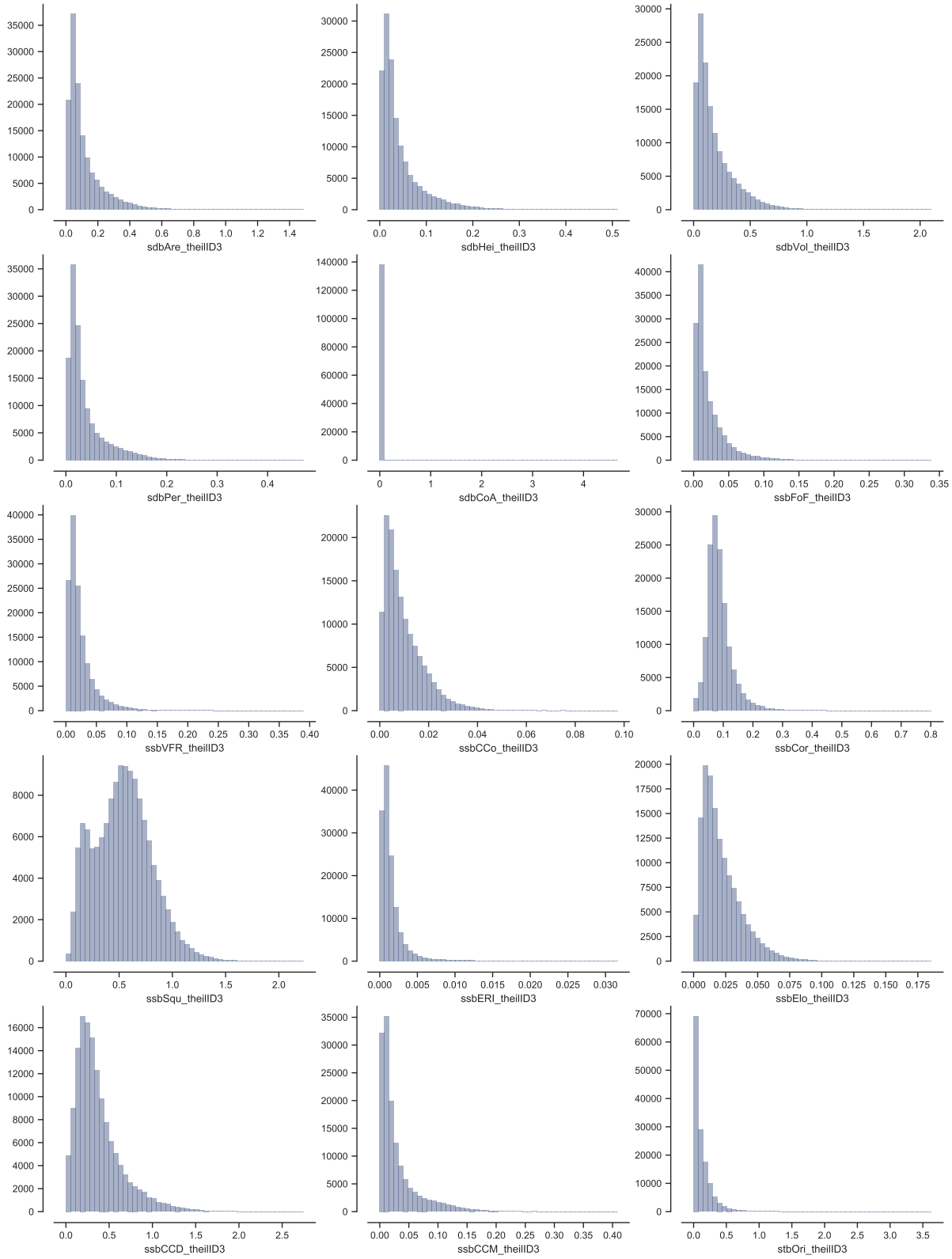


Figure A7.32: Histograms of interdecile Theil index for characters 1-15 are showing the variety of distributions within the measured contextual data.

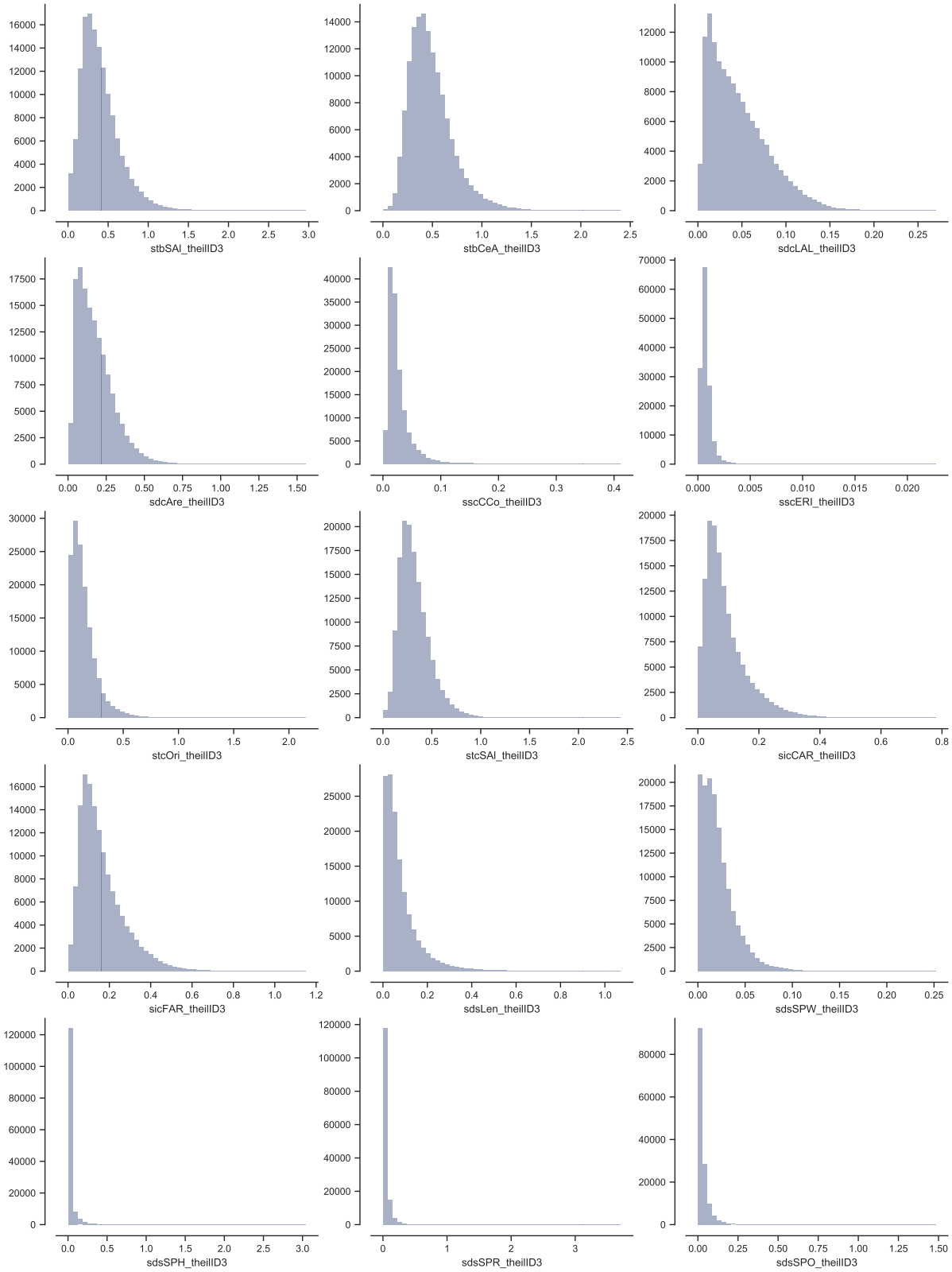


Figure A7.33: Histograms of interdecile Theil index for characters 16-30 are showing the variety of distributions within the measured contextual data.

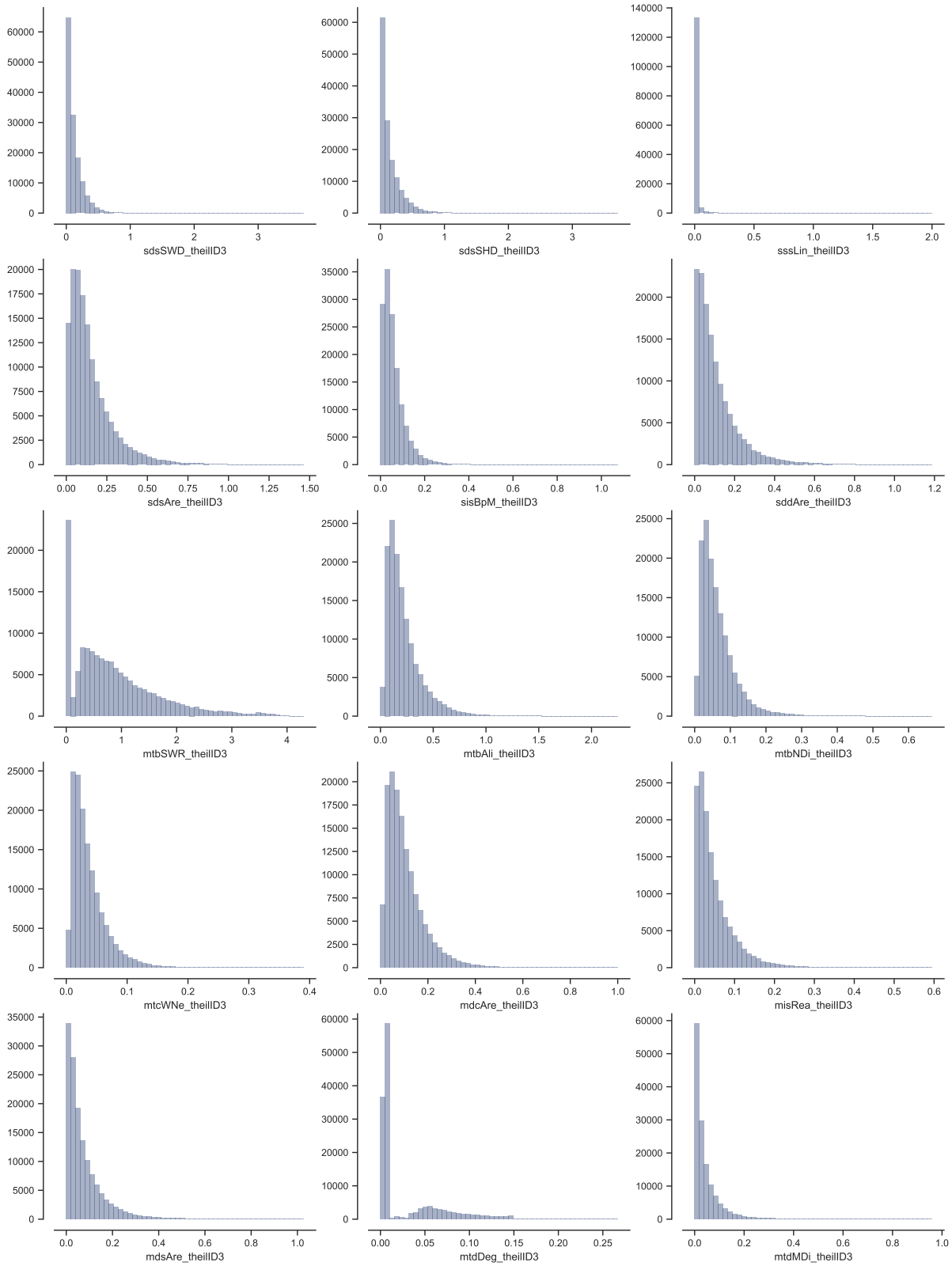


Figure A7.34: Histograms of interdecile Theil index for characters 31-45 are showing the variety of distributions within the measured contextual data.

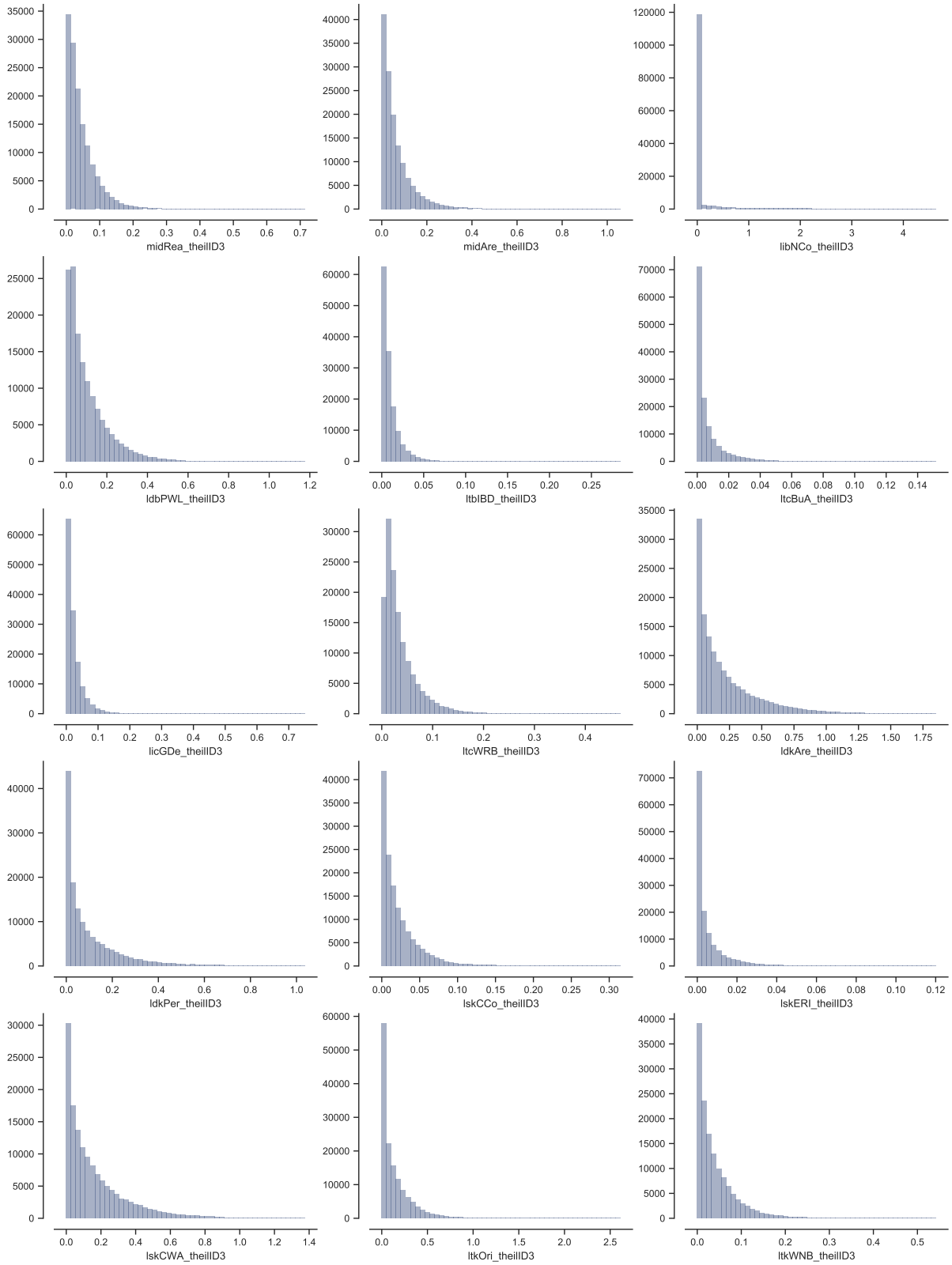


Figure A7.35: Histograms of interdecile Theil index for characters 46-60 are showing the variety of distributions within the measured contextual data.

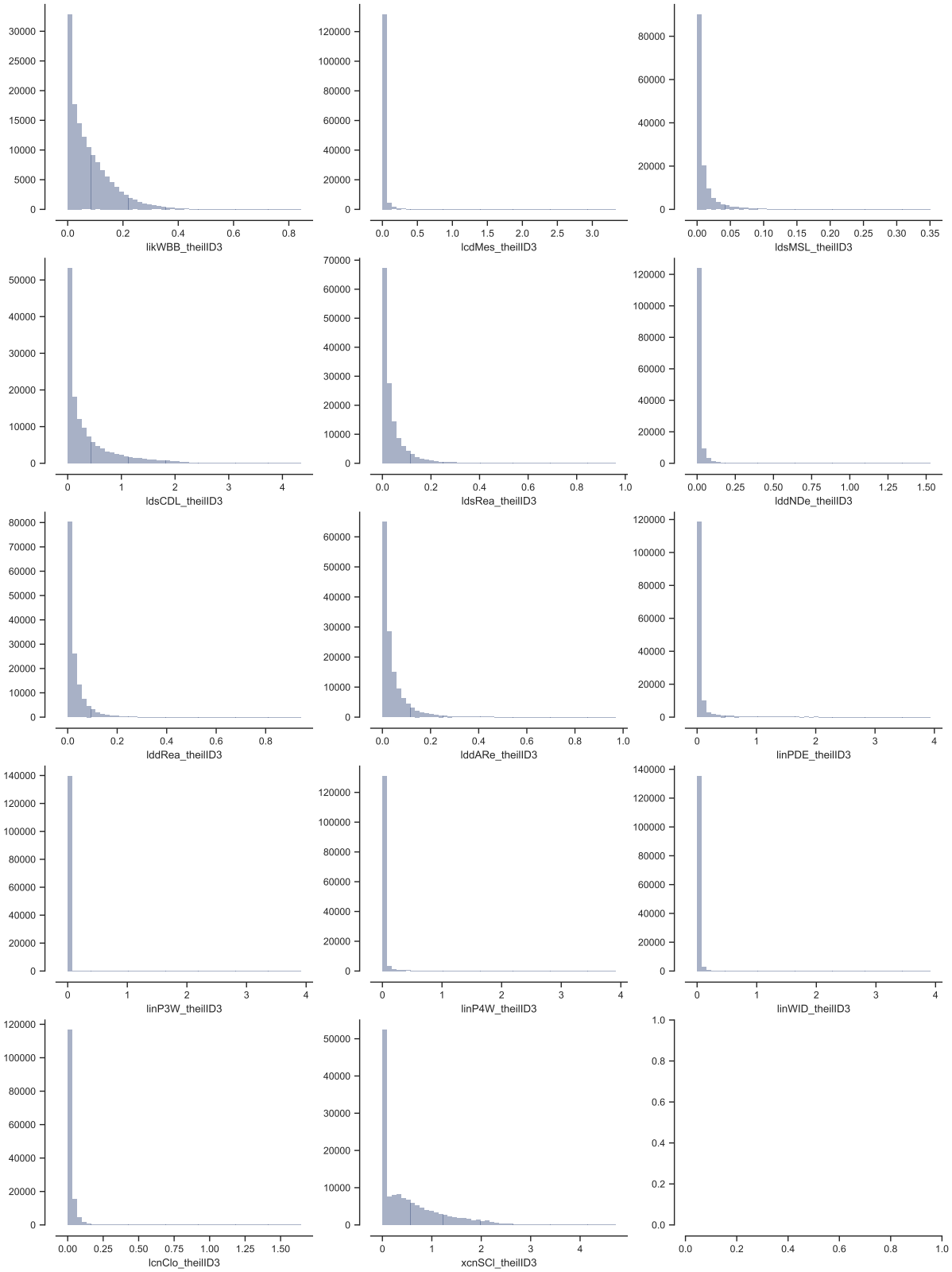


Figure A7.36: Histograms of interdecile Theil index for characters 61-74 are showing the variety of distributions within the measured contextual data.

Simpson index

Table A7.9: Overview of the contextual morphometric values of Simpson index for the whole case study. Key to character IDs is available in table A7.3.

	mean	std	min	25%	50%	75%	max
sdcLAL	0.57	0.22	0.15	0.39	0.54	0.74	1
sdcAre	0.67	0.22	0.15	0.49	0.67	0.85	1
stcSAl	0.5	0.16	0.14	0.38	0.47	0.6	1
sicCAR	0.58	0.21	0.13	0.42	0.56	0.73	1
sicFAR	0.7	0.26	0.13	0.5	0.75	0.95	1
mdcAre	0.62	0.24	0.12	0.42	0.6	0.85	1
licGDe	0.83	0.22	0.12	0.62	1	1	1
ltcWRB	0.67	0.26	0.1	0.44	0.61	1	1
sdbHei	0.68	0.25	0.13	0.46	0.67	0.94	1
sdbAre	0.71	0.22	0.16	0.51	0.73	0.91	1
sdbVol	0.73	0.23	0.17	0.51	0.78	0.96	1
sdbPer	0.61	0.21	0.17	0.44	0.6	0.79	1
sdbCoA	0.99	0.048	0.43	1	1	1	1
ssbFoF	0.52	0.18	0.11	0.4	0.49	0.64	1
ssbVFR	0.59	0.21	0.14	0.41	0.56	0.78	1
ssbCor	0.53	0.14	0.16	0.43	0.52	0.62	1
ssbSqu	0.63	0.18	0.16	0.5	0.64	0.77	1
ssbCCM	0.62	0.23	0.14	0.44	0.6	0.83	1
ssbCCD	0.53	0.15	0.18	0.42	0.51	0.62	1
stbSAl	0.59	0.21	0.13	0.42	0.55	0.76	1
stbCeA	0.53	0.16	0.15	0.4	0.5	0.63	1
mtbAli	0.54	0.23	0.12	0.35	0.48	0.7	1
mtbNDi	0.53	0.21	0.13	0.37	0.49	0.66	1
libNCo	0.92	0.16	0.2	0.94	1	1	1
ldbPWL	0.74	0.23	0.15	0.53	0.78	1	1
linPDE	0.75	0.23	0.18	0.53	0.77	1	1
ldsCDL	0.75	0.23	0.18	0.53	0.77	1	1
xcnSCl	0.66	0.23	0.2	0.47	0.6	0.9	1

	mean	std	min	25%	50%	75%	max
mtdMDi	0.74	0.22	0.22	0.53	0.76	1	1
lddNDe	0.7	0.23	0.19	0.51	0.68	0.97	1
linWID	0.7	0.23	0.17	0.5	0.66	0.96	1
lddARe	0.76	0.22	0.21	0.54	0.8	1	1
sddAre	0.78	0.22	0.21	0.56	0.83	1	1
midRea	0.65	0.22	0.17	0.49	0.59	0.85	1
midAre	0.77	0.22	0.17	0.56	0.82	1	1
sdsLen	0.73	0.22	0.2	0.53	0.72	1	1
sdsSPH	0.75	0.25	0.15	0.52	0.82	1	1
sdsSPR	0.74	0.25	0.15	0.51	0.78	1	1
sdsSHD	0.69	0.25	0.15	0.48	0.66	1	1
sdsAre	0.78	0.22	0.2	0.57	0.83	1	1
sisBpM	0.6	0.21	0.2	0.45	0.54	0.74	1
misRea	0.65	0.22	0.16	0.49	0.59	0.85	1
mdsAre	0.78	0.22	0.17	0.56	0.83	1	1
ldsMSL	0.79	0.22	0.2	0.58	0.87	1	1
ldsRea	0.77	0.22	0.17	0.56	0.82	1	1
ldkAre	0.82	0.21	0	0.61	0.94	1	1
ldkPer	0.78	0.21	0	0.57	0.82	1	1
lskCWA	0.77	0.21	0.25	0.56	0.8	1	1
likWBB	0.67	0.24	0.16	0.5	0.61	0.95	1
stcOri	0.3	0.13	0.17	0.21	0.25	0.34	1
sscCCo	0.2	0.05	0.14	0.18	0.19	0.22	1
sscERI	0.2	0.041	0.13	0.18	0.2	0.22	1
mtcWNe	0.27	0.093	0.13	0.21	0.25	0.3	1
ssbCCo	0.24	0.081	0.14	0.19	0.22	0.28	1
ssbERI	0.32	0.11	0.14	0.25	0.29	0.36	1
ssbElo	0.24	0.075	0.17	0.2	0.23	0.27	1
stbOri	0.47	0.22	0.17	0.3	0.41	0.59	1
mtbSWR	0.49	0.17	0.25	0.36	0.44	0.58	1
ltbIBD	0.45	0.19	0.15	0.31	0.4	0.52	1
ltcBuA	0.56	0.21	0.17	0.41	0.51	0.67	1

	mean	std	min	25%	50%	75%	max
mtdDeg	0.63	0.18	0.33	0.5	0.59	0.76	1
lcdMes	0.55	0.19	0.16	0.41	0.51	0.66	1
linP3W	0.54	0.19	0.13	0.4	0.5	0.63	1
linP4W	0.59	0.2	0.16	0.45	0.54	0.72	1
lcnClo	0.52	0.2	0.16	0.37	0.48	0.62	1
lddRea	0.47	0.18	0.15	0.34	0.43	0.55	1
sdsSPW	0.38	0.15	0.15	0.27	0.33	0.43	1
sdsSPO	0.38	0.16	0.15	0.27	0.34	0.43	1
sdsSWD	0.37	0.15	0.17	0.27	0.33	0.42	1
sssLin	0.74	0.22	0.2	0.54	0.75	1	1
lskCCo	0.47	0.19	0.15	0.33	0.42	0.54	1
lskERI	0.49	0.19	0.16	0.36	0.45	0.57	1
ltkOri	0.53	0.21	0.17	0.37	0.48	0.63	1
ltkWNB	0.53	0.22	0.15	0.36	0.47	0.64	1

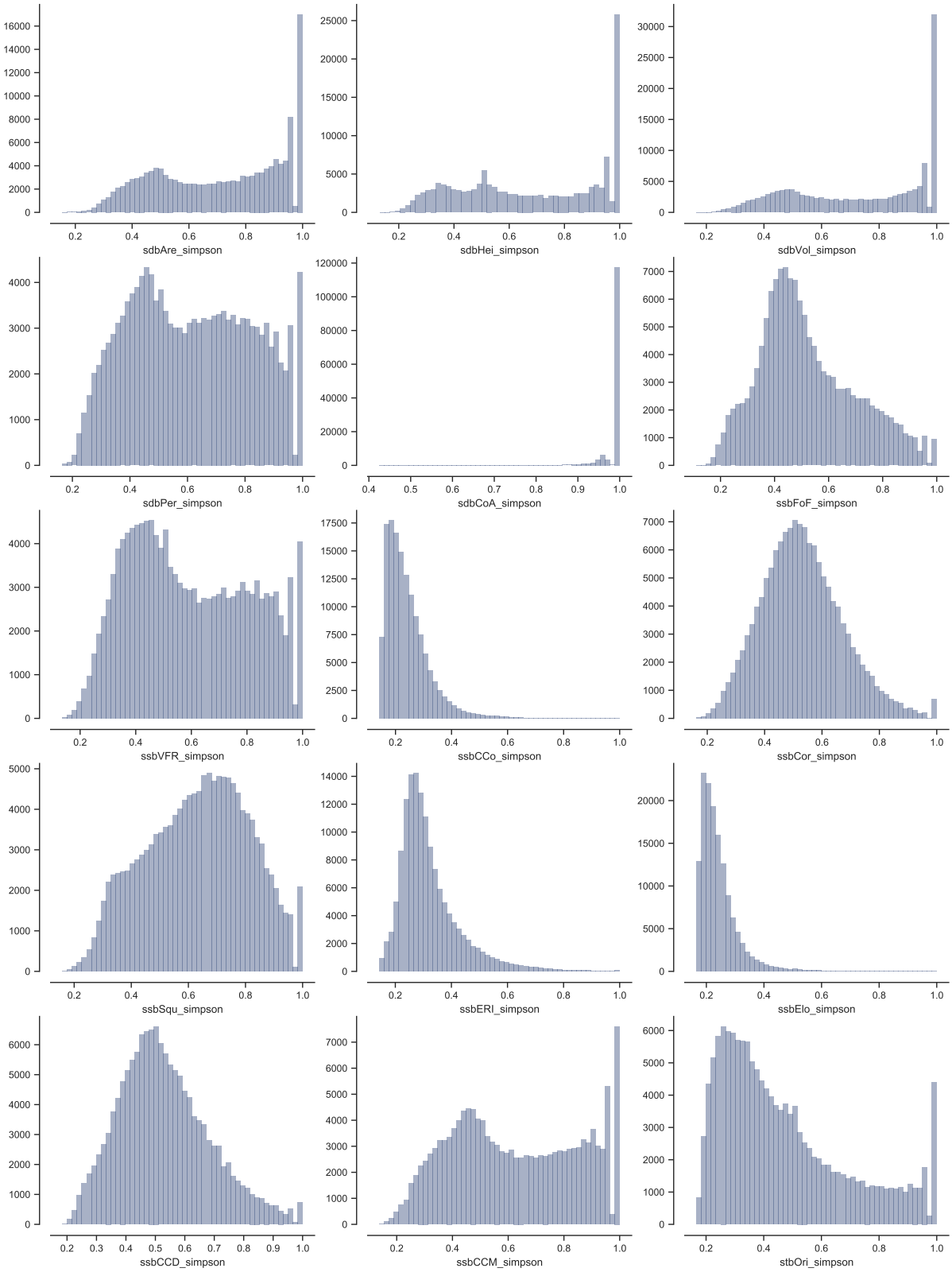


Figure A7.37: Histograms of Simpson index for characters 1-15 are showing the variety of distributions within the measured contextual data.

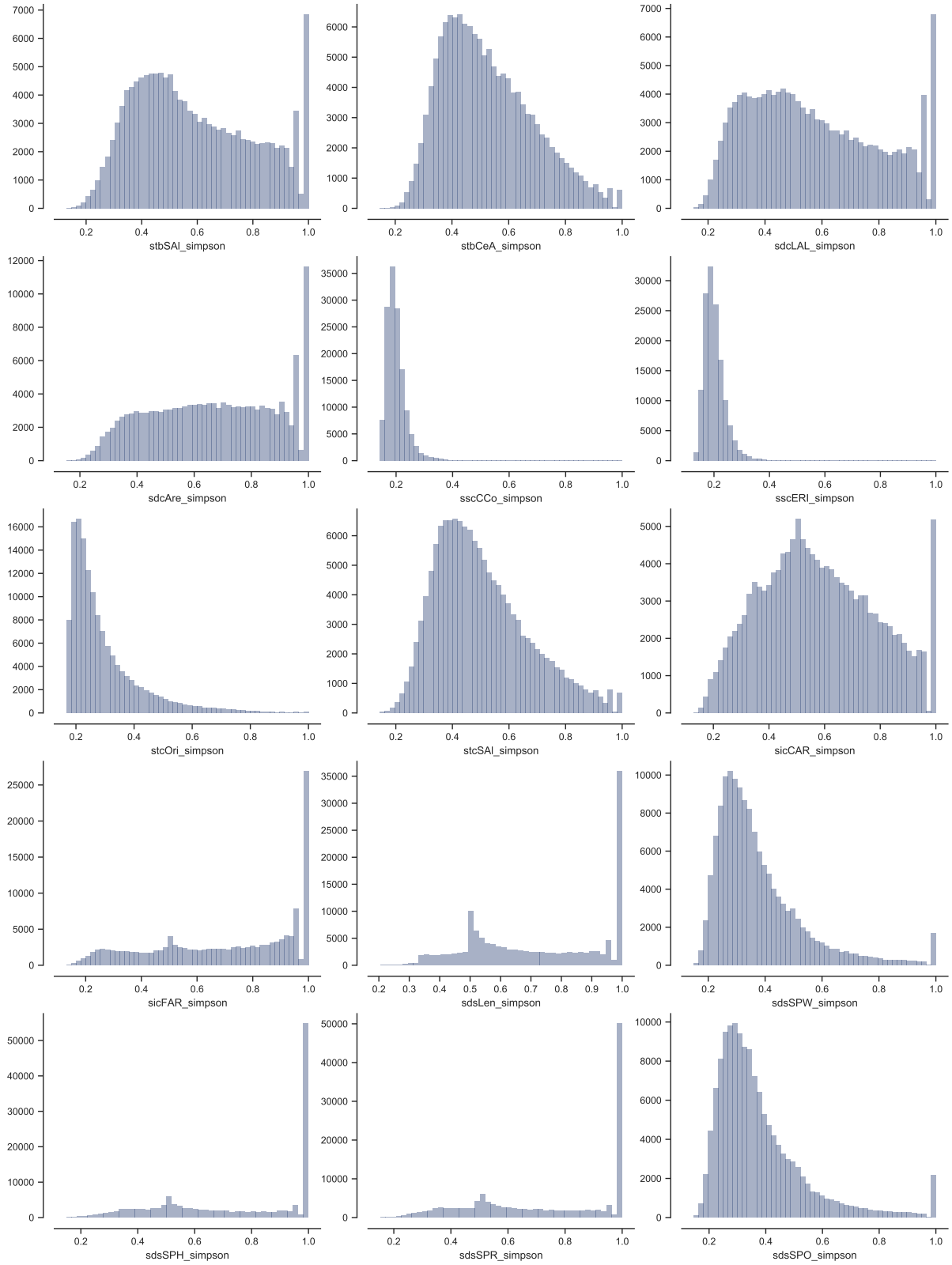


Figure A7.38: Histograms of Simpson index for characters 16-30 are showing the variety of distributions within the measured contextual data.

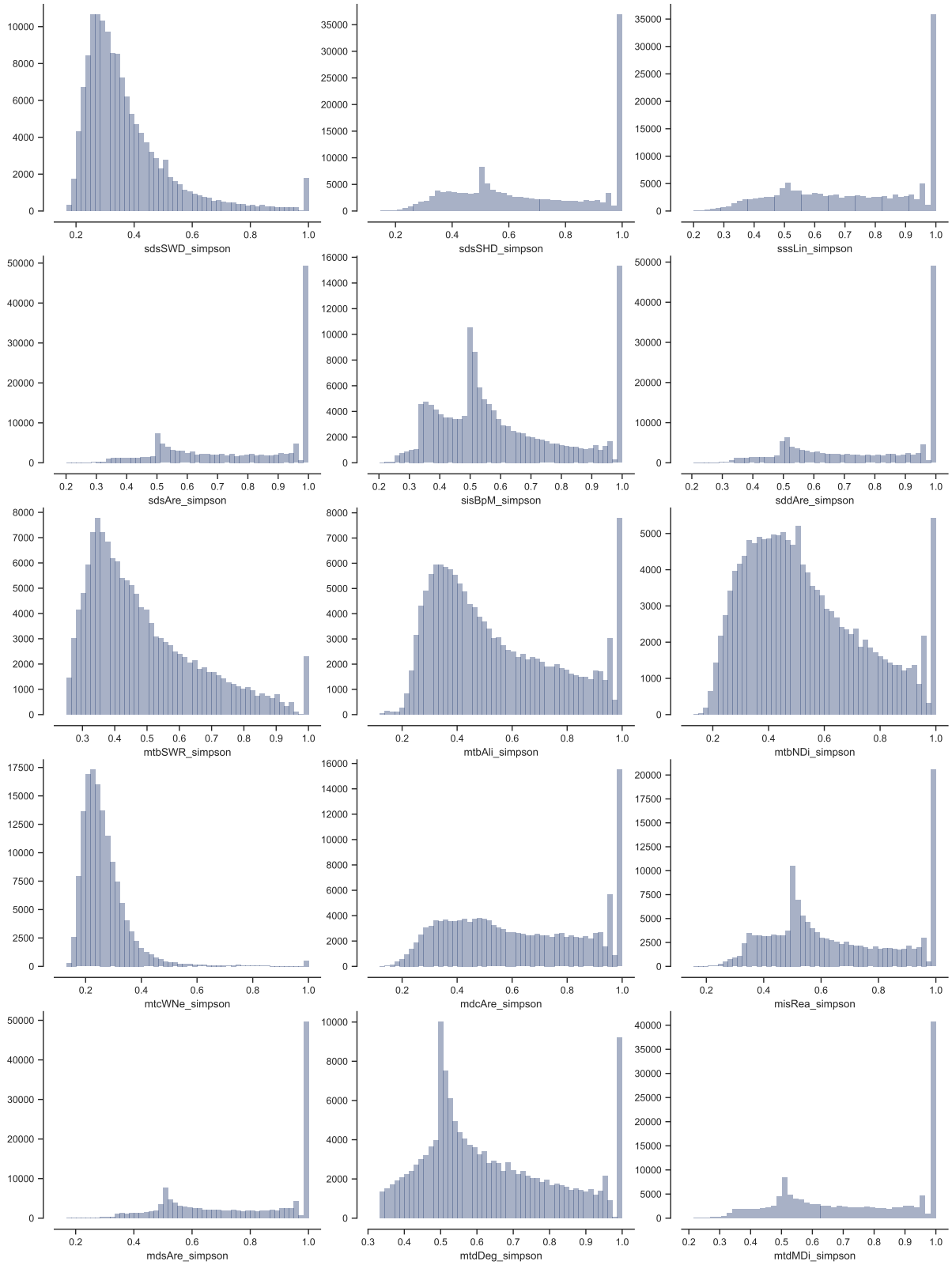


Figure A7.39: Histograms of Simpson index for characters 31-45 are showing the variety of distributions within the measured contextual data.

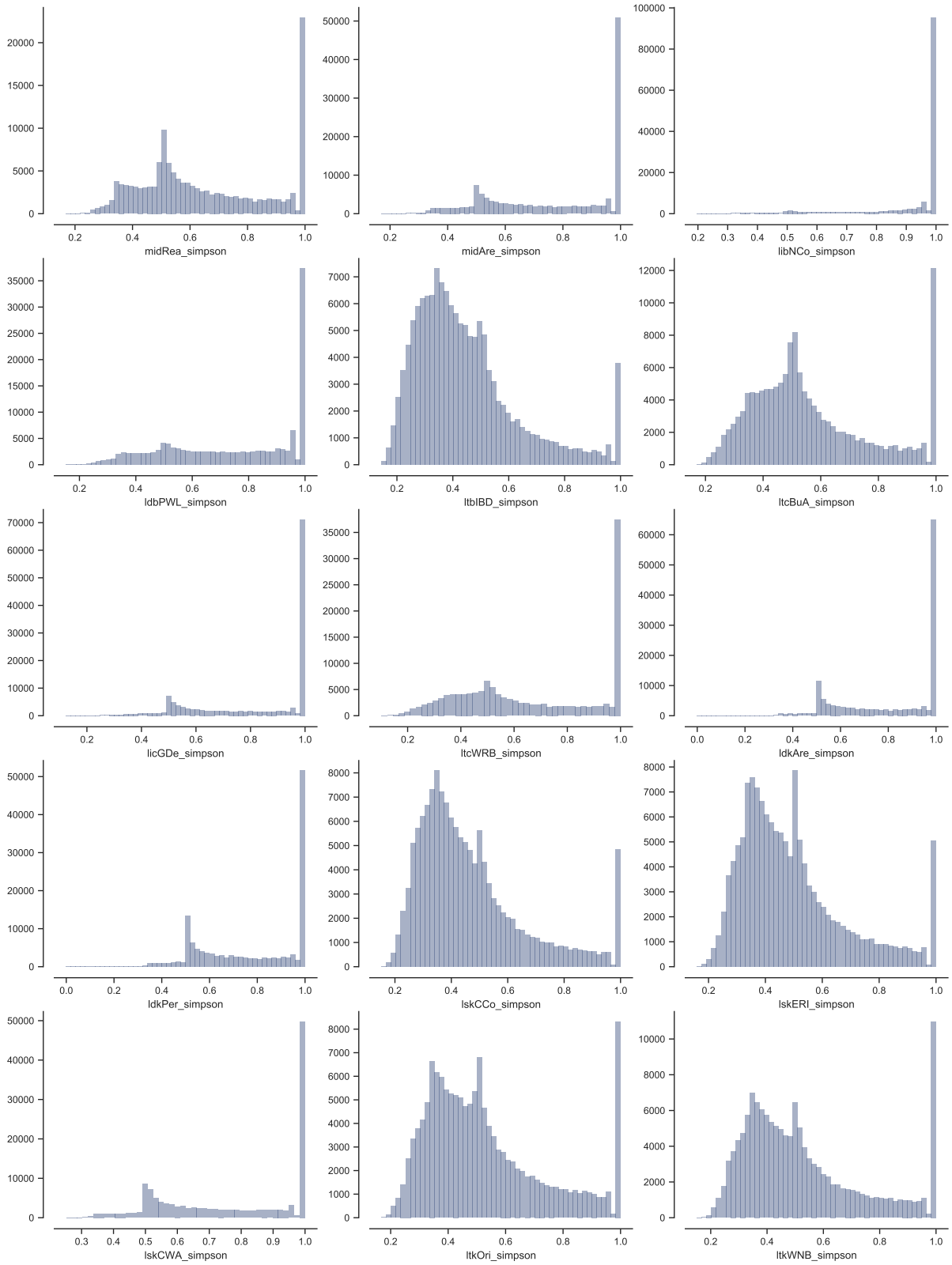


Figure A7.40: Histograms of Simpson index for characters 46-60 are showing the variety of distributions within the measured contextual data.

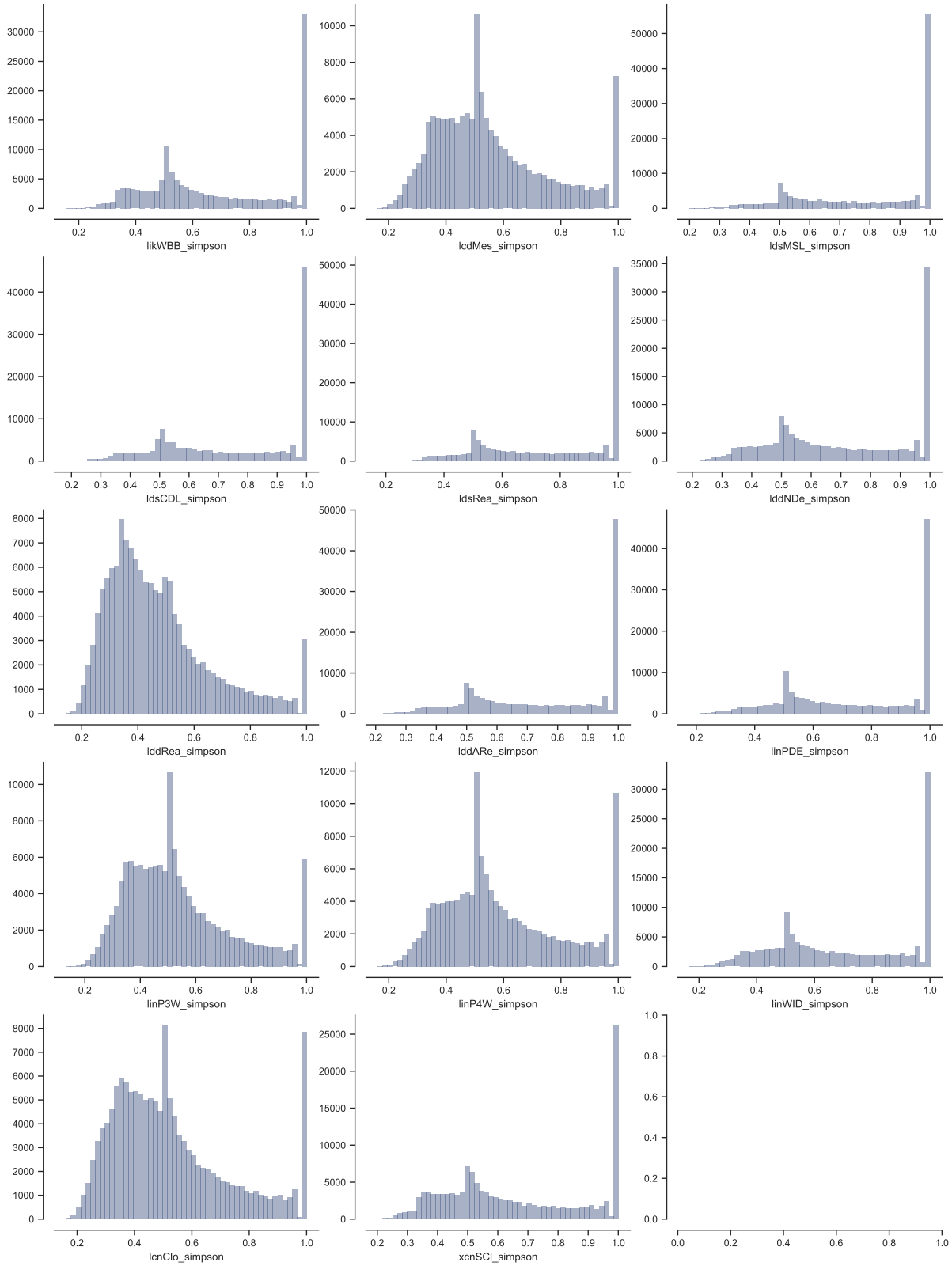


Figure A7.41: Histograms of Simpson index for characters 61-74 are showing the variety of distributions within the measured contextual data.

7.8 Correlation matrix of contextual characters

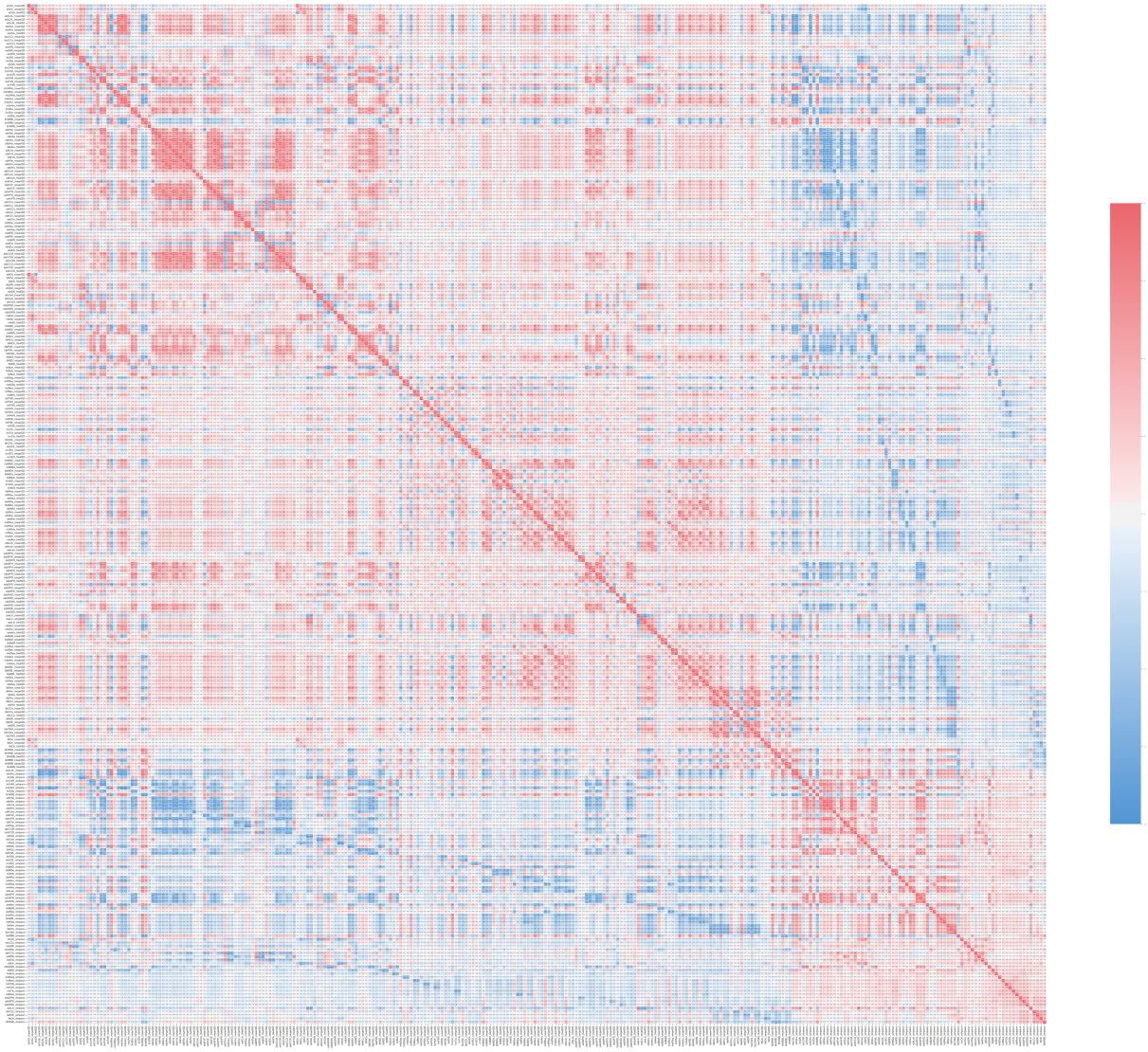


Figure A7.42: Correlation matrix of contextual characters. Please refer to a PDF version for a better clarity.

7.9 Structure of clusters of sampled and complete clustering

Supplementary material for section 7.2.3.2.2.1.

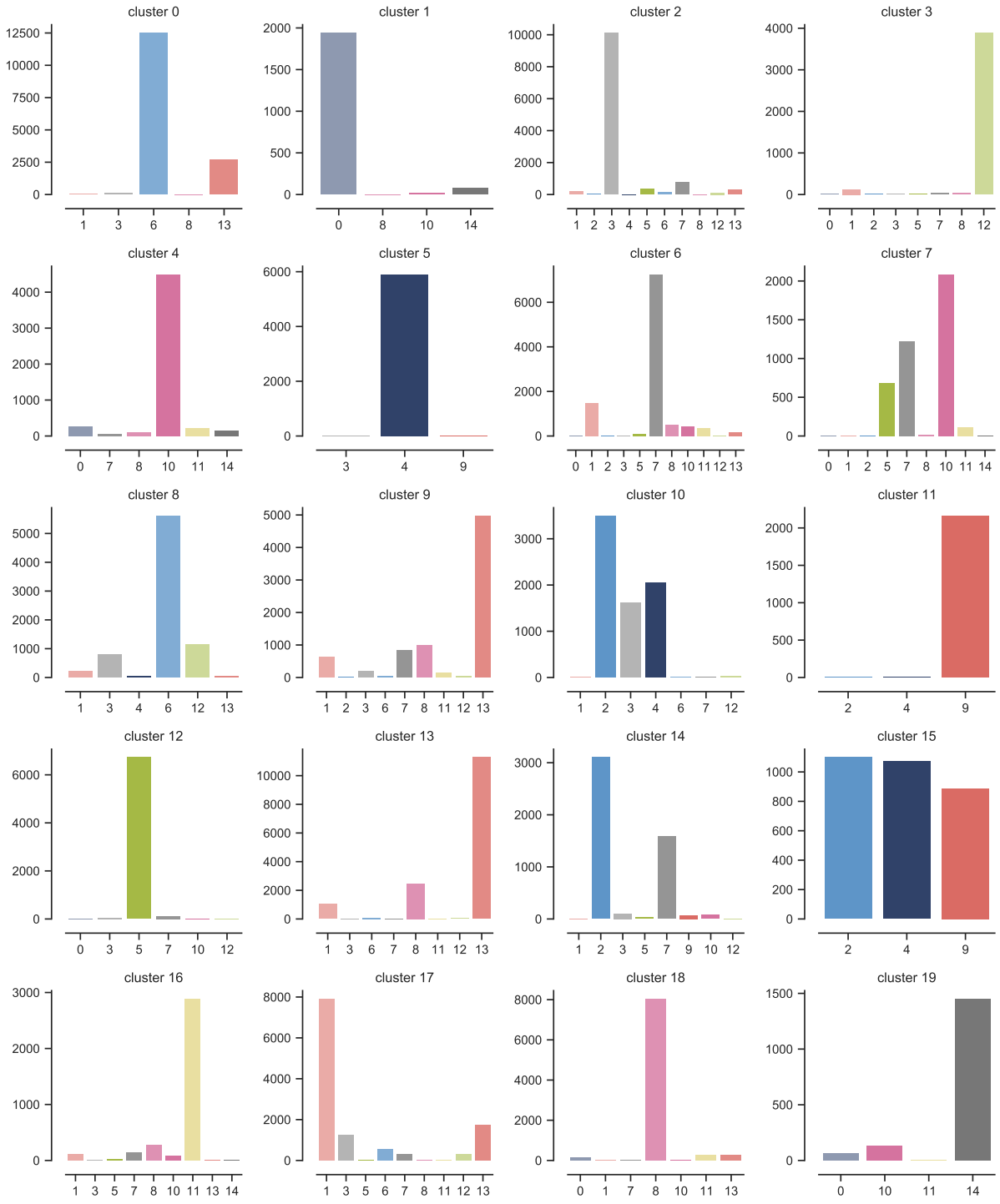


Figure A7.43: Composition of clusters in relation to sampled clustering. Shows number of features labeled as studied cluster and their labels in the other clustering variant.

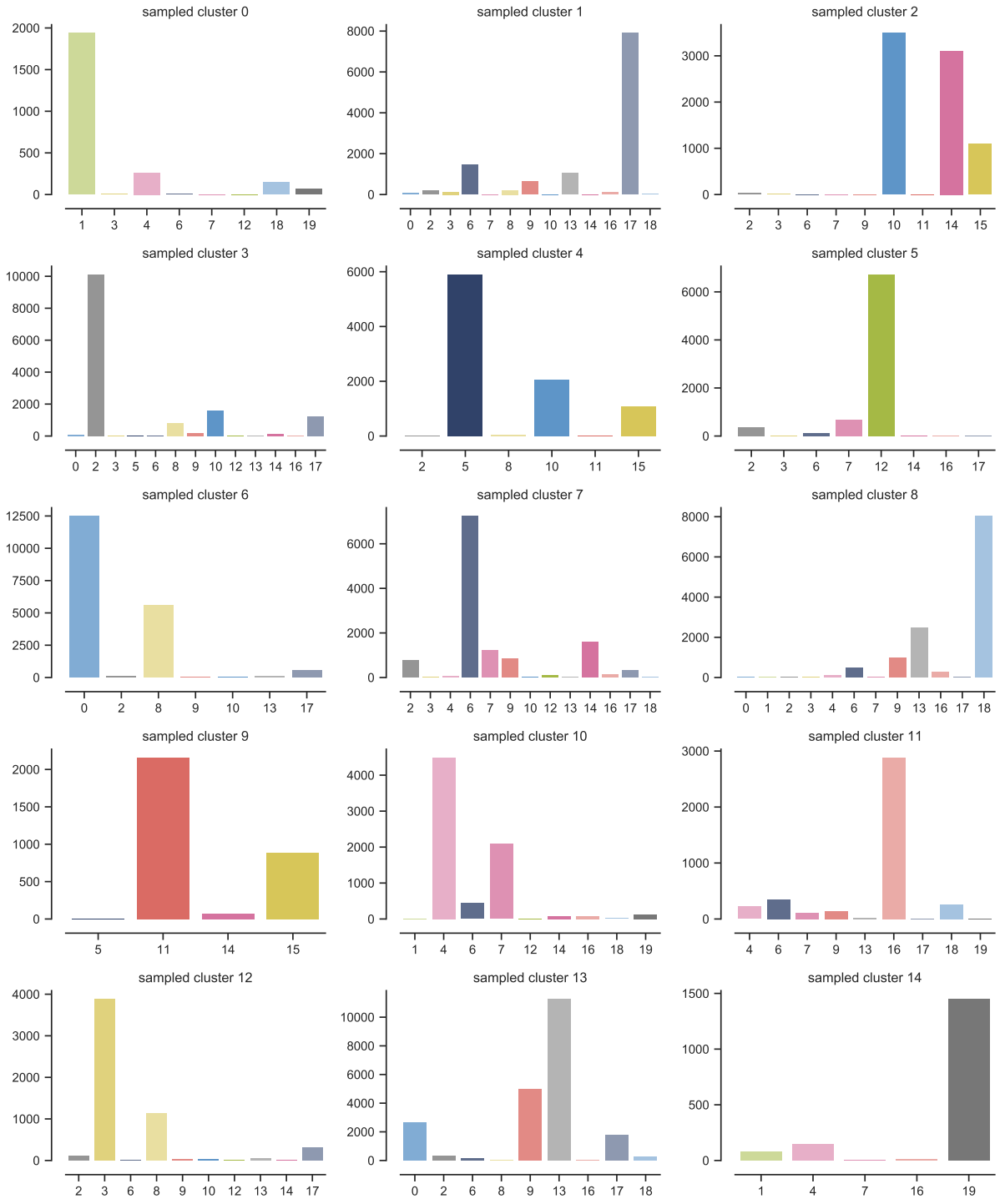


Figure A7.44: Composition of sampled clusters in relation to original clustering. Shows number of features labeled as studied cluster and their labels in the other clustering variant.

Appendix A8: Supplementary material for chapter 8

A8.1 Individual branches in Prague

Spatial distribution of individual branches of the dendrogram. Selection of branches was presented in the section 8.3.1. Complete overview of branches is presented below. For a better clarity please refer to the electronic version of the document.

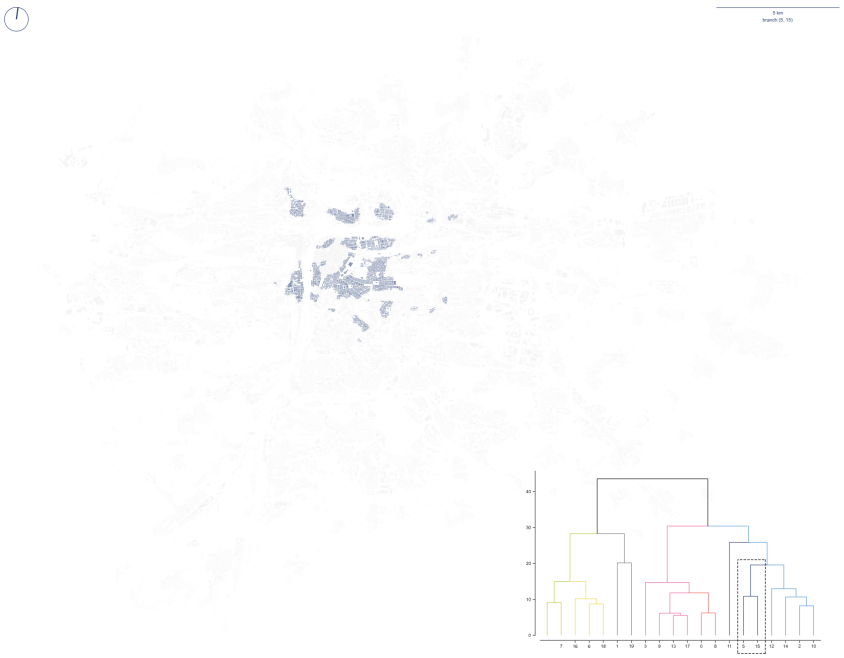


Figure A8.1: Spatial distribution of clusters within a branch 0

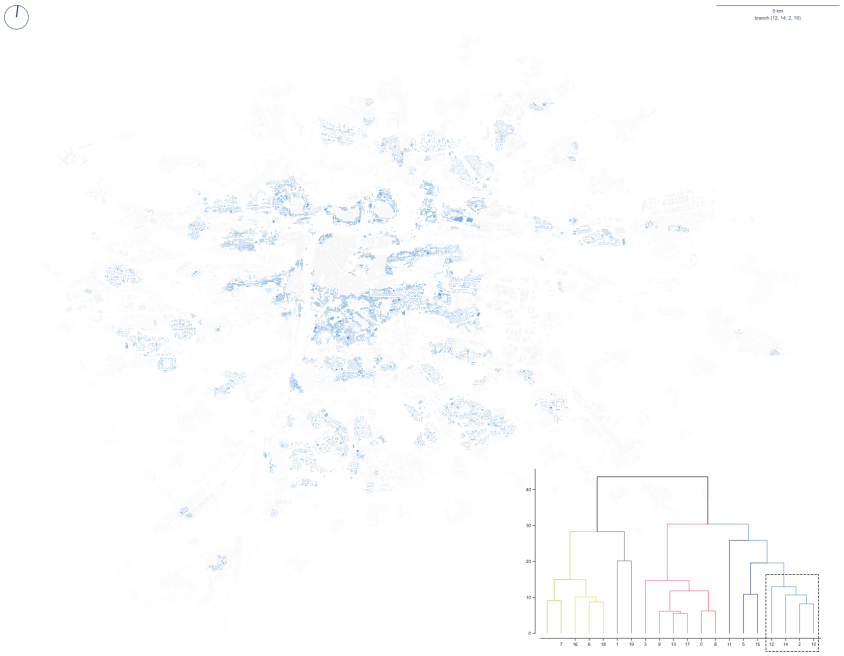


Figure A8.2: Spatial distribution of clusters within a branch 1

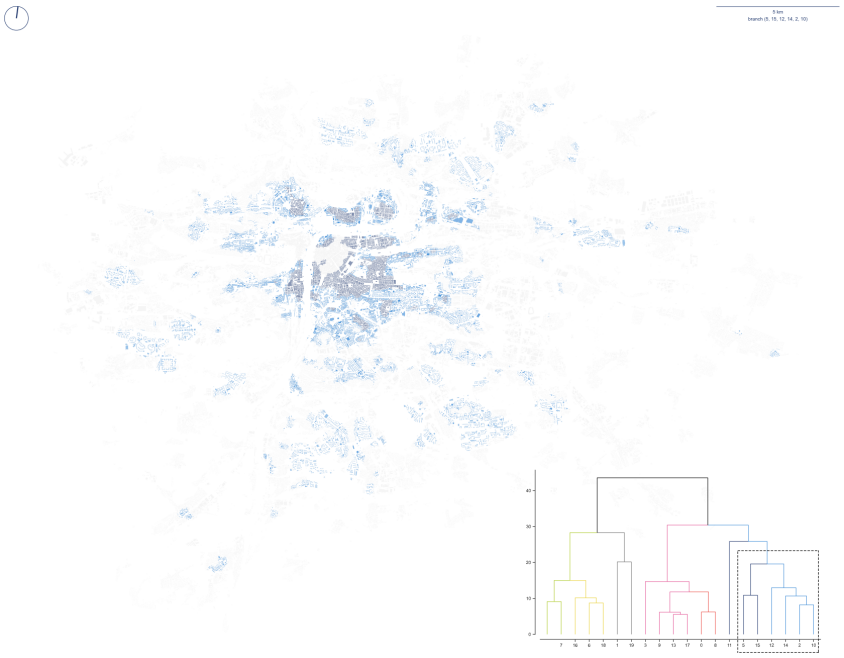


Figure A8.3: Spatial distribution of clusters within a branch 2

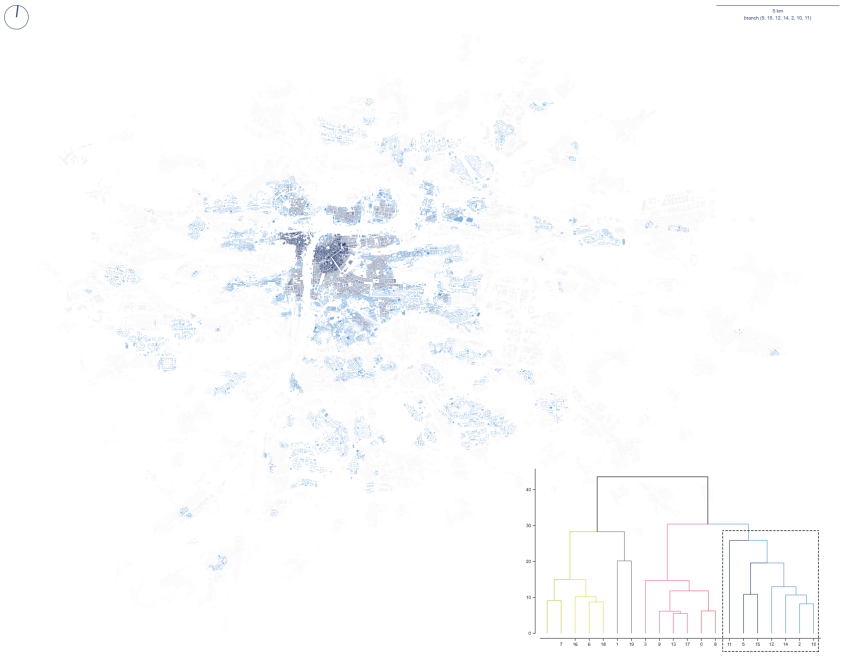


Figure A8.4: Spatial distribution of clusters within a branch 3

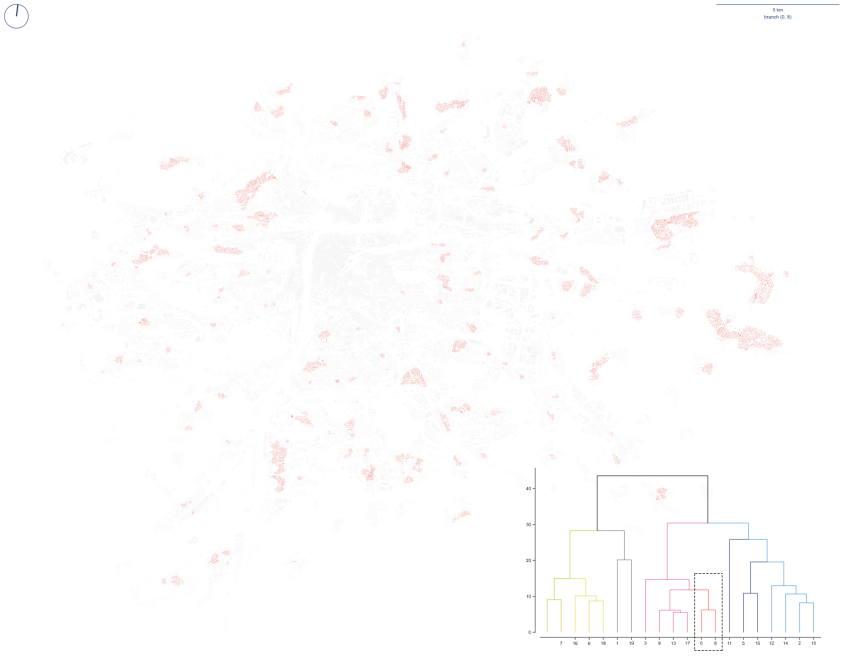


Figure A8.5: Spatial distribution of clusters within a branch 4

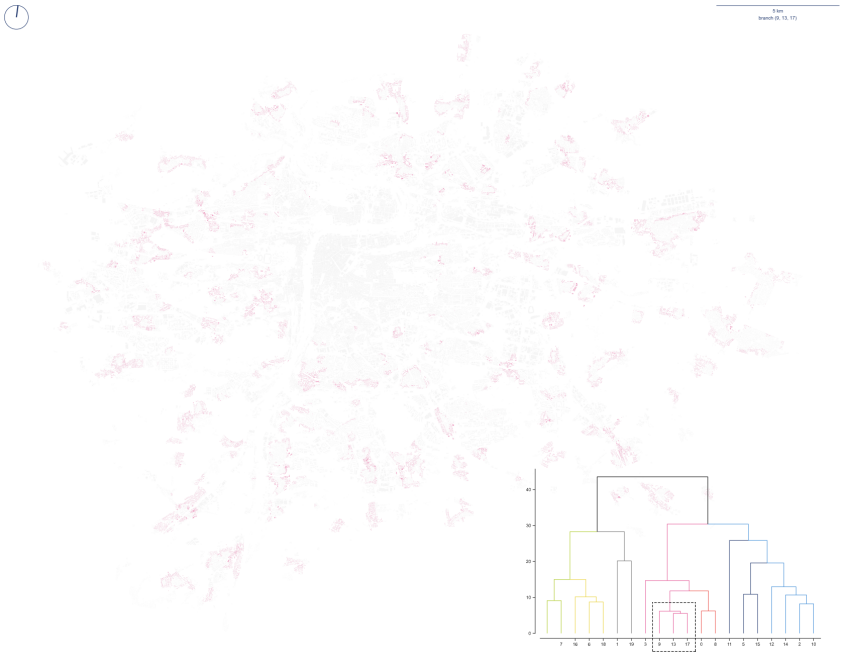


Figure A8.6: Spatial distribution of clusters within a branch 5

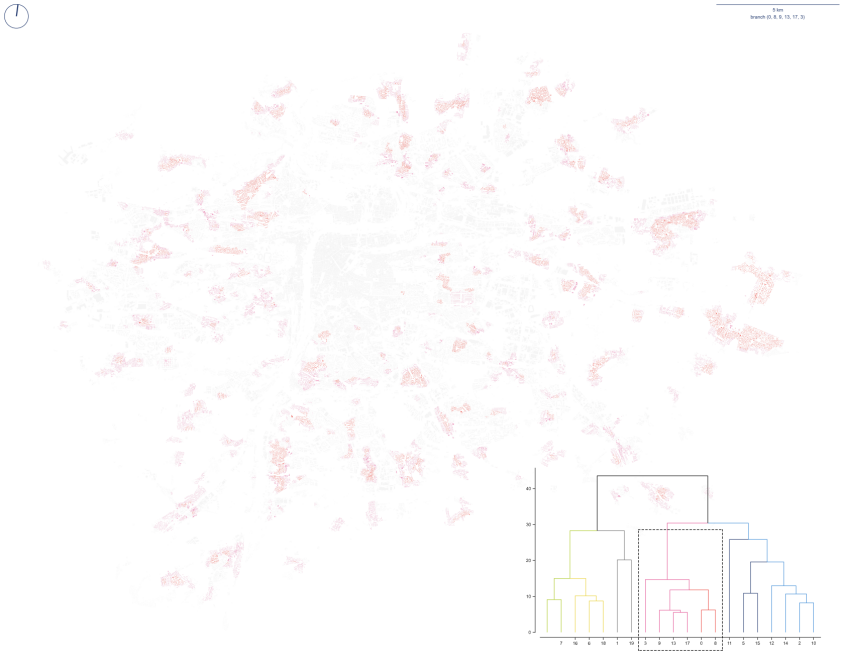


Figure A8.7: Spatial distribution of clusters within a branch 6

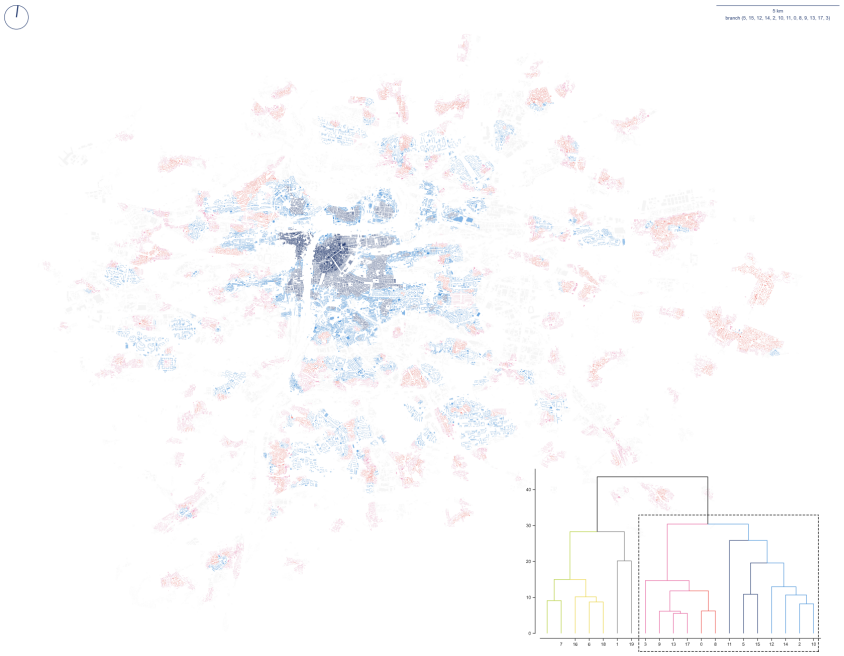


Figure A8.8: Spatial distribution of clusters within a branch 7

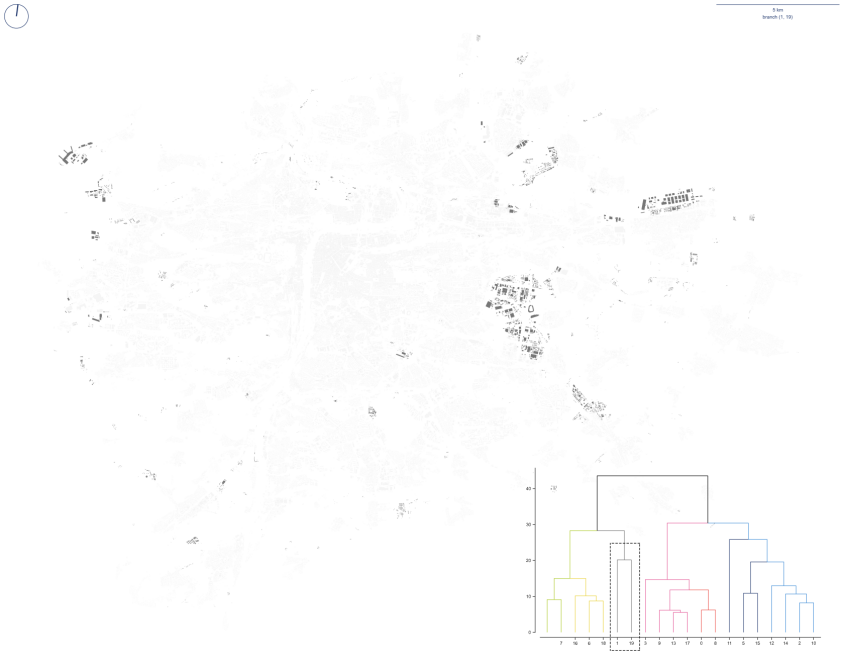


Figure A8.9: Spatial distribution of clusters within a branch 8

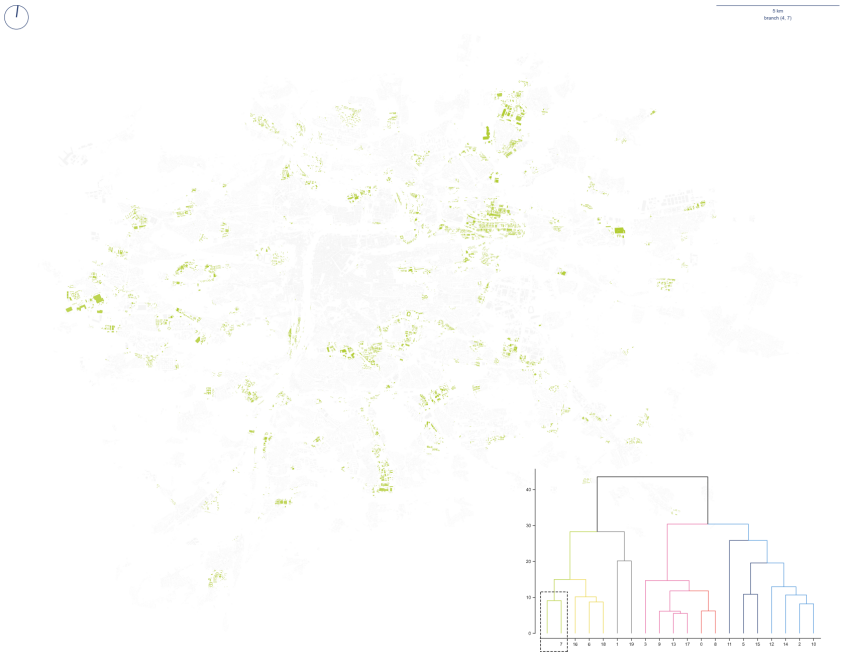


Figure A8.10: Spatial distribution of clusters within a branch 9

Figure A8.11: Spatial distribution of clusters within a branch 10

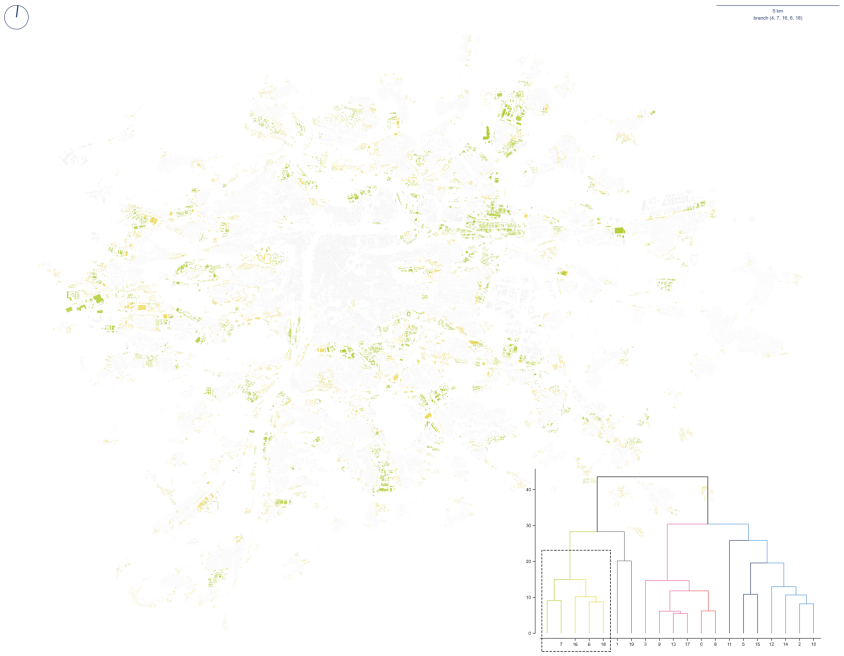


Figure A8.12: Spatial distribution of clusters within a branch 11

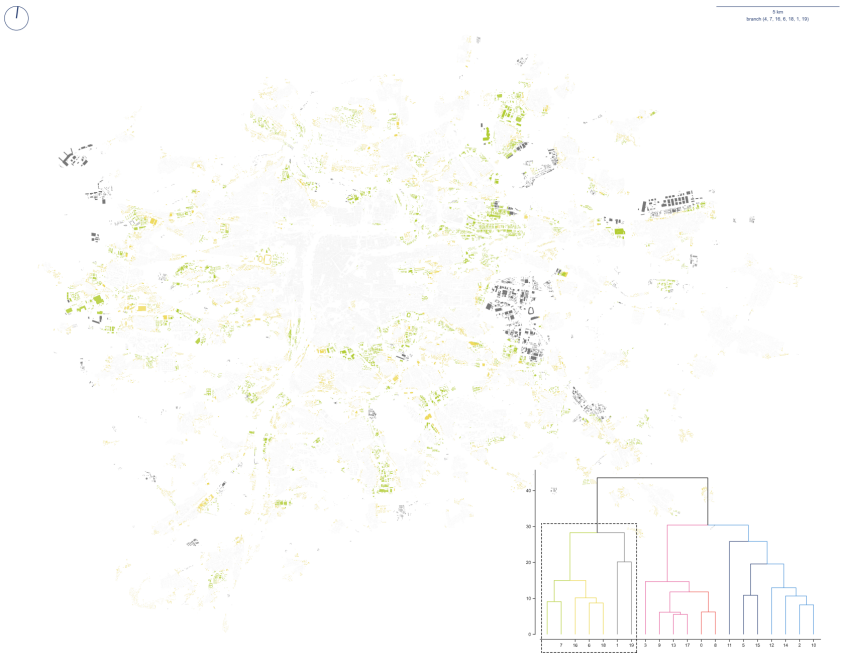


Figure A8.13: Spatial distribution of clusters within a branch 12

A8.2 Morphometric characters in Amsterdam

Reported morphometric data used in the section 8.3.2.2.

Descriptive summary values of all primary characters measured on Amsterdam data are presented in the table below.

Table A8.1: Overview of the primary morphometric values for the Amsterdam case study. The key to character IDs is available in Appendix 7.

id	mean	std	min	25%	50%	75%	max
sdbAre	180	1300	30	51	64	96	430000
sdbHei	8.3	4.8	3	4.8	7.1	11	120
sdbVol	1900	16000	90	300	440	980	3.7e+06
sdbPer	47	53	19	30	36	45	4700
sdbCoA	0.93	57	0	0	0	0	22000
ssbFoF	1.3	0.71	0.24	0.92	1.1	1.5	22
ssbVFR	2.3	1.8	0.66	1.6	1.8	2.2	92
ssbCCo	0.5	0.11	0.022	0.43	0.52	0.58	1
ssbCor	7.2	10	0	4	6	8	3000
ssbSqu	2.9	7.7	0	0.026	0.14	0.76	85
ssbERI	0.95	0.073	0.26	0.92	0.99	1	1.1
ssbElo	0.6	0.18	0.019	0.47	0.58	0.73	1
ssbCCD	1	2.1	0	0.0029	0.45	1.4	150
ssbCCM	7.4	6.5	0.35	5.3	5.9	7.1	360
stbOri	21	13	0	10	20	30	45
stbSAl	5.5	8.5	0	0.39	1.3	6.9	45
stbCeA	3.1	6.1	0	0.04	0.43	2.9	45
sdcLAL	51	38	8.3	30	37	54	1100
sdcAre	1100	4000	33	180	280	600	600000
sscCCo	0.31	0.15	0.027	0.2	0.26	0.41	0.99
sscERI	0.95	0.068	0.047	0.92	0.98	1	1.1
stcOri	21	13	0	10	20	30	45
stcSAl	6.6	8.8	0	0.63	2.4	9.4	45
sicCAR	0.26	0.15	0.00094	0.15	0.26	0.35	1.4

id	mean	std	min	25%	50%	75%	max
sicFAR	0.69	0.71	0.001	0.22	0.43	0.86	16
sdsLen	200	330	6.4	73	120	190	5500
sdsSPW	28	8.5	2.8	22	30	34	50
sdsSPH	8.6	4.7	0	5.5	7.1	11	86
sdsSPR	0.37	0.34	0	0.18	0.25	0.45	9.1
sdsSPO	0.41	0.22	0	0.24	0.38	0.56	1
sdsSWD	3	1.8	0	1.6	2.9	4.2	11
sdsSHD	1.2	1.4	0	0.26	0.81	1.4	47
sssLin	0.93	0.15	0	0.96	1	1	1
sdsAre	27000	88000	47	3700	7200	15000	1.2e+06
sisBpM	0.16	0.085	0.00053	0.096	0.15	0.22	1.3
sddAre	23000	59000	47	4700	7900	15000	640000
mtbSWR	0.42	0.24	0	0.27	0.5	0.62	1.7
mtbAli	2.1	3.8	0	0.075	0.4	2.6	45
mtbNDi	18	15	0	10	14	21	200
mtcWNe	0.069	0.033	0.0006	0.046	0.068	0.091	0.3
mdcAre	9800	21000	210	2000	3400	8400	920000
misRea	69	57	1	35	57	86	630
mdsAre	63000	130000	47	16000	26000	44000	1.4e+06
mtdDeg	3	0.79	1	3	3	3	6
mtdMDi	140	200	6.4	70	94	130	5500
midRea	81	63	1	41	68	100	620
midAre	70000	140000	47	20000	31000	53000	1.3e+06
libNCo	0.53	2.4	0	0	0	0	45
ldbPWL	240	280	20	88	140	280	9800
ltbIBD	20	9.7	0	14	18	24	200
ltcBuA	0.36	0.21	0.029	0.22	0.3	0.44	1
licGDe	0.64	0.62	0.0011	0.22	0.39	0.85	5.2
ltcWRB	0.00018	0.00014	1.2e-06	6.8e-05	0.00014	0.00025	0.0013
ldkAre	82000	180000	90	11000	20000	61000	1.8e+06
ldkPer	1200	1500	41	470	680	1300	12000
lskCCo	0.44	0.13	0.07	0.34	0.44	0.54	0.98

id	mean	std	min	25%	50%	75%	max
lskERI	0.87	0.14	0.14	0.8	0.91	0.97	1.1
lskCWA	300	500	0.43	75	140	280	4100
ltkOri	21	13	0.0019	10	20	30	45
ltkWNB	0.0092	0.0051	0	0.0055	0.0087	0.012	0.049
likWBB	0.0025	0.0016	4.3e-06	0.0011	0.0024	0.0036	0.013
lcdMes	0.14	0.061	-0.33	0.1	0.14	0.18	0.38
ldsMSL	2900	2400	6.4	1600	2300	3400	22000
ldsCDL	210	600	0	0	71	190	13000
ldsRea	230000	340000	6600	82000	120000	200000	3.7e+06
lddNDe	0.13	0.22	0	0.046	0.077	0.13	5.5
lddRea	280	180	1	150	240	370	1300
lddARe	240000	350000	4400	86000	130000	220000	3.6e+06
linPDE	0.11	0.093	0	0.042	0.095	0.16	1
linP3W	0.67	0.12	0	0.59	0.67	0.75	1
linP4W	0.21	0.12	0	0.13	0.2	0.29	0.71
linWID	0.25	0.42	0	0.092	0.16	0.26	10
lcnClo	4.4e-06	2e-06	0	2.9e-06	4.4e-06	5.6e-06	1.4e-05
xcnSCl	0.055	0.1	0	0	0	0.08	1

Figures A8.14 - A8.18 show histograms capturing the (truncated) distribution of all measured characters. Note the differences outlined above and overall variety of distributions.

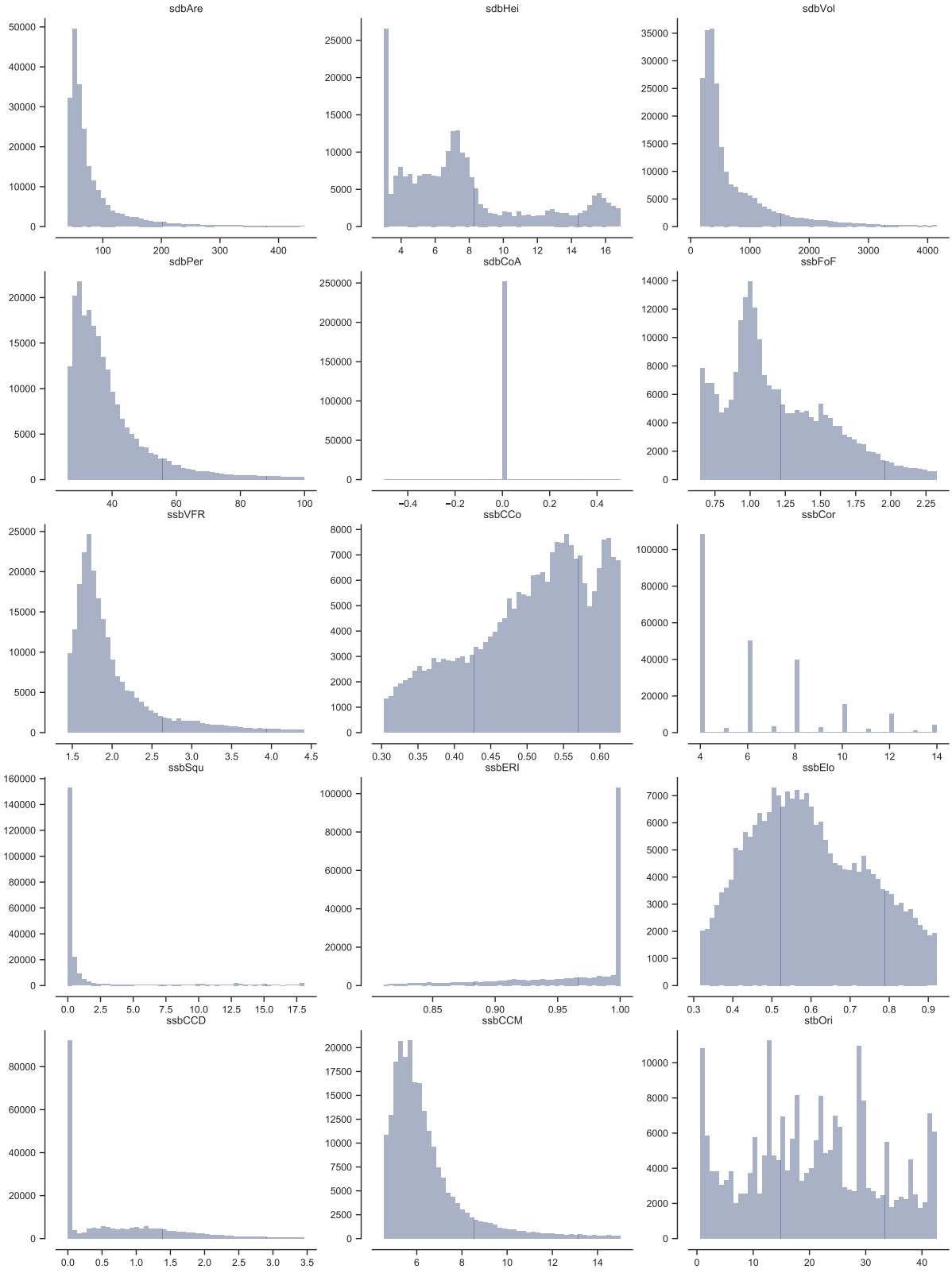


Figure A8.14: Histograms of characters 1-15 are showing the variety of distributions within the measured primary data (Amsterdam). Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table above are presented complete for reference.

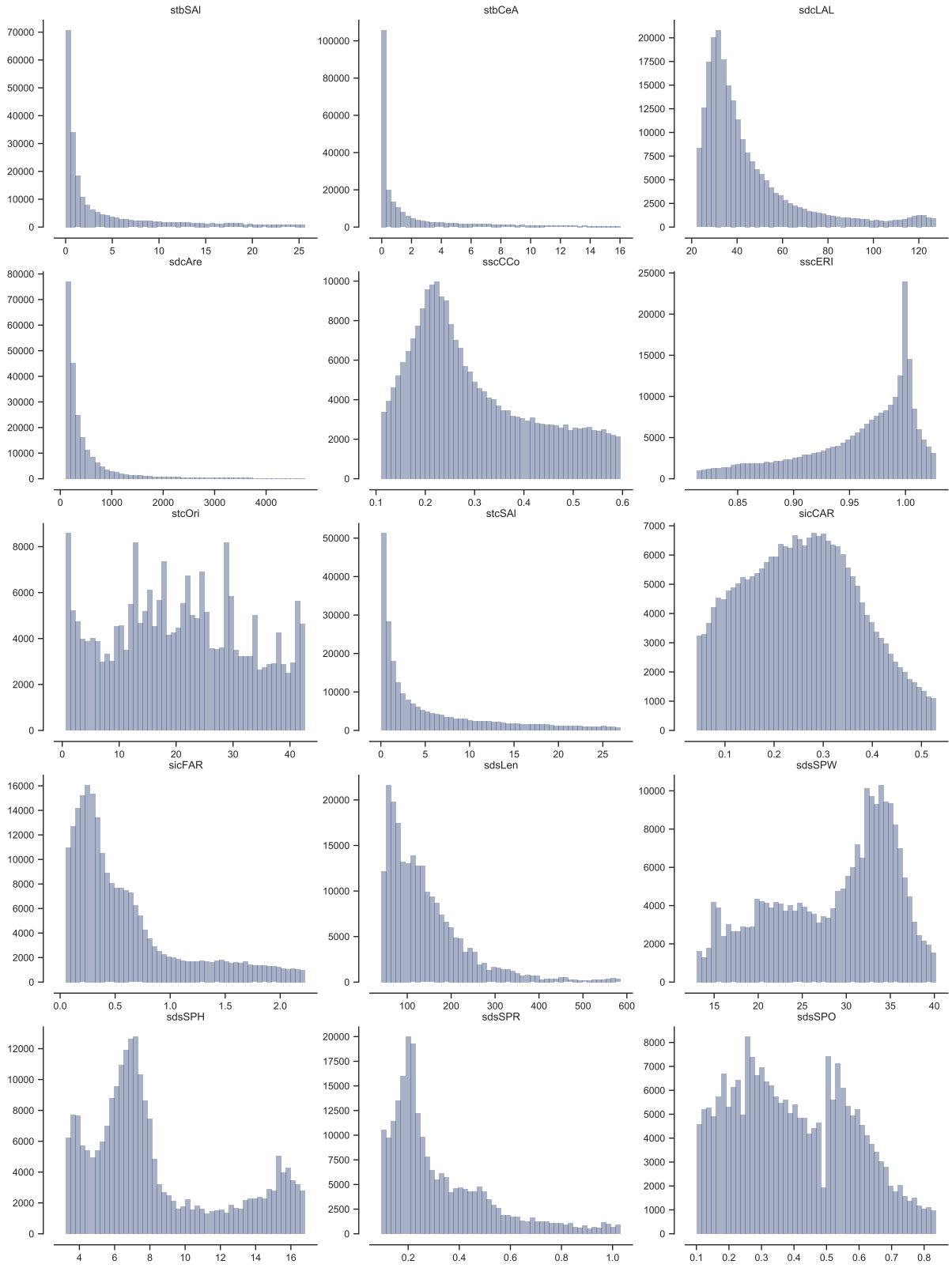


Figure A8.15: Histograms of characters 16-30 are showing the variety of distributions within the measured primary data (Amsterdam). Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table above are presented complete for reference.

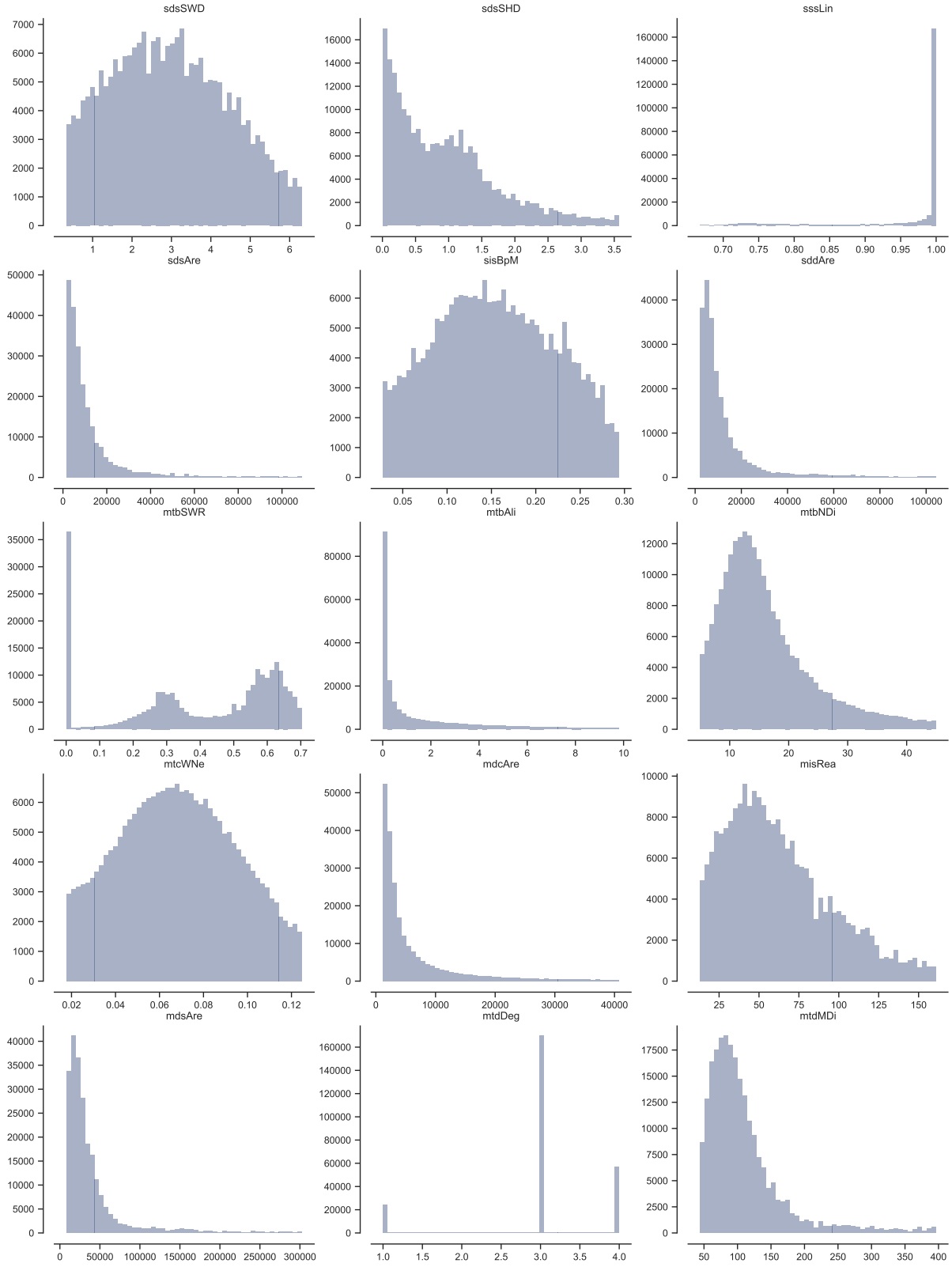


Figure A8.16: Histograms of characters 31-45 are showing the variety of distributions within the measured primary data (Amsterdam). Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table above are presented complete for reference.

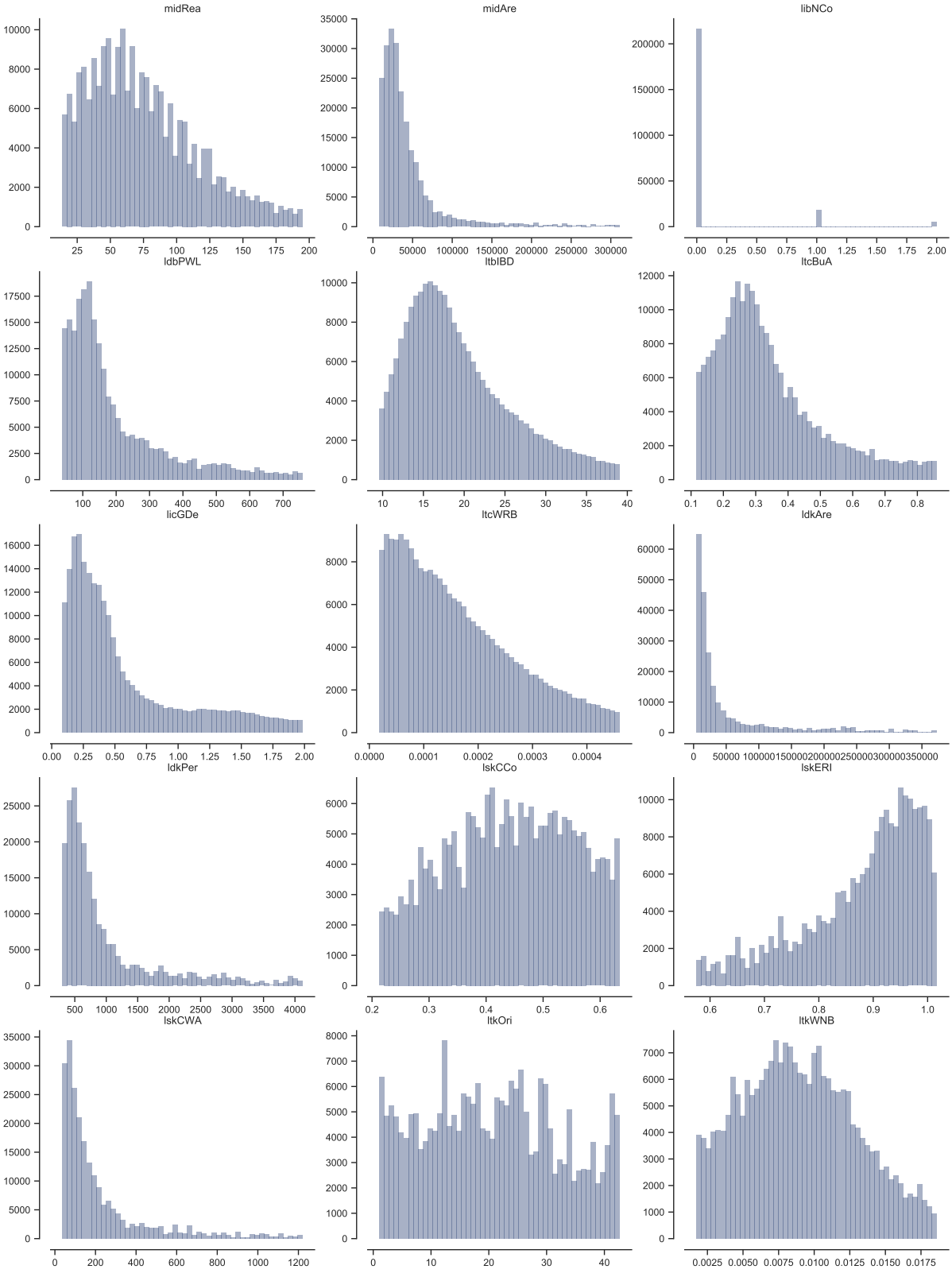


Figure A8.17: Histograms of characters 45-60 are showing the variety of distributions within the measured primary data (Amsterdam). Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table above are presented complete for reference.

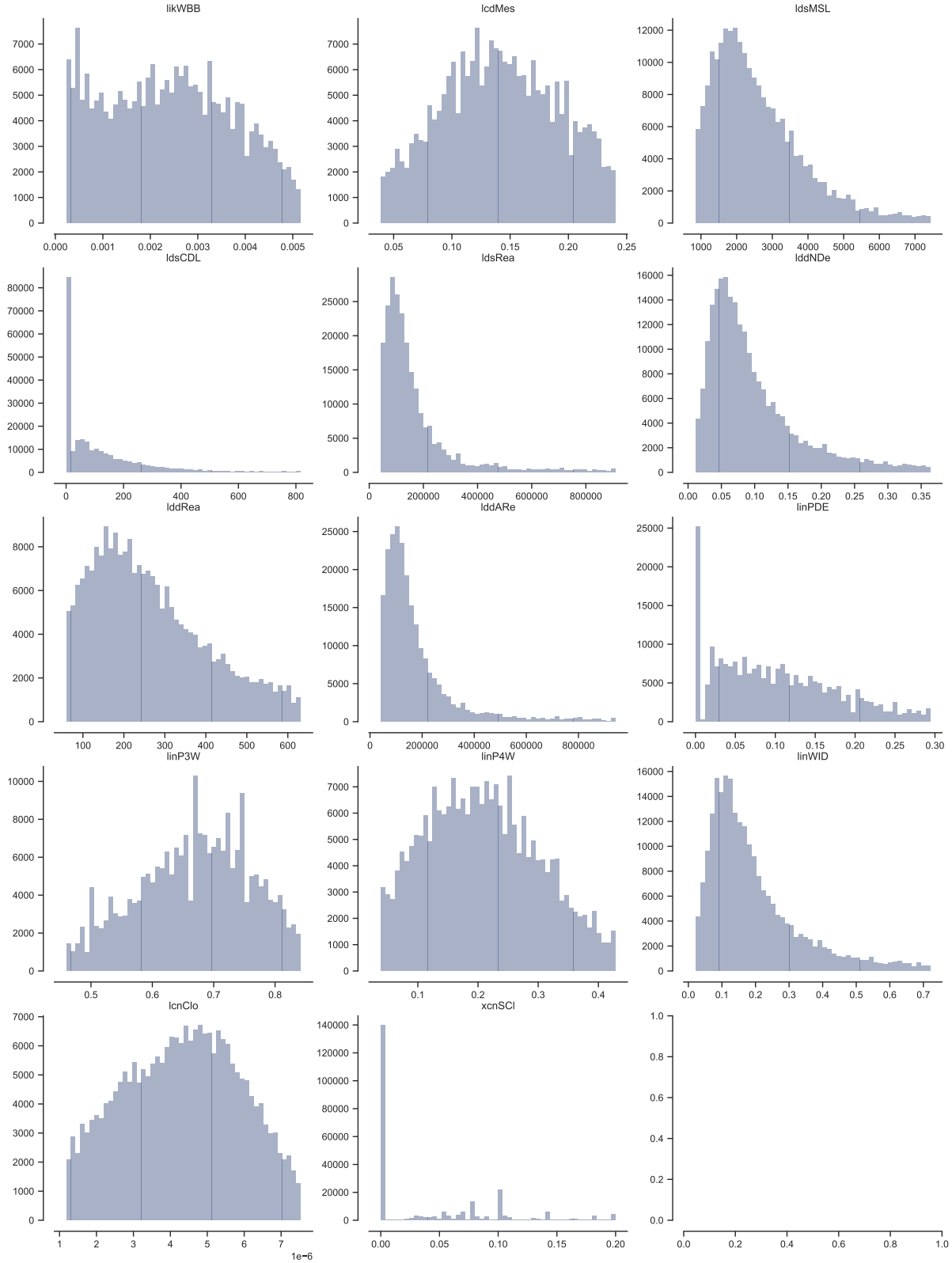


Figure A8.18: Histograms of characters 61-74 are showing the variety of distributions within the measured primary data (Amsterdam). Histograms illustrate data within percentiles (5, 95) to avoid extreme skewing due to the presence of outliers. Data in table above are presented complete for reference.

Appendix N: Reproducible computational Jupyter notebooks

Appendix N comprises a selection of key Jupyter notebooks used in the thesis. Notebooks contain reproducible Python code largely dependent on momepy, which code is not included in notebooks and is documented in Annexe 1. The goal of this appendix is to allow easy reproducibility of each critical step of the research. Notebooks mostly consist of code actual production code, which means that it may not always follow the optimal patterns. However, the major contribution of this research on the ground of reproducibility are not notebooks, but momepy as an independent fully documented and tested open-source package.

The PDF version of Jupyter notebooks spans across more than 100 pages, therefore is not directly included in the main submission. The complete set is available at:

martinfleischmann.net/thesis_notebooks.zip

Notebooks related to chapter 4 and part of chapter 6 have been already published alongside relevant peer-reviewed articles. Notebooks related to chapters 7 and 8 will be published alongside article summarising their contents.

The archive contains following notebooks:

- Chapter 4 - Matrix of Literature
- Chapter 4 - Quantitative assessment of the database
- Chapter 6 - Generate tessellation diagram
- Chapter 6 - Parameters optimisation analysis

- Chapter 6 - Generating morphological tessellation and measure morphometric characters
- Chapter 6 - Analysis of similarity of measured data
- Chapter 6 - Spatial autocorrelation
- Chapter 6 - Contiguity diagram
- Chapter 6 - Aggregation models
- Chapter 7 - Generate additional morphometric elements
- Chapter 7 - Measure contextual - spatially lagged characters
- Chapter 7 - Measure primary characters
- Chapter 7 + 8 - Cluster analysis + taxonomy
- Chapter 8 - Comparison of taxa and period of building origin
- Chapter 8 - Comparison of taxa and land use
- Chapter 8 - Comparison of taxa and qualitative classification

Annexe 1: momepy: Urban Morphology Measuring Toolkit

Annexe 1 consists of the open-source Python package `momepy`, which accompanies the work presented in this thesis, allows its reproducibility and lays a foundation for further morphometric research. Due to its nature, it is not possible to attach it directly to the thesis in an informative way. Therefore, please follow the links to the online presence of its source code and documentation.

Source code:

github.com/martinfleis/momepy

Online documentation:

docs.momepy.org

References

Abadi, M. *et al.* (2016) 'Tensorflow: A system for large-scale machine learning', in *12th usenix symposium on operating systems design and implementatiom*, pp. 265–283.

Adams, D. C., Rohlf, F. J. and Slice, D. E. (2004) 'Geometric morphometrics: Ten years of progress following the "revolution"', *Italian Journal of Zoology*, 71(1), pp. 5–16. doi: 10/cm9f7h.

Agresti, A. (2018) *An introduction to categorical data analysis*. John Wiley & Sons.

Agryzcov, T., Tortosa, L. and Vicent, J. F. (2018) 'An algorithm to compute data diversity index in spatial networks', *Applied Mathematics and Computation*, 337, pp. 63–75. doi: 10.1016/j.amc.2018.04.068.

Agryzcov, T. *et al.* (2017) 'A centrality measure for urban networks based on the eigenvector centrality concept', *Environment and Planning B: Urban Analytics and City Science*, 219(1), p. 239980831772444. doi: 10.1177/2399808317724444.

Ahlfeldt, G. M. and Pietrostefani, E. (2019) 'The economic effects of density: A synthesis', *Journal of Urban Economics*, 111, pp. 93–107. doi: 10/gg4vf4.

Ai, T. and Zhang, X. (2007) 'The Aggregation of Urban Building Clusters Based on the Skeleton Partitioning of Gap Space', in *OpenStreetMap in GIScience*. Berlin, Heidelberg: Springer International Publishing, pp. 153–170. doi: 10.1007/978-3-540-72385-1_9.

Akaike, H. (1973) 'Information theory and an extension of the maximum likelihood principle,[w:] Proceedings of the 2nd international symposium on information, bn petrow, f', *Czaki, Akademiai Kiado, Budapest*.

Alexander, C. (1966) 'A City is not a tree', p. 17.

Alexander, D. L. J., Tropsha, A. and Winkler, D. A. (2015) 'Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models', *Journal of chemical information and modeling*, 55(7), pp. 1316–1322. doi: 10.1021/acs.jcim.5b00206.

Alexiou, A., Singleton, A. and Longley, P. A. (2016) 'A Classification of Multidimensional Open Data for Urban Morphology', *Built Environment*, 42(3), pp. 382–395. doi: 10/gddwsn.

Andersson, E. and Colding, J. (2014) 'Understanding how built urban form influences biodiversity', *Urban Forestry & Urban Greening*, 13(2), pp. 221–226. doi: 10/f22sg2.

Angel, S. *et al.* (2018) 'The shape compactness of urban footprints', *Progress in Planning*. doi: 10.1016/j.progress.2018.12.001.

Angel, S. and Parent, J. (2007) 'Urban sprawl metrics: An analysis of global urban expansion using GIS', in *ASPRS 2007 Annual Conference*. Tampa.

Ankerst, M. *et al.* (1999) 'OPTICS: Ordering points to identify the clustering structure', *ACM Sigmod record*, 28(2), pp. 49–60. doi: 10/cnj38n.

Anselin, L. (2001) 'Spatial econometrics', *A companion to theoretical econometrics*, 310330.

Anselin, L. (2010) 'Local Indicators of Spatial Association-LISA', *Geographical Analysis*, 27(2), pp. 93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x.

Araldi, A. and Fusco, G. (2017) 'Decomposing and Recomposing Urban Fabric: The City from the Pedestrian Point of View', in *Computational Science and Its Applications ICCSA 2017*. Cham: Springer International Publishing, pp. 365–376. doi: 10.1007/978-3-319-62401-3_27.

Araldi, A. and Fusco, G. (2019) 'From the street to the metropolitan region: Pedestrian perspective in urban fabric analysis:' *Environment and Planning B: Urban Analytics and City Science*, 46(7), pp. 1243–1263. doi: 10.1177/2399808319832612.

Ariza-Villaverde, A. B., Jiménez-Hornero, F. J. and Ravé, E. G. D. (2013) 'Multifractal analysis of axial maps applied to the study of urban morphology', *Computers, Environment and Urban Systems*, 38(1), pp. 1–10. doi: 10.1016/j.compenvurbsys.2012.11.001.

Assunção, R. M. *et al.* (2006) 'Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees', *International Journal of Geographical Information Science*, 20(7), pp. 797–811. doi: 10/b5pw7c.

Bailey, K. D. (1994) *Typologies and Taxonomies - An Introduction to Classification Techniques*. Sage Publications.

Banister, D., Watson, S. and Wood, C. (1997) 'Sustainable cities: Transport, energy, and urban form', *Environment and Planning B: planning and design*, 24(1), pp. 125–143. doi: 10/fcww7z.

Barnsley, M. J. and Barr, S. L. (1996) 'Inferring Urban Land Use from Satellite Sensor Images Using Kernel-Based Spatial Reclassification', p. 10.

Barrington-Leigh, C. and Millard-Ball, A. (2017) 'The world's open-source street map is more than 80% complete', *PLoS ONE*, 12(8), pp. 1–20. doi: 10.1371/journal.pone.0180698.

Barron, C., Neis, P. and Zipf, A. (2014) 'A comprehensive framework for intrinsic OpenStreetMap quality

analysis', *Transactions in GIS*, 18(6), pp. 877–895. doi: 10/f6thzc.

Basaraner, M. and Cetinkaya, S. (2017) 'Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS', *International Journal of Geographical Information Science*, 31(10), pp. 1952–1977. doi: 10.1080/13658816.2017.1346257.

Basaraner, M. and Selcuk, M. (2004) 'An attempt to automated generalization of buildings and settlement areas in topographic maps', in *International Archives of Photogrammetry and Remote Sensing*. Istanbul.

Batty, M. (1997) 'Cellular Automata and Urban Form: A Primer', *Journal of the American Planning Association*, 63(2), pp. 266–274. doi: 10.1080/01944369708975918.

Batty, M. (2012) 'Building a science of cities', *Cities*, 29, pp. S9–S16. doi: 10.1016/j.cities.2011.11.008.

Batty, M. and Longley, P. A. (1987) 'Fractal-based description of urban form', *Environment and Planning B: Planning and Design*, 14(1961), pp. 123–134. doi: 10.1068/b140123.

Beresford, M. W. (1971) 'The back-to-back house in Leeds, 1787-1937', *The History of Working Class Housing*, pp. 93–132.

Berghauser Pont, M. and Haupt, P. (2010) *Spacematrix: Space, density and urban form*.

Berghauser Pont, M. and Marcus, L. H. (2014) 'Innovations in measuring density: From area and location density to accessible and perceived density', *Nordisk Arkitekturforskning*, (2).

Berghauser Pont, M. and Olsson, J. (2017) 'Typology based on three density variables central to Spacematrix using cluster analysis', in *ISUF XXIV international conference City and territory in the globalization age*. Universitat Politècnica de València, pp. 1337–1348. doi: 10.4995/ISUF2017.2017.5319.

Berghauser Pont, M. *et al.* (2019) 'The spatial distribution and frequency of street, plot and building types across five european cities', *Environment and Planning B: Urban Analytics and City Science*, 46(7), pp. 1226–1242.

Berghauser Pont, M. *et al.* (2017) 'Quantitative comparison of cities: Distribution of street and building types based on density and centrality measures', in *11th Space Syntax Symposium*.

Berghauser Pont, M., Stavroulaki, G. and Marcus, L. (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*, 46(8), pp. 1549–1564. doi: 10/gghf42.

Biljecki, F., Ledoux, H. and Stoter, J. (2016) 'An improved LOD specification for 3D building models', *Computers, Environment and Urban Systems*, 59, pp. 25–37. doi: 10/f83fz4.

Bingham, E. *et al.* (2019) 'Pyro: Deep universal probabilistic programming', *Journal of Machine Learning Research*, 20, pp. 28:1–28:6.

Biology-online.org (2020) 'Individual'. https://www.biologyonline.com/dictionary/individual.

Bobkova, E., Berghauser Pont, M. and Marcus, L. (2019) 'Towards analytical typologies of plot systems: Quantitative profile of five European cities', *Environment and Planning B: Urban Analytics and City Science*, p. 239980831988090. doi: 10/ggbgsm.

Bobkova, E., Marcus, L. and Berghauser Pont, M. (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*, (July).

Bodenstein, L. and Sidman, R. L. (1987) 'Growth and development of the mouse retinal pigment epithelium: I. Cell and tissue morphometrics and topography of mitotic activity', *Developmental biology*, 121(1), pp. 192–204.

Boeing, G. (2017a) 'Measuring the Complexity of Urban Form and Design', *SSRN Electronic Journal*. doi: 10.2139/ssrn.2958923.

Boeing, G. (2017b) 'OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks', *Computers, Environment and Urban Systems*, 65, pp. 126–139. doi: 10/gbvjxq.

Boeing, G. (2018) 'A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood', *Environment and Planning B: Urban Analytics and City Science*, 219(4), p. 239980831878459. doi: 10.1177/2399808318784595.

Boeing, G. (2020a) 'Off the grid… and back again? The recent evolution of american street network planning and design', *Journal of the American Planning Association*, pp. 1–15. doi: 10/ghf423.

Boeing, G. (2020b) 'The right tools for the job: The case for spatial science tool-building', *Transactions in GIS*, 24(5), pp. 1299–1314. doi: 10/ghhjgp.

Borruso, G. (2003) 'Network Density and the Delimitation of Urban Areas', *Transactions in GIS*, 7(2), pp. 177–191. doi: 10.1111/1467-9671.00139.

Bourdic, L., Salat, S. and Nowacki, C. (2012) 'Assessing cities: A new system of cross-scale spatial indicators', *Building Research & Information*, 40(5), pp. 592–605. doi: 10.1080/09613218.2012.703488.

Bramley, G. and Power, S. (2009) 'Urban form and social sustainability: The role of density and housing type', *Environment and Planning B: Planning and Design*, 36(1), pp. 30–48. doi: 10/dm7h5x.

Breno, M., Leirs, H. and Van Dongen, S. (2011) 'Traditional and geometric morphometrics for studying skull morphology during growth in *Mastomys Natalensis* (Rodentia: Muridae)', *Journal of Mammalogy*, 92(6), pp. 1395–1406. doi: 10/dz4nrn.

Burton, E. (2002) 'Measuring urban compactness in UK towns and cities', *Environment and Planning B: Planning and Design*, 29(2), pp. 219–250. doi: 10.1068/b2713.

Cambridge Dictionary (2020) 'Individual'. https://dictionary.cambridge.org/dictionary/english/individual.

Caniggia, G. and Maffei, G. L. (1979) 'Compositione architettonica e tipologia edilizia. I Lettura dell'edilizia di base', *Venice, Italy*.

Caniggia, G. and Maffei, G. L. (2001) *Architectural composition and building typology: Interpreting basic building*. Firenze: Alinea Editrice.

Carneiro, C. *et al.* (2010) 'Digital Urban Morphometrics: Automatic Extraction and Assessment of Morphological Properties of Buildings: Digital Urban Morphometrics', *Transactions in GIS*, 14(4), pp. 497–531. doi: 10/c5w6fp.

Caruso, G., Hilal, M. and Thomas, I. (2017) 'Measuring urban forms from inter-building distances: Combining MST graphs with a Local Index of Spatial Association', *Landscape and Urban Planning*, 163, pp. 80–89. doi: 10.1016/j.landurbplan.2017.03.003.

Castro, K. B. de *et al.* (2019) 'New perspectives in land use mapping based on urban morphology: A case study of the Federal District, Brazil', *Land Use Policy*, 87, p. 104032. doi: 10.1016/j.landusepol.2019.104032.

Cheng, J. (2011) 'Exploring urban morphology using multi-temporal urban growth data: A case study of Wuhan, China', *Asian Geographer*, 28(2), pp. 85–103. doi: 10.1080/10225706.2011.623399.

Coblentz, D., Pabian, F. and Prasad, L. (2014) 'Quantitative geomorphometrics for terrain characterization', *International Journal of Geosciences*, 2014.

Colaninno, N., Roca, J. and Pfeffer, K. (2011) 'An automatic classification of urban texture: Form and compactness of morphological homogeneous structures in Barcelona', in *51st Congress of the European Regional Science Association*. Louvain-la-Neuve: European Regional Science Association (ERSA).

Conzen, M. (1960) *Alnwick, Northumberland: A study in town-plan analysis*. London: George Philip & Son.

Conzen, M. R. (2004) *Thinking about urban form: Papers on urban morphology, 1932-1998*. Peter Lang.

Conzen, M. R. G. (1969) *Alnwick, Northumberland: A study in town-plan analysis*. Institute of British Geographers.

Cooper, J. (2003) 'Fractal assessment of street-level skylines: A possible means of assessing and comparing character', *Urban Morphology*, 7(2), pp. 73–82.

Cooper, J. (2005) 'Assessing urban character: The use of fractal analysis of street edges', *Urban Morphology*, 9(2), pp. 95–107. doi: 10.1108/17506200710779521.

Crewson, P. (2006) 'Applied statistics handbook', *AcaStat Software, Leesburg*.

Cuthbert, A. R. (2007) 'Urban design: Requiem for an eraReview and critique of the last 50 years', *Urban Design International*, 12(4), pp. 177–223. doi: 10/c8csb2.

D'Agostino, R. B. (1986) *Goodness-of-fit-techniques*. CRC press.

De Queiroz, K. (1998) 'The general lineage concept of species, species criteria, and the process of speciation', *Endless forms: species and speciation*.

Dibble, J. (2016) *Urban Morphometrics: Towards a Quantitative Science of Urban Form*. PhD thesis.

Dibble, J. *et al.* (2015) 'Urban Morphometrics: Towards a Science of Urban Evolution', *arXiv.org*, physics.soc-ph.

Dibble, J. *et al.* (2017) 'On the origin of spaces: Morphometric foundations of urban form evolution', *Environment and Planning B: Urban Analytics and City Science*, 46(4), pp. 707–730. doi: 10.1177/2399808317725075.

Dictionary.com (2020) 'Individual'. https://www.dictionary.com/browse/individual?s=t.

Dill, J. (2004) 'Measuring network connectivity for bicycling and walking', *83rd Annual Meeting of the Transportation …*, (1), pp. 20–20.

Dogrusoz, E. and Aksoy, S. (2007) 'Modeling urban structures using graph-based spatial patterns', in. IEEE, pp. 4826–4829. doi: 10.1109/IGARSS.2007.4423941.

Dong, J., Li, L. and Han, D. (2019) 'New Quantitative Approach for the Morphological Similarity Analysis of Urban Fabrics Based on a Convolutional Autoencoder', *IEEE Access*, 7, pp. 138162–138174. doi: 10.1109/ACCESS.2019.2931958.

Duchêne, C., Bard, S. and Barillot, X. (2003) 'Quantitative and qualitative description of building orientation', in *He 5th ICA workshop on progress in automated map generalization*. Paris.

Dukai, B. (2020) '3D Registration of Buildings and Addresses (BAG) / 3D Basisregistratie Adressen en Gebouwen (BAG)', *4TU.ResearchData*. doi: 10.4121/uuid:f1f9759d-024a-492a-b821-07014dd6131c.

Duque, J. C., Anselin, L. and Rey, S. J. (2012) 'The max-p-regions problem', *Journal of Regional Science*, 52(3), pp. 397–419. doi: 10/cf9h6h.

Ester, M. *et al.* (1996) 'A density-based algorithm for discovering clusters in large spatial databases with noise.', in *Kdd*. (34), pp. 226–231.

European Environment Agency (1990) 'CORINE Land Cover', pp. 1–163.

Ewing, R. *et al.* (2016) 'Does urban sprawl hold down upward mobility?', *Landscape and Urban Planning*, 148, pp. 80–88. doi: 10/f8ghxs.

Ewing, R. *et al.* (2006) 'Identifying and Measuring Urban Design Qualities Related to Walkability', *Journal of Physical Activity and Health*, 3(Suppl 1), pp. 223–240. doi: 10.13072/midss.126.

Ewing, R. and Rong, F. (2008) 'The impact of urban form on US residential energy use', *Housing policy*

debate, 19(1), pp. 1–30. doi: 10/fgf7q8.

Feliciotti, A. (2018) *RESILIENCE AND URBAN DESIGN: A SYSTEMS APPROACH TO THE STUDY OF RESILIENCE IN URBAN FORM*. PhD thesis. University of Strathclyde.

Feliciotti, A., Romice, O. and Porta, S. (2016) 'Design for change : Five proxies for resilience in the urban form', *Open House International*, 41(4), pp. 22–30.

Feliciotti, A., Romice, O. and Porta, S. (2017) 'Urban regeneration , masterplans and resilience : The case of Gorbals , Glasgow', *Urban Morphology*, 21(1), pp. 61–79.

Feng, J. and Xie, S. (2013) 'Numerical Taxonomy of Species in the Genus Mallomonas (Chrysophyta) from China', *ISRN Biodiversity*. Edited by I. Bisht and P. De los Ríos Escalante, 2013, p. 653958. doi: 10/gb67xz.

Fernbach, P. and Sloman, S. (2017) *The knowledge illusion*. Penguin Random House Audio Publishing Group.

Fleischmann, M. (2019) 'Momepy: Urban Morphology Measuring Toolkit', *Journal of Open Source Software*, 4(43), p. 1807. doi: 10.21105/joss.01807.

Fleischmann, M. *et al.* (2020) 'Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale', *Computers, Environment and Urban Systems*, 80, p. 101441. doi: 10.1016/j.compenvurbsys.2019.101441.

Fleischmann, M., Romice, O. and Porta, S. (2020) 'Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities', *Environment and Planning B: Urban Analytics and City Science*, p. 239980832091044. doi: 10/ggngw6.

Forster, C. A. (1972) *Court housing in Kingston upon Hull: An example of cyclic processes in the morphological development of nineteenth-century bye-law housing*. University of Hull (19).

Frankhauser, P. (2004) 'Comparing the morphology of urban patterns in Europe a fractal approach', *European Cities Insights on outskirts*, 2, pp. 79–105.

Galster, G. C. *et al.* (2001) 'Wrestling Sprawl to the Ground: Defining and measuring an elusive concept', *Housing Policy Debate*, 12(4), pp. 681–717. doi: 10.1080/10511482.2001.9521426.

Gielen, E. *et al.* (2017) 'An urban sprawl index based on multivariate and Bayesian factor analysis with application at the municipality level in Valencia', *Environment and Planning B: Urban Analytics and City Science*. doi: 10.1177/2399808317690148.

Gil, J. (2014) 'Analyzing the configuration of multimodal urban networks', *Geographical Analysis*, 46(4), pp. 368–391. doi: 10.1111/gean.12062.

Gil, J. (2016) 'Street network analysis "edge effects": Examining the sensitivity of centrality

measures to boundary conditions', *Environment and Planning B: Planning and Design*. doi: 10.1177/0265813516650678.

Gil, J. *et al.* (2012) 'On the Discovery of Urban Typologies: Data Mining the Multi-dimensional Character of Neighbourhoods', *Urban Morphology*, 16(1), pp. 27–40.

Gillies, S. and others (2007) 'Shapely: Manipulation and analysis of geometric objects'. toblerity.org.

González-Bailón, S. (2013) 'Big data and the fabric of human geography', *Dialogues in Human Geography*, 3(3), pp. 292–296. doi: 10/ggnpg2.

Hagberg, A. A., Schult, D. A. and Swart, P. J. (2008) 'Exploring Network Structure, Dynamics, and Function using NetworkX', in el Varoquaux, G., Vaught, T., and Millman, J. (eds) *Proceedings of the 7th Python in Science Conference*. Pasadena, CA USA, pp. 11–15.

Haggag, M. A. and Ayad, H. M. (2002) 'The urban structural units method: A basis for evaluating environmental prospects for sustainable development', *URBAN DESIGN International*, 7(2), pp. 97–108. doi: 10.1057/palgrave.udi.9000071.

Haklay, M. and Weber, P. (2008) 'Openstreetmap: User-generated street maps', *IEEE Pervasive Computing*, 7(4), pp. 12–18. doi: 10/b24hbq.

Hallowell, G. D. and Baran, P. K. (2013) 'Suburban change: A time series approach to measuring form and spatial configuration', *The Journal of Space Syntax*, 4(1), pp. 74–91.

Hamaina, R., Leduc, T. and Moreau, G. (2012) 'Towards Urban Fabrics Characterization Based on Buildings Footprints', in *Bridging the Geographic Information Sciences*. Berlin, Heidelberg: Springer, Berlin, Heidelberg, pp. 327–346. doi: 10.1007/978-3-642-29063-3_18.

Hamaina, R., Leduc, T. and Moreau, G. (2013) 'A New Method to Characterize Density Adapted to a Coarse City Model', in *OpenStreetMap in GIScience*. Berlin, Heidelberg: Springer International Publishing, pp. 249–263. doi: 10.1007/978-3-642-31833-7_16.

Hartmann, A. *et al.* (2016) 'A Workflow for Automatic Quantification of Structure and Dynamic of the German Building Stock Using Official Spatial Data', *ISPRS International Journal of Geo-Information*, 5(8), p. 142. doi: 10/f872vh.

Hausleitner, B. and Berghauser Pont, M. (2017) 'Development of a configurational typology for micro-businesses integrating geometric and configurational variables', in *11th Space Syntax Symposium*.

Herold, M., Scepan, J. and Clarke, K. C. (2002) 'The Use of Remote Sensing and Landscape Metrics to Describe Structures and Changes in Urban Land Uses', *Environment and Planning A*, 34(8), pp. 1443–1458. doi: 10.1068/a3496.

Hijazi, I. *et al.* (2016) 'Measuring the homogeneity of urban fabric using 2D geometry data', *Environment and Planning B: Planning and Design*, pp. 1–25. doi: 10.1177/0265813516659070.

Hillier, B. (1996) *Space is the machine : A configurational theory of architecture*. Cambridge: Cambridge University Press.

Hinkle, D. E., Wiersma, W. and Jurs, S. G. (2003) *Applied statistics for the behavioral sciences*. Houghton Mifflin College Division.

Hrůza, J. (2003) 'Urbanismus světových velkoměst', *Díl: Praha. Praha: Vydavatelství*.

Huang, J., Lu, X. X. and Sellers, J. M. (2007) 'A global comparative analysis of urban form: Applying spatial metrics and remote sensing', *Landscape and Urban Planning*, 82(4), pp. 184–197. doi: 10.1016/j.landurbplan.2007.02.010.

Hughes, J. N. *et al.* (2015) 'Geomesa: A distributed architecture for spatio-temporal fusion', in *Geospatial informatics, fusion, and motion video analytics v*. International Society for Optics and Photonics, p. 94730F.

Huijboom, N. and Van den Broek, T. (2011) 'Open data: An international comparison of strategies', *European journal of ePractice*, 12(1), pp. 4–16.

Institut plánování a rozvoje hlavního města Prahy (2018) *Územní plán hlavního města Prahy Metropolitní plán Návrh k projednání dle 50 stavebního zákona*. Praha: IPR Praha.

Jacobs, J. (1961) *The death and life of American cities*.

Jain, A. K. and Dubes, R. C. (1988) *Algorithms for clustering data*. Prentice-Hall, Inc.

Jenks, G. F. (1967) 'The data model concept in statistical mapping', *International yearbook of cartography*, 7, pp. 186–190.

Jiang, B. (2013) 'Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution', *The Professional Geographer*, 65(3), pp. 482–494. doi: 10/f24r6j.

Jochem, W. C. *et al.* (2020) 'Classifying settlement types from multi-scale spatial patterns of building footprints', *Environment and Planning B: Urban Analytics and City Science*, p. 239980832092120. doi: 10/ggtsbn.

Jones, E. *et al.* (2001) 'SciPy'.

Jordahl, K. *et al.* (2020) 'Geopandas/geopandas: V0.8.1'. Zenodo. doi: 10.5281/zenodo.3946761.

Jost, L. (2006) 'Entropy and diversity', *Oikos*, 113(2), pp. 363–375. doi: 10.1111/j.2006.0030-1299.14714.x.

Khirfan, L. (2011) 'Understanding the links between inherited built forms and urban design: Athens and Alexandria as case studies', *Urban Morphology*, 15(1), pp. 39–53.

Kokoska, S. and Zwillinger, D. (2000) *CRC standard probability and statistics tables and formulae*. Hobo-

ken, NJ: Taylor and Francis.

Kong, H. and Sui, D. Z. (2016) 'Integrating the normative with the positive dimension of the new science for cities: A geodesign-based framework for Cellular Automata modeling', *Environment and Planning B: Planning and Design*. doi: 10.1177/0265813516651085.

Krenz, K. (2018) *NETWORK CENTRALITIES IN POLYCENTRIC URBAN REGIONS: Methods for the Measurement of Spatial Metrics*. PhD thesis. Bartlett Faculty of the Built Environment University College London.

Krizek, K. J. (2003) 'Operationalizing Neighborhood Accessibility for Land Use-Travel Behavior Research and Regional Modeling', *Journal of Planning Education and Research*, 22(3), pp. 270–287. doi: 10.1177/0739456X02250315.

Kropf, K. (1996) 'Urban tissue and the character of towns', *URBAN DESIGN International*, 1(3), pp. 247–263. doi: 10.1057/udi.1996.32.

Kropf, K. (1997) 'When is a plot not a plot: Problems in representation and interpretation', in *Fourth International Seminar on Urban Form*. Birmingham.

Kropf, K. (2017) *The handbook of urban morphology*. Chichester: John Wiley & Sons.

Lai, P.-C. *et al.* (2018) 'Neighborhood Variation of Sustainable Urban Morphological Characteristics', 15(3), pp. 465–413. doi: 10.3390/ijerph15030465.

Larkham, P. J. and Jones, A. N. (1991) *A glossary of urban form*. Urban Morphology Research Group, School of Geography, University of Birmingham.

Laskari, A., Hanna, S. and Derix, C. (2008) 'Urban identity through quantifiable spatial attributes Coherence and dispersion of local identity through the automated comparative analysis of building block plans', *Design Computing and Cognition '08*, (June), pp. 615–634. doi: 10.1007/978-1-4020-8728-8_32.

Lehner, A. and Blaschke, T. (2019) 'A Generic Classification Scheme for Urban Structure Types', *Remote Sensing*, 11(2), p. 173. doi: 10.3390/rs11020173.

Levy, A. (1999) 'Urban morphology and the problem of the modern urban fabric: Some questions for research', *Urban Morphology*, 3, pp. 79–85.

Li, X. *et al.* (2020) 'Data analytics of urban fabric metrics for smart cities', *Future Generation Computer Systems*, 107, pp. 871–882. doi: 10.1016/j.future.2018.02.017.

Li, Z. *et al.* (2004) 'Automated building generalization based on urban morphology and Gestalt theory', *International Journal of Geographical Information Science*, 18(5), pp. 513–534. doi: 10.1080/13658810410001702021.

Lind, P. G., González, M. C. and Herrmann, H. J. (2005) 'Cycles and clustering in bipartite networks',

Physical Review E, 72(5), p. 056127. doi: 10/c6m9xd.

Liqiang, Z. *et al.* (2013) 'A spatial cognition-based urban building clustering approach and its applications', *International Journal of Geographical Information Science*, 27(4), pp. 721–740. doi: 10.1080/13658816.2012.700518.

Liu, Y. *et al.* (2014) 'A Combined Approach to Cartographic Displacement for Buildings Based on Skeleton and Improved Elastic Beam Algorithm', *PLoS ONE*. Edited by R. Ji, 9(12), pp. e113953–27. doi: 10.1371/journal.pone.0113953.

Louf, R. and Barthelemy, M. (2014) 'A typology of street patterns', *Journal of the Royal Society Interface*, 11. doi: http://dx.doi.org/10.1098/rsif.2014.0924.

Lowry, J. H. and Lowry, M. B. (2014) 'Comparing spatial metrics that quantify urban form', *Computers, Environment and Urban Systems*, 44, pp. 59–67. doi: 10.1016/j.compenvurbsys.2013.11.005.

Lynch, K. (1960) *The image of the city*. MIT press.

MacQueen, J. and others (1967) 'Some methods for classification and analysis of multivariate observations', in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA, pp. 281–297.

Marcus, L. H. (2010) 'Spatial Capital', *The Journal of Space Syntax*, 1(1), pp. 30–40.

Marcus, L. H., Berghauser Pont, M. and Bobkova, E. (2017) 'Cities as accessible densities and diversities', in *11th Space Syntax Symposium*, pp. 1–13.

Marradi, A. (1990) 'Classification, typology, taxonomy', *Quality and Quantity*, 24(2), pp. 129–157. doi: 10.1007/BF00209548.

Maxar (2020) 'Optical Imagery'.

McInnes, L., Healy, J. and Astels, S. (2017) 'Hdbscan: Hierarchical density based clustering', *Journal of Open Source Software*, 2(11), p. 205. doi: 10/ggfp85.

Mehaffy, M. W. *et al.* (2010) 'Urban nuclei and the geometry of streets: The "emergent neighborhoods" model', *URBAN DESIGN International*, 15(1), pp. 22–46. doi: 10.1057/udi.2009.26.

Merriam-Webster.com (2020) 'Individual'. https://www.merriam-webster.com/dictionary/individual.

Moran, P. A. (1950) 'Notes on continuous stochastic phenomena', *Biometrika*, 37(1/2), pp. 17–23.

Mottelson, J. and Venerandi, A. (2020) 'A fine-grain multi-indicator analysis of the urban form of five informal settlements in east africa', *Urban Science*, 4(3), p. 31. doi: 10/ghhffj.

Moudon, A. V. (1986) *Built for change: Neighborhood architecture in San Francisco*. Mit Press.

Moudon, A. V. (1997) 'Urban morphology as an emerging interdisciplinary field', *Urban Morphology*, 1(1), pp. 3–10.

Muratori, S. (1959) 'Studi per una operante storia urbana di Venezia', *Palladio. Rivista di storia dell'architettura*, 1959, pp. 1–113.

Nedovic-Budic, Z. *et al.* (2016) 'Measuring urban form at community scale: Case study of Dublin, Ireland', *Cities*, 55, pp. 148–164. doi: 10.1016/j.cities.2016.02.014.

Neidhart, H. and Sester, M. (2004) 'Identifying building types and building clusters using 3-D laser scanning and GIS-data', *Int Arch Photogramm Remote Sens Spatial Inf Sci*, 35, pp. 715–720.

Novotnỳ, J. (2007) 'On the measurement of regional inequality: Does spatial dimension of income inequality matter?', *The Annals of Regional Science*, 41(3), pp. 563–580. doi: 10/dtt96d.

Oliveira, V. (2013) 'Morpho: A methodology for assessing urban form', *Urban Morphology*, 17(1), pp. 21–33.

Oliveira, V. (2016) *Urban Morphology: An Introduction to the Study of the Physical Form of Cities*. Cham: Springer International Publishing.

Oliveira, V. and Yaygin, M. A. (2020) 'The concept of the morphological region: Developments and prospects', *Urban Morphology*, 24(1), p. 18.

Omer, I. and Kaplan, N. (2018) 'Structural properties of the angular and metric street network's centralities and their implications for movement flows', *Environment and Planning B: Urban Analytics and City Science*, 17(2), p. 239980831876057. doi: 10.1177/2399808318760571.

Openshaw, S. (1984) *The Modifiable Areal Unit Problem*.

Oxford English Dictionary (2020a) '"Classification, n."' https://www.oed.com/view/Entry/33896?redirectedFrom=classifi

Oxford English Dictionary (2020b) '"Primary, adj. And n."'. https://www.oed.com/view/Entry/151280.

Panerai, P. *et al.* (2004) *Urban forms: The death and life of the urban block*. Routledge.

Park, H.-S. and Jun, C.-H. (2009) 'A simple and fast algorithm for K-medoids clustering', *Expert systems with applications*, 36(2), pp. 3336–3341. doi: 10/bdqjkt.

Paszke, A. *et al.* (2019) 'Pytorch: An imperative style, high-performance deep learning library', in *Advances in neural information processing systems*, pp. 8026–8037.

Pauleit, S. and Duhme, F. (2000) 'Assessing the environmental performance of land cover types for urban planning', *Landscape and Urban Planning*, 52(1), pp. 1–20. doi: 10.1016/s0169-2046(00)00109-2.

Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine learning in python', *the Journal of machine Learning research*, 12, pp. 2825–2830.

Peponis, J. *et al.* (2007) 'Measuring the configuration of street networks: The spatial profiles of 118 urban areas in the 12 most populated metropolitan regions in the US', in *6th International Space Syntax Symposium*.

Pham, H. M., Yamaguchi, Y. and Bui, T. Q. (2011) 'A case study on the relation between city planning and urban growth using remote sensing and spatial metrics', *Landscape and Urban Planning*, 100(3), pp. 223–230. doi: 10.1016/j.landurbplan.2010.12.009.

planet (2020) '50cm SKYSAT IMAGERY'.

Porta, S., Crucitti, P. and Latora, V. (2006) 'The network analysis of urban streets: A primal approach', *Environment and Planning B: Planning and Design*, 33(5), pp. 705–725. doi: 10.1068/b32045.

Porta, S., Latora, V. and Strano, E. (2010) 'Networks in Urban Design. Six Years of Research in Multiple Centrality Assessment', in *Network Science*. London: Springer London, pp. 107–129.

Porta, S. and Romice, O. (2014) 'Plot-based urbanism : Towards time-consciousness in place-making', in Mäckler, C. and Sonne, W. (eds) *Dortmunder Vorträge zur Stadtbaukunst [Dortmunder Lectures on Civic Art]*. Sulgen, DE: Niggli, pp. 82–111.

Porta, S. *et al.* (2014) 'Alterations in scale: Patterns of change in main street networks across time and space', *Urban Studies*, 51(16), pp. 3383–3400. doi: 10.1177/0042098013519833.

Randall, T. A. and Baetz, B. W. (2001) 'Evaluating Pedestrian Connectivity for Suburban Sustainability', *Journal of Urban Planning and Development*, 127(1), pp. 1–15. doi: 10.1061/(ASCE)0733-9488(2001)127:1(1).

Rey, S. J. (2019) 'PySAL: The first 10 years', *Spatial Economic Analysis*, 14(3), pp. 273–282. doi: 10/ghhffd.

Rey, S. J. and Anselin, L. (2007) 'PySAL: A Python Library of Spatial Analytical Methods', *The Review of Regional Studies*, 37(1), pp. 5–27.

Reynolds, D. A. (2009) 'Gaussian mixture models.', *Encyclopedia of biometrics*, 741. doi: 10/cqtzqm.

Rocklin, M. (2015) 'Dask: Parallel computation with blocked algorithms and task scheduling', in Huff, K. and James Bergstra (eds) *Proceedings of the 14th python in science conference*, pp. 130–136.

Rohlf, F. J. and Marcus, L. F. (1993) 'A revolution in morphometrics', *Trends in Ecology & Evolution*, 8(4), pp. 129–132. doi: 10.1016/0169-5347(93)90024-j.

Romice, O., Porta, S. and Feliciotti, A. (2020) *Masterplanning for change: Designing the resilient city*. RIBA Publishing.

Salat, S. (2017) 'A systemic approach of urban resilience: Power laws and urban growth patterns', *International Journal of Urban Sustainable Development*, 9(2), pp. 107–135. doi:

10.1080/19463138.2016.1277227.

Schirmer, P. M. and Axhausen, K. W. (2015) 'A multiscale classification of urban morphology', *Journal of Transport and Land Use*, 9(1), pp. 101–130. doi: 10.5198/jtlu.2015.667.

Schirmer, P. M. and Axhausen, K. W. (2019) 'A Multiscale Clustering of the Urban Morphology for Use in Quantitative Models', in D'Acci, L. (ed.) *The Mathematics of Urban Morphology*. Cham: Springer, pp. 355–382. doi: 10.1007/978-3-030-12381-9_16.

Schonlau, M. (2002) 'The clustergram: A graph for visualizing hierarchical and nonhierarchical cluster analyses', *The Stata Journal*, 2(4), pp. 391–402. doi: 10/ghh97z.

Schwarz, G. and others (1978) 'Estimating the dimension of a model', *The annals of statistics*, 6(2), pp. 461–464.

Sehra, S. S. *et al.* (2020) 'Extending Processing Toolbox for assessing the logical consistency of OpenStreetMap data', *Transactions in GIS*, 24(1), pp. 44–71. doi: 10/ggc6d5.

Serra, M., Gil, J. and Pinho, P. (2016) 'Towards an understanding of morphogenesis in metropolitan street-networks', *Environment and Planning B: Urban Analytics and City Science*, 0(0), pp. 1–22. doi: 10.1177/0265813516684136.

Serra, M., Psarra, S. and O'Brien, J. (2018) 'Social and Physical Characterization of Urban Contexts: Techniques and Methods for Quantification, Classification and Purposive Sampling', *Urban Planning*, 3(1), pp. 58–74. doi: 10.17645/up.v3i1.1269.

Seto, K. and Fragkias, M. (2005) 'Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics', *Landscape Ecology*, 20(7), pp. 871–888. doi: 10.1007/s10980-005-5238-8.

Sevtsuk, A., Kalvo, R. and Ekmekci, O. (2016) 'Pedestrian accessibility in grid layouts: The role of block, plot and street dimensions', *Urban Morphology*, 20(2), pp. 89–106.

Sevtsuk, A. and Mekonnen, M. (2012) 'Urban network analysis. A new toolbox for ArcGIS', *Revue internationale de géomatique*, 22(2), pp. 287–305. doi: 10.3166/rig.22.287-305.

Singleton, A. and Arribas-Bel, D. (2019) 'Geographic Data Science', *Geographical Analysis*, p. gean.12194. doi: 10/gfx787.

Singleton, A. D. and Longley, P. A. (2009) 'Geodemographics, visualisation, and social networks in applied geography', *Applied Geography*, 29(3), pp. 289–298. doi: 10/dg6t8r.

Smith, K. B. (2002) 'Typologies, Taxonomies, and the Benefits of Policy Classification', *Policy Studies Journal*, 30(3), pp. 379–395. doi: 10.1111/j.1541-0072.2002.tb02153.x.

Sneath, P. H. A. and Sokal, R. R. (1973) *Numerical Taxonomy*. San Francisco: Freeman.

Song, Y. and Knaap, G.-J. (2004) 'Measuring Urban Form: Is Portland Winning the War on Sprawl?', *Journal of the American Planning Association*, 70(2), pp. 210–225. doi: 10.1080/01944360408976371.

Song, Y. and Knaap, G.-J. (2007) 'Quantitative Classification of Neighbourhoods: The Neighbourhoods of New Single-family Homes in the Portland Metropolitan Area', *Journal of Urban Design*, 12(1), pp. 1–24. doi: 10.1080/13574800601072640.

Song, Y., Popkin, B. and Gordon-Larsen, P. (2013) 'A national-level analysis of neighborhood form metrics', *Landscape and Urban Planning*, 116, pp. 73–85. doi: 10.1016/j.landurbplan.2013.04.002.

Spaan, B. and Waag Society (2015) 'All buildings in Netherlands shaded by a year of construction'.

Spearman, C. (1961) 'The proof and measurement of association between two things.' doi: 10/b368ns.

Ståhle, A., Marcus, L. H. and Karlström, A. (2005) 'Place Syntax : Geographic accessibility with axial lines in GIS', in *Proceedings, Fifth international space syntax symposium*, pp. 131–144.

Steiniger, S. *et al.* (2008) 'An Approach for the Classification of Urban Building Structures Based on Discriminant Analysis Techniques', *Transactions in GIS*, 12(1), pp. 31–59. doi: 10.1111/j.1467-9671.2008.01085.x.

Stewart, I. D. and Oke, T. R. (2012) 'Local Climate Zones for Urban Temperature Studies', *Bulletin of the American Meteorological Society*, 93(12), pp. 1879–1900. doi: 10.1175/BAMS-D-11-00019.1.

Taubenböck, H. *et al.* (2020) 'Seven city types representing morphologic configurations of cities across the globe', *Cities*, 105, p. 102814. doi: 10/gg2jv4.

Thienel, I. (2013) *Städtewachstum im industrialisierungsprozess des 19. Jahrhunderts: Das berliner beispiel*. Walter de Gruyter.

Thomas, I. *et al.* (2010) 'Clustering patterns of urban built-up areas with curves of fractal scaling behaviour', *Environment and Planning B: Planning and Design*, 37(5), pp. 942–954. doi: 10.1068/b36039.

Tratalos, J. *et al.* (2007) 'Urban form, biodiversity potential and ecosystem services', *Landscape and urban planning*, 83(4), pp. 308–317. doi: 10/dqhj8n.

Tsai, Y.-H. (2005) 'Quantifying Urban Form: Compactness versus 'Sprawl' ', *Urban Studies*, 42(1), pp. 141–161. doi: 10.1080/0042098042000309748.

UN-Habitat (2020) *World Cities Report 2020*. Nairobi: United Nations Human Settlements Programme.

Usui, H. and Asami, Y. (2013) 'Estimation of Mean Lot Depth and Its Accuracy', *Journal of the City Planning Institute of Japan*, 48(3), pp. 357–362.

Usui, H. and Asami, Y. (2017) 'Size distribution of urban blocks in the Tokyo Metropolitan Region: Estimation by urban block density and road width on the basis of normative plane tessellation', *International Journal of Geographical Information Science*, 32(1), pp. 120–139. doi: 10.1080/13658816.2017.1384550.

Usui, H. and Asami, Y. (2019) 'Size Distribution of Building Lots and Density of Buildings and Road Networks: Theoretical Derivation Based on Gibrat's Law and Empirical Study of Downtown Districts in Tokyo', *International Regional Science Review*, 10, pp. 016001761982627–25. doi: 10.1177/0160017619826270.

Vanderhaegen, S. and Canters, F. (2017) 'Mapping urban form and function at city block level using spatial metrics', *Landscape and Urban Planning*, 167, pp. 399–409. doi: 10.1016/j.landurbplan.2017.05.023.

Venerandi, A. *et al.* (2016) 'Form and urban change An urban morphometric study of five gentrified neighbourhoods in London', *Environment and Planning B: Urban Analytics and City Science*, 44(6), pp. 1056–1076. doi: 10.1177/0265813516658031.

Wang, P. P. and Jernigan, T. L. (1994) 'Morphometric studies using neuroimaging', *Neurologic clinics*, 12(4), pp. 789–802. doi: 10/ghfw6v.

Ward Jr, J. H. (1963) 'Hierarchical grouping to optimize an objective function', *Journal of the American statistical association*, 58(301), pp. 236–244. doi: 10/fz95kg.

Weisberg, H. and Weisberg, H. F. (1992) *Central tendency and variability*. Sage (Sage University Paper Series on Quantitative Applications in the Social Sciences, 83).

Whitehand, J. W. (2012) 'Issues in urban morphology', *Urban Morphology*, 16(1), pp. 55–65.

Whitehand, J. W. R. (1981) 'Background to the urban morphogenetic tradition', in *The urban landscape: Historical development and management*. London: Academic Press, pp. 1–24.

Wilcox, R. R. and Keselman, H. (2003) 'Modern robust data analysis methods: Measures of central tendency.', *Psychological methods*, 8(3), p. 254. doi: 10/cvq8sw.

Williams, K. (2014) *Urban form and infrastructure: A morphological review*. Future of Cities: Working Paper. Government Office for Science.

Wurm, M., Schmitt, A. and Taubenbock, H. (2016) 'Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), pp. 1901–1912. doi: 10.1109/JSTARS.2015.2465131.

Xu, J.-h. *et al.* (2003) 'Quantitative analysis and fractal modeling on the mosaic structure of landscape in the central area of Shanghai metropolis', *Chinese Geographical Science*, 13(3), pp. 199–206. doi: 10.1007/s11769-003-0017-4.

Yan, H., Weibel, R. and Yang, B. (2007) 'A Multi-parameter Approach to Automated Building Grouping and Generalization', *GeoInformatica*, 12(1), pp. 73–89. doi: 10.1007/s10707-007-0020-5.

Yoshida, H. and Omae, M. (2005) 'An approach for analysis of urban morphology: Methods to derive morphological properties of city blocks by using an urban landscape model and their interpretations', *Computers, Environment and Urban Systems*, 29(2), pp. 223–247. doi: 10.1016/j.compenvurbsys.2004.05.008.

Yu, J., Wu, J. and Sarwat, M. (2015) 'Geospark: A cluster computing framework for processing large-scale spatial data', in *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, pp. 1–4.

Yu, W. *et al.* (2017) 'The analysis and measurement of building patterns using texton co-occurrence matrices', 31(6), pp. 1079–1100. doi: 10.1080/13658816.2016.1265121.

Zelditch, M. L., Lundrigan, B. L. and Garland Jr, T. (2004) 'Developmental regulation of skull morphology. I. Ontogenetic dynamics of variance', *Evolution & development*, 6(3), pp. 194–206. doi: 10/fkzt4w.