# 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.

## 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 [@dogrusoz2007] to comprehensive small scale studies based on detailed manual digitisation of urban form elements [@dibble2017], 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 @dibble2017 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 [@moudon1997], 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.

## 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 @dibble2017.

### 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.

### 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.

#### 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.

#### 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.

## momepy: Urban Morphology Measuring Toolkit

*The content of this section was partially published in @fleischmann2019a.*

Urban morphology is based on the analysis of space, traditionally mostly visual and qualitative [@dibble2015]; its objects are the fundamental elements of urban form (building, plot, street) [@moudon1997] as well as a range of analytical constructs such as axial maps [@ariza-villaverde2013] or proximity bands [@araldi2019]. 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 [@dibble2017]. 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 [@sevtsuk2012] or Place Syntax Tool [@stahle2005]. 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 [@kelsey\_jordahl\_2020\_3946761], using its GeoSeries and GeoDataFrame objects to store and handle large amounts of geospatial information. Under the hood uses PySAL [@doi:10.1080/17421772.2019.1593495; @rey2007], mostly taking care of spatial weights matrices capturing the adjacency of elements of urban form. The graph module uses the capabilities of networkX [@hagberg2008] 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 [@mottelson2020fine]. 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](https://docs.momepy.org).

## 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 ), which still extends the built-up area of the city by a large margin. That ensures the minimisation of the edge effect [@gil2016] 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.

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, Amsterdam 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 ). The data are obtained from @dukai2020 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 @krenz2018.

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 [@fleischmann2020] prior specification of Prague and Amsterdam as case studies for numerical taxonomy as a boundary is used administrative area of Zurich (figure ). 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.

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

## 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.