# 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.

## 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.

### 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 @araldi2019. Both methods measure primary characters on selected elements (street segment in the case of @araldi2019), then include contextualisation layer (LISA/ILINCS patterns in the case of @araldi2019), 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 @araldi2019). Similarly to @dibble2017, @dong2019, @li2020 or @serra2018a, 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 @berghauserpont2019a, 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 @berghauserpont2019a, 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 @berghauserpont2019a 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.

### 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.

### 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 .

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 . 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 b), 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).

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.

### 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 @dibble2017 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 @alexander1966.

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 [@serra2018a; @dibble2017; @jochem2020].

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 [@oliveira2016]? 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* [@kropf2017], 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.

## 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 [@bingham2019pyro; @matthew\_rocklin-proc-scipy-2015], 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).

## 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 [@feliciotti2018] or other performance-based indicators (e.g. sustainability [@bourdic2012]), 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.

## 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 current 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 [@schonlau2002clustergram]) 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 foundations 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.

## 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.