THE URBAN ATLAS: TOWARDS A MORPHOMETRIC TAXONOMY OF URBAN FORM

Martin Fleischmann, PhD Candidate, University of Strathclyde; Research Associate, University of Liverpool, UK

Ombretta Romice, Senior Lecturer in Urban Design, University of Strathclyde, UK

Sergio Porta, Professor of Urban Design, University of Strathclyde, UK

ABSTRACT

A key aim of urban morphology is to understand the different morphological patterns characterizing cities and their part. However, the discipline is still lacking quantitative methods of pattern detection rich in information and scalable (applicable up to regional and national scales).

We present progress towards this goal, through an urban morphometrics approach: the quantitative, systematic and comprehensive analysis of urban form. Linking numerical taxonomy (evolutionary biology) and urban morphology we develop a procedure which identifies distinct homogenous patterns in cities based on spatial components only. This procedure allows large-scale analysis which is a) unsupervised, machine-learning, and b) requires minimal input-data requirements, i.e. street network and buildings (footprint and height) only, while generating an abundant numerical characterization of urban patterns. From these two sets of entries we: measure a comprehensive set of 360 unique morphometric descriptors of urban form elements' relationships in space; identify homogenous clusters (taxa) of distinct urban tissues on the ground; determine their phenetic similarity. This constitutes a taxonomy of urban form organized in a phenetic tree, a format which other fields, notably evolutionary biology, use to precisely describe their object of research.

We demonstrate the cross-cultural validity of the method through application to different spatial contexts.

Keywords: urban morphometrics, classification, taxonomy, urban tissue, machine learning.

INTRODUCTION

One of the main aims of urban morphology is to describe patterns of urban form and attempt to classify them into a coherent systematics, to reduce the inherent complexity of urban form into manageable types, enable the identification of similarities and differences between types and, capture the structural relationships between them.

Our effort sits in a rich tradition of urban morphology studies aimed to identify and classify components of the urban form along with their functional, geographical or even phenetic relationship. For example, Karl Kropf has proposed a general terminological framework along the Conzenian tradition and beyond (Kropf 1993, 2014), while Caniggia and Maffei (2001; c. 1979) have offered a speculative kernel for a general theory of evolution in urban form developing the Muratorian tradition. Along the same lines, we set out to progress towards a general classification of urban form by taking full advantage of digitization and machine learning (ML) techniques and processes, in order to: 1. Establish an unsupervised process of recognition of both the identity of homogeneous areas in cities (taxa) and their hierarchical relationship of morphological similarity (numerical taxonomy of urban form); 2. Scale-up the ground of evidence on which the classification is based, by designing a method that is fit for the automated analysis of the urban fabric at regional and national scale of extent.

CLASSIFICATION IN LITERATURE

Existing literature offers a variety of methods aiming to provide a classification of urban form. From an urban morphometrics point of view, we are still missing an *optimal* method of analysis, which is *exhaustive*, *empirical*, *hierarchical*, *comprehensive*, *detailed* and *scalable*, and uses classes that are, ideally, *mutually exclusive* (derived from Bailey (1994)). This method should also be reproducible, scale-dependent flexible and universal, describing small- and large-scale case studies using the same set of tools. While the extent of this paper does not allow for a full overview of existing attempts in this direction, we scrutinize two recently published classification methods, which are the closest to achieving such *optimal model*: 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; Berghauser Pont, Stavroulaki and Marcus, 2019; Bobkova, Berghauser Pont and Marcus, 2019) and the Multiple Fabric Assessment proposed by Araldi and Fusco (2019, 2017).

The Chalmers School proposes to use three individual typologies of morphological elements: plots, streets and buildings. Each typology is defined through a handful of morphometric characters, thus making the outputs influenced by this particular selection. Compared to the optimal criteria above, their model is not hierarchical and, importantly, not comprehensive (due to the limited number of morphometric characters it uses).

The work of Araldi and Fusco (2019, 2017) 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 greater 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 the whole segment, which is rarely the case in urban contexts of almost all periods.

This research proposes a new method of hierarchical classification of urban form which fulfils all the criteria of the *optimal model* above, implementing the principles of numerical taxonomy and detailed morphometrics.

METHOD

The method consists of three major parts:

1. Morphometric characterization of urban form
2. Spatial identification of distinct homogenous clusters (DHCs) of urban form, representing taxa of urban tissues.
3. Hierarchical classification of resulting taxa to generate a taxonomy of urban form.

The process behind the method builds on the theory of numerical taxonomy established in the second half of the 20th century in biology (Sneath and Sokal, 1973) for the description and classification of species. While biology itself now prefers DNA sequencing for the same purpose, urban morphology can build on the established theory based on the quantification of shape, i.e. morphometrics.

The morphometric characterization is done based on two input vector data layers. The first is a polygon representation of building footprints with the attribute of building height at the Level of Detail (LoD) 1 (Biljecki, Ledoux and Stoter, 2016). The second is the centerline representation of a street network, cleared of transport planning related structures (e.g. roundabouts, dual carriageways). These are used to generate additional morphometric elements, namely morphological tessellation (Fleischmann *et al.*, 2020) and tessellation-based blocks (Fleischmann, 2019).

The spatial patterns made of morphometric elements are measured to capture their structural complexity represented by six categories of morphometric characters (Fleischmann, Romice and Porta, 2020) - dimension, shape, spatial distribution, intensity, connectivity, diversity, and to capture cross-scale complexity represented by a range of topological relations of selected elements. That means that characters capturing the form of individual elements (e.g. *elongation of a building*) are captured alongside the characters capturing the spatial distribution of elements within the building's broader topological context (e.g. *mean inter-building distance*). Moreover, characters are linked to the level of individual building either directly or based on the proximity between morphometric elements.

This research implements 74 primary characters, both extracted from literature and newly developed, which capture individual features of urban form elements and their fundamental relations. Their spatial distribution could be abrupt and do not express patterns . Hence the characters defined as primary have to be expressed using contextual, spatially lagged versions for pattern detection. The context here is defined as a neighborhood of each building within three topological steps on morphological tessellation. Four types of contextual characters are proposed. One captures a local central tendency (*interquartile mean*) and three capture the properties of the distribution of values within the context (*interquartile range*, *interdecile Theil index of inequality*, *Simpson's diversity index*). For each of the primary characters, each of the contextual is calculated and used within the clustering algorithm. The resulting set of used characters is composed of 4 x 74 = 296 contextual characters.

The second step of the method is the ML model of identification of urban tissue taxa using the Gaussian Mixture Model (GMM)..

The ideal outcome of DHC recognition is to find each cluster as a distinct taxon of urban tissue. However, the definition of urban tissue (Kropf, 2017) 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 applied here. The ideal outcome of clustering is the optimal number of clusters based on the actual structure of the observed data. GMM clustering requires specification of a number of components of the model (i.e., clusters). However, that number is not known in the case of urban form. The way around this problem is to estimate the number of components based on the goodness of fit (Bayesian Information Criterion (BIC) is used) of the model. 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 are clusters distinguished).

The last part of the method is the determination of the phenetic relations between distinct taxa, i.e. their morphological similarity. We use a form of Ward's hierarchical clustering aimed to develop a hierarchy of similarities between observations based on their morphometric profiles (using a centroid of each cluster). We use this procedure to generate a tree-like classification of urban form, which lays the foundation for the taxonomy of urban tissues.

Finally, the method proposed is tested in the case of two historical heterogeneous cities, Prague, CZ and Amsterdam, NL, and in the case of the suburban city of Grand Rapids, MI, to test its universality and cross-cultural validity.

RESULTS.

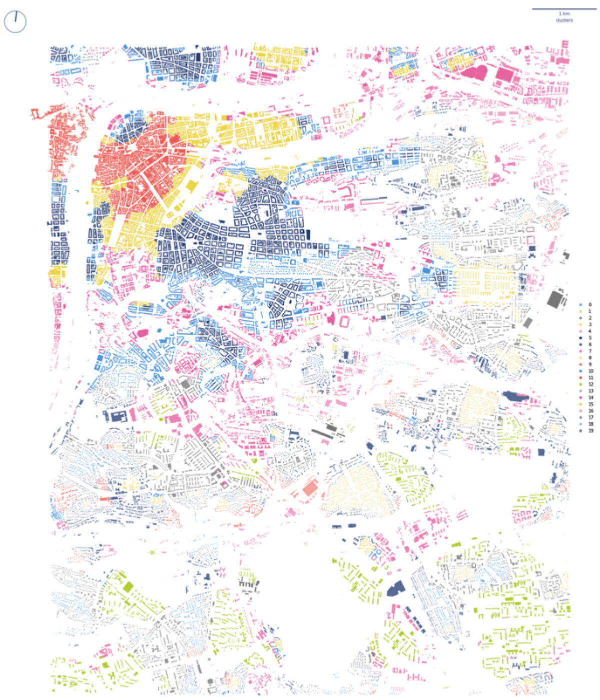

Figure 2. Distinct homogenous clusters identified within city center area of Prague. Each color represents a single cluster.

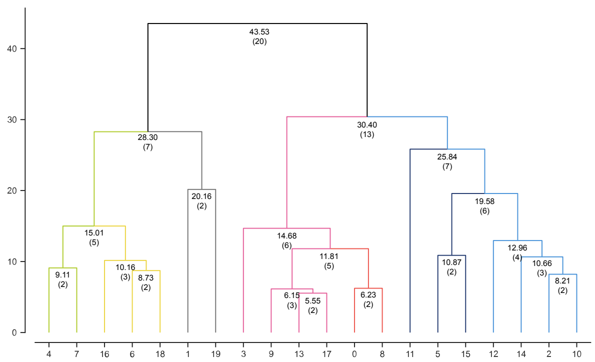

Figure 1. Dendrogram illustrating relationship between individual clusters (x-axis) identified in Prague and their cophenetic distance (y-axis).

The first application of the proposed recognition method is the case study of Prague, where clustering is applied within its administrative boundary, ensuring that the edge effect caused by street network cutting (Gil, 2016) is minimised.

Based on the BIC results, GMM clustering identifies 20 clusters. The clusters are well defined and reflect a homogenous form. Even though there is no spatial constraint imposed in the clustering itself, results show an apparent contiguity derived from the contextual characters. Figure 1 shows detail of the section of Prague covering the City Centre and the area towards the southern boundary for a better understanding of results.

Starting from the top left corner, which represents the historical core of Prague, 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 a pure observation, DHCs seems to be very precise and detailed and, most importantly, meaningful in terms of their link to the concept of the urban tissue type.

The centroid values of each DHC, obtained as a mean value of each morphometric character,

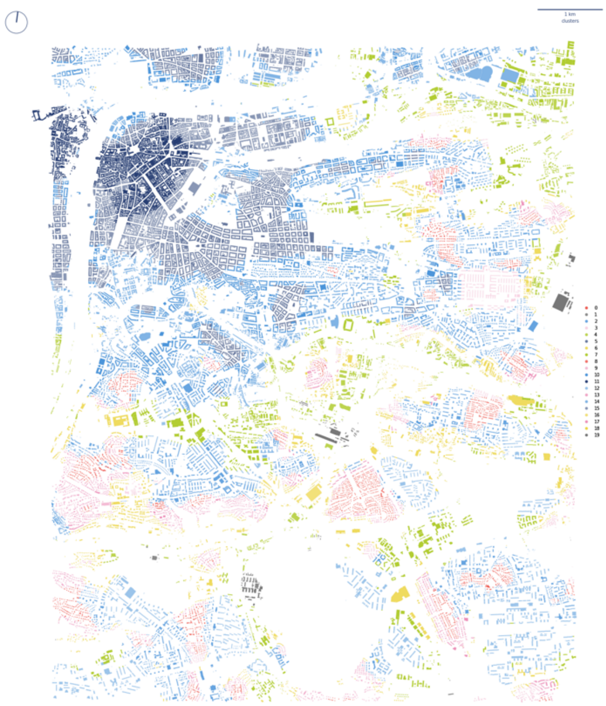

Figure 3. Hierarchically represented distinct homogenous clusters in the area of Prague's city center with colors representing the phenetic relation captured in figure 2.

Figure 4. Hierarchically represented distinct homogenous clusters in the metropolitan area of Amsterdam with colors representing the phenetic relation captured in embedded dendrogram.

are used as taxonomic characters within Ward's hierarchical clustering. The resulting relationship between centroids, representing the relationship between identified urban tissue taxa, is illustrated on the dendrogram on figure 2. The horizontal axis represents each 20 DHCs (or types of urban form) detected in Prague, while 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.

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 clusters in the detail of the city centre is shown in figure 3.

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 types. The right side of the tree represents what we could call "the organised city". It consists of areas of mixed origin, spanning from the historical core to modernist and contemporary developments. On the other side lies "the unorganised city". It contains both industrial and fringe areas as well as contemporary

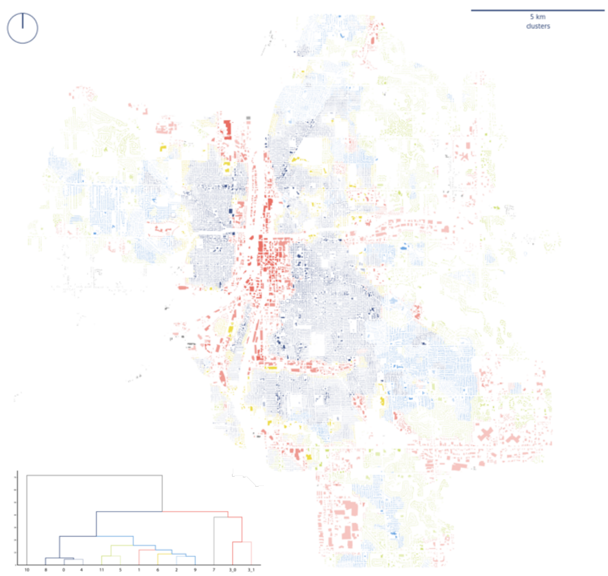

Figure 5. Hierarchically represented distinct homogenous clusters in the metropolitan area of Grand Rapid with colors representing the interpreted phenetic relation.

office parks. The yellow branch is "heterogenous unorganised", but it may as well be linked to low density. The disorder here is the driver.

After validating the method in the city of Prague, we want to test the degree to which the same is transferable to other contexts. This is critical to the overarching obective of this morphometric research, which is designed with universality in mind. The hypothesis is that the method should be able to capture a similar level of information in other geographic and historical contexts as well. As a test of this hypothesis, we applied the same procedure to the case studies of Amsterdam (NL) and Grand Rapids (MI).

The spatial distribution of clusters identified in both cases shows potentially meaningful clusters. From this first perspective, it looks promising, and there is no reason to think that the method identification of clusters is not transferable. The tree of Amsterdam urban form shows similar characteristics as we have seen in Prague, with the significant bifurcation into two branches and then consequent bifurcations lower in the tree distinguishing different rules of the organisation. The tree of Grand Rapids, modern US city, is characterised by a different structure, which in this case reflects very different patterns of historical development.

Results from all cases show how the spatial distribution of taxonomic branches can be mapped to initial clusters to allow hierarchical reading of city structure and comparison of macro patterns of development across the different cities.

SYNTHESIS

The method tends to identify homogenous clusters which could be interpreted as urban tissues. However, the relation of the distinct homogenous cluster to urban tissue is dependent on the definition of the urban tissue. While the definition used by Kropf (2017), adopted within this study, is based on internal homogeneity and contiguity, we can say that each *contiguous part* of each distinct homogenous cluster is a single urban tissue.

The proposed hierarchical classification is paving the way towards the development of a comprehensive taxonomy of urban form. As shown in the context of three case studies, it fulfils all requirements of the *optimal model* of urban form classification. The significant advantage of the method is the data richness derived from the used set of morphometric characters, allowing detailed morphometric profiling of each tissue, each cluster, or each branch, and its hierarchical flexibility.

The main limitation of the method is the availability of input data at a specific level of resolution. The input data for Prague and Amsterdam were close to the optimal situation, while building footprints for Grand Rapids do not distinguish individual buildings in configurations of adjacent buildings, so the model is adapted accordingly. Even though this article shows results based on two different resolutions of input data and both result in meaningful classification, the data limitations are still the crucial constraint of the method's applicability.

Urban morphology may benefit from urban morphometrics attempts to identify urban tissues and classify them. The proposed method offers a way to accomplish a quantitative, systematic and comprehensive analysis of urban form, based on both richness of information and extra-large scale of geographical coverage. The results show that the method is valid and applies to different morphological contexts. We believe that urban morphometrics opens a new avenue of research in urban morphology. It builds on and expands the core-traditions of the discipline, and its potential is still largely unexplored.

REFERENCES

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.

Bailey, K.D., (1994). Typologies and Taxonomies - An Introduction to Classification Techniques. Sage Publications.

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. doi: 10/gf8x8j.

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*. *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., 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.

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.

Caniggia, G. and Maffei, GL. (2001, c.1979) *Architectural Composition and Building Typology: Interpreting Basic Building,* Firenze, IT: Alinea

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.

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.

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.

Kropf, K. (1993) *The definition of built form in urban morphology*. University of Birmingham.

Kropf, K. (2014) *Ambiguity in the definition of built form*. Urban Morphology 18(1): 41-57.

Kropf, K. (2017) *The handbook of urban morphology*. Chichester: John Wiley & Sons.

Sneath, P. H. A. and Sokal, R. R. (1973) *Numerical Taxonomy*. San Francisco.

CORRESPONDING AUTHOR

Martin Fleischmann, Research Associate / Data Scientist, Geographic Data Science Lab, University of Liverpool, Roxby Building, Liverpool, L69 7ZT, United Kingdom. m.fleischmann@liverpool.ac.uk @martinfleis (twitter)