# 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 @dibble2017, who tried to provide the first link between the numerical taxonomy and urban morphology.

## Learning from biology

From a conceptual point of view, this research builds on the previous of @dibble2017 (partially published as @dibble2015 and @dibble2016), 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 @dibble2017, 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.

### Morphometrics as a descriptive tool

Morphometrics is a quantitative analysis of form, i.e. size and shape [@rohlf1993] 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 [@sneath1973]‌. 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*” [@rohlf1993, p.129]. The variables used within the analysis usually represent widths, lengths and angles of features or distances between specific landmarks [@rohlf1993], hence containing little information about the geometry of the structure [@zelditch2004developmental]. The recreation of the original shape is not possible based on the used set of variables [@adams2004geometric]. Among the statistical techniques used within traditional morphometrics @rohlf1993 list principal component analysis, discriminant functions of canonical variate analysis. An example of measurements used within traditional morphometrics is shown in figure .

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 [@breno2011].

The second major branch of morphometrics, which tries to overcome some limitations of the previous one is geometric morphometrics [@rohlf1993], focusing on the position of landmarks and semi-landmarks on the grid and its deformation (figure ). The description of both size and shape is then captured through the series of coordinates either on a 2D plane or in 3D space.

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 [@batty1987] or different aspects of streetscape [@cooper2003; @cooper2005]. Furthermore, there are other branches of morphometrics (tissue morphometrics [@bodenstein1987growth], geomorphometrics [@coblentz2014quantitative], neuroimaging [@wang1994morphometric]) which are applying the same principles in their respective fields.

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

### 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 @sneath1973 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*” [@sneath1973, 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 @sneath1973, “*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.

@sneath1973 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.

1. *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.

1. *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.

1. *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.

1. *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.

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

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

1. *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.

#### 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. @sneath1973 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.

#### 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*” [@sneath1973, 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 @sneath1973 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.

#### 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. ). However, the methods of cluster analysis vary, even though the tendency is to employ hierarchical methods.

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*” [@sneath1973, p.199]. If the OTU in place will be an aggregation as is Sanctuary area in case of @dibble2017 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.

### Criticism of numerical taxonomy

As any other classification, numerical taxonomy has its drawbacks and limitations, which should be carefully considered before its application. @sneath1973 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.

## 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.

### 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 [@schirmer2015] to network-based characters [@porta2010] and beyond [@araldi2019]. 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 @carneiro2010 and then further elaborated by @dibble2017, 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.

### 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* [@cambridgedictionary]
* *a single organism capable of independent existence* [@dictionary.com]
* *a member of a compound organism or colony* [@dictionary.com]
* *a particular being or thing as distinguished from a class, species, or collection* [@merriam-webster.com]
* *a single organism as distinguished from a group* [@merriam-webster.com]
* *a single, separate organism (animal or plant) distinguished from others of a same kind* [@biology-online.org]

Morphological literature used various elements as the unit of analysis, from a building [@schirmer2015] or plot [@bobkova2017] to sanctuary area @dibble2017 or neighbourhood [@song2007]. 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 [@moudon1997] 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 . 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 [@de1998general]. 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 [@sneath1973, 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.

### 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 @dibble2017. 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.

#### Sanctuary area as OTU

As an OTU, @dibble2017 use Sanctuary area (SA), defined as the portion of the urban fabric enclosed by main streets [@mehaffy2010]. 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. The whole fabric of the town is composed of few SAs. The working hypothesis @dibble2017 use is that it is composed of several classes of urban patterns following the division of urban fabric into SAs as shown on figure .

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 ) 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 @dibble2017).

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” [@mehaffy2010]. 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.

#### Selection of taxonomic characters

While selecting morphometric characters used for classification, one should avoid the empirical correlation of resulting values [@sneath1973]. 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 @dibble2017, 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.

## 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 @dibble2017 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.