This is the title of the thesis

Martin Fleischmann

A thesis presented for the degree of
Doctor of Philosophy

Supervised by:
Dr Ombretta Romice
Professor Sergio Porta

Urban Design Studies Unit
Department of Architecture
University of Strathclyde, UK
Month 2020

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

Signed:

Date:

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam et turpis gravida, lacinia ante sit amet, sollicitudin erat. Aliquam efficitur vehicula leo sed condimentum. Phasellus lobortis eros vitae rutrum egestas. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Donec at urna imperdiet, vulputate orci eu, sollicitudin leo. Donec nec dui sagittis, malesuada erat eget, vulputate tellus. Nam ullamcorper efficitur iaculis. Mauris eu vehicula nibh. In lectus turpis, tempor at felis a, egestas fermentum massa.

Acknowledgements

Interdum et malesuada fames ac ante ipsum primis in faucibus. Aliquam congue fermentum ante, semper porta nisl consectetur ut. Duis ornare sit amet dui ac faucibus. Phasellus ullamcorper leo vitae arcu ultricies cursus. Duis tristique lacus eget metus bibendum, at dapibus ante malesuada. In dictum nulla nec porta varius. Fusce et elit eget sapien fringilla maximus in sit amet dui.

Mauris eget blandit nisi, faucibus imperdiet odio. Suspendisse blandit dolor sed tellus venenatis, venenatis fringilla turpis pretium. Donec pharetra arcu vitae euismod tincidunt. Morbi ut turpis volutpat, ultrices felis non, finibus justo. Proin convallis accumsan sem ac vulputate. Sed rhoncus ipsum eu urna placerat, sed rhoncus erat facilisis. Praesent vitae vestibulum dui. Proin interdum tellus ac velit varius, sed finibus turpis placerat.

Table of Contents

Table of Contents

List of figures

List of tables

Abbreviations

API	**A**pplication **P**rogramming **I**nterface
JSON	**J**ava**S**cript **O**bject **N**otation

Chapter 1

Introduction

Chapter 2

Existing approaches to classification of urban form

6 000 words (if less, better)

2.1 Introduction

- Explain prior focus on quantitative morphology (link to introduction), but say that the chapters gives overview of all, with the focus on quantitative.

2.2 The need for the classification

- *Why is classification important, what can it bring to the table, why should we bother doing it.*
- What is classification
  - a bit of definitions
  - different ways of making classification
    - typology/taxonomy distinction **important**
- Why is classification useful in general

- Why is classification useful in urban morphology

2.3 Existing methods of classification of urban form

- *Literature review of existing methods of classification and its analysis and description of patterns within the field.*
- Introduction

2.3.1 The history of classification attempts

- *A brief overview of the history of classification of urban form focusing on its origins and early attempts. People like Lynch, Kostof.*
- **Research TO DO**
- link between history and qualitative

2.3.2 Qualitative

- Traditional schools of urban morphology
  - Conzen
  - Muratori
  - Duany
- City-based approaches (Portland, Berlin, Prague)
- Spatial typology
  - Kohout, a+t
- The qualities of such approaches, their limits.
  - **Research TO DO (a bit)**
  - expert knowledge needed
  - concepts based
  - might be biased (not necessarily)

  - good in interpretation, could be detailed
  - time consuming, information demanding
  - limited applicability

2.3.3 Mixed (predominantly non-morphological)

- Socio-demography as a main branch
- Additional (energy)
- The qualities of such approaches, their limits
  - capturing non-morphological classes
  - good for specific purposes
  - good source for link between form and soft data
- **Research TO DO (a bit)**

2.3.4 Quantitative

- introduction
  - what does it mean quantitative method
  - two major groups divided by the data source
    * remote sensing — raster data
    * morphometrics — vector data
      · *morphometrics can in theory be done on remote sensing as well, so it might be better to use another term*

2.3.4.1 Remote sensing

- Introduce RS
  - satellite or aerial data, automatic (multi-spectral) image recognition, supervised ML
- Units of analysis

  - patch
  - block
  - grid
  - *add some figures as an illustration*

- Number of categories
  - 1 - 10

- The qualities of such approaches, their limits
  - possible extent
  - only "visible" spectrum - roofs can make a lot of difference in RS but minimal in reality
  - mostly supervised nature - you have to predefine ground truth
  - the aspect of resolution and data availability
  - number of categories is generally low related to low number of actual indicators (like Copernicus)

2.3.4.2 Urban Morphology (quantitative)

- *This is the key focus of the whole chapter, and the majority of scrutinised works fall into this category. The rest mentioned above and below is to draw a full picture, but it does not aim to provide an in-depth understanding, unlike this part.*
- **Research TO DO - check recent papers, some might be included**
- Introduce quantitative morphology
- units of classification
  - *Assessment based on the unit of classification and its placement on the scale.*
  - gradient of scales
  - from city scale to building and plot
  - *do some quantitative assessment of the db*

- number of classes

  - generally low, in few cases higher
  - *do some quantitative assessment of the db*

- mention number of characters used for classification (scrutinised in the next chapter)
- Synthesis of the corpus of works
  - *taxonomic relations between types?*
  - The qualities of such approaches, their limits

2.4 The gap in the systematic classification

- lack of systematic classification based on the small-scale unit
- gap in unsupervised classification
- gap in detailed classification (i.e. number of classes)
- gap in exploration of relationships between classes (*check before writing*)

2.5 Conclusion

- *conclusion: the existing approaches and methods have gaps: the lack of systematic classification based on the small-scale units ~~using an extensive, inclusive set of indicators~~ enabling detailed classification into larger number of types/taxa/classes. That should help position my work within the field and say what I am bringing new in later stages. BE CAREFUL TO CONCLUDE ONLY BASED ON THE CONTENT OF THE CHAPTER NOT MORE. FIND A GAP WHICH MAKES SENSE. THIS TEXT IS MIXING TOGETHER RESULTS OF THIS AND THE NEXT CHAPTER. THIS LOOKS AT THE UNIT AND NUMBER OF CLASSES MOSTLY. NUMBER OF CHARACTERS SHOULD BE LEFT TO THE NEXT CHAPTER.*

Chapter 3

Measuring of urban form

- The need for measuring
- Based mostly on my MSc
- How others measured form?
- Where is the gap?

Chapter 4

Evolution and urban form

- Evolutionary perspective in the context of urban design
  - Biological as well as cultural
  - Taxonomy
- Explain the principles of evolution in the context of urban morphology
- deifne viable analogies
- use cultural evolution alongside with biological
- introduce evolutionary approach to classification - cladistics, taxonomy
- current views of evolution and cities (e.g. Marshall)

Chapter 5

Propositions

Chapter 6

Urban tissue as an individual

- ~~Identification of "individual" (OTU)~~
  - ~~Urban Tissue~~
- ~~Exploration of concept of DHC~~
- What is urban tissue
- Why is it worth studying
- How others approached it
- City as an ecosystem
- What is an individual within this ecosystem
- Principles of identification of individuals
- Introduction of DHC (as a theoretical concept)

6.1 What is the *individual* in the urban form?

- We are trying to identify and describe distinct kinds of urban form. *Individuals* forming the population - the city.
- Dibble et al. used Sanctuary Areas (SA), I argue that the concept of SA is limited. By qualitative definition of SA, and by possible heterogeneity of it.

  - The problem with SAs is that their definition and identification is *phylogenic* process, it is based on the process of development of the settlement. The rest of the taxonomic systematisation is, however, based on purely *phenetic* attributes.
  - Use of SA as OTU assumes that whole cities are in fact ideal according to 'Emergent Neighbourhood Model' (Mehaffy et al). Even the authors states that they are not (e.g. the three pathologies). While the concept of SA in this model perfectly works, in case of unrestricted taxonomy it doesn't.

- We are looking for the structure indicating the smallest distinct kind (Sneath and Sokal) of urban form, which urban morphologists define as urban tissue (Kropf)

6.2 Urban Tissue and similar concepts

- Urban tissue has several definitions
  - principal unit of growth
  - the ensemble of aggregated buildings, spaces and access routes (Cannigian analysis)(Samuels, 1982, p.3)
  - a distinct area of a settlement in all three dimensions, characterised by a unique combination of streets, blocks/plot series, plots, buildings, structures and materials and usually the result of a distinct process of formation at a particular time or period (Kropf, 2017)
- Urban morphologists are using a few concepts which are very similar. Those are **urban tissue**, conzenian **plan unit**, cannigian **tessuto urbano**, and **urban structural unit**. However, there are minor differences.
  - We can say, that **urban tissue** is a broad theoretical concept, which is defined above.
  - All the other, are methodological terms capturing urban tissue in different ways. Conzenian plan unit is based on qualitative analysis of two-dimensional town plans, 'tessuto urbano' by Muratori, Cannigia

and Maffei uses the principle of aggregation of smaller hierarchical elements (also qualitative approach), urban structural unit, originating in studies of metabolism of urban systems (Pauliet and Duhme, 1998 + some others) is mixed-use method incorporating, beside the built form, also structure of open spaces (Osmond).

- I am proposing another concept of capturing the urban tissue.
  * **the smallest distinct physiognomically homogenous cluster** (DHC)
    · In short, DHC is formed by clustering method based on measurable characters.
  * Unlike methods described above, DHC is purely quantitative one

One of the result of the research should therefore be the taxonomy of urban tissues (defined as DHC).

~~###### 05.x.x Complexity (of urban form)~~

Chapter 7

Identification of urban tissues through urban morphometrics

7.1 Theory: The relational model of urban form

7.2 Principles of systematic morphometric description

7.3 Methodological proposition

7.3.1 Selection of morphometric characters

7.3.1.1 Diversity as a statistical dispersion

7.3.1.1.1 Diversity in urban morphology

- *reiterate the issue of capturing diversity (ch3, EPB paper)*
- *importance of diversity in urban environment (Jacobs)*

7.3.1.1.2 Existing methods of measuring diversity

- *a bit of introduction here*
- *check other methods in the table*
- *mention Dibble's IQR*

to be edited, text below comes from EPB paper (unused) - expand description of methods

The frequent use of the Simpson's diversity index, for example, could be worrying, as it relies on pre-defined "bins" (classes of values). For example, Bobkova, Marcus and Berghauser Pont (2017) use this index to measure the diversity of plot sizes, but their binning into intervals based on the actual case-specific values makes the outcomes hardly comparable to any other: if we apply the same formula to another place, we will get different binning. This appears to be a rather ubiquitous problem in applying the Simpson's diversity index, i.e. it is necessary to set a finite set of pre-established bins prior to undertaking the analysis, however despite the need for urban morphology analysis to produce comparable outcomes, it is difficult to ensure specific descriptiveness to "universal" predefined bins. The use of the Simpson's diversity index in ecology is encouraged (Jost, 2006) because ecologists have a finite number of groups enabling them to pre-define all bins appropriately (moreover, bins are usually not defined on a continuous numerical scale), however this is not often the case in urban morphology.

Recent literature shows that we now have alternative ways to measure the diversity of morphological characters. Caruso, Hilal and Thomas (2017) applied the Local Index of Spatial Autocorrelation (LISA) in a form of local Moran's I, defined as "the weighted product of the difference to the mean of the value of a variable at a certain observation and the same difference for all other observations, with more weight given to the observations in close spatial proximity. (Caruso et al., 2017: 84) LISA aims to identify clusters of similar values in space, describing their similarity or dissimilarity, which could be seen as a proxy for diversity.

Another approach grounds the character on the statistical distribution of all measured values and compares it to the ideal distribution. One example is a test whether such distribution follows the principle of the Power Law used by Salat (2017). Another is an application of the Gini index initially used to measure in-

equality. In the case of diversity, the more unequal the distribution is, the more diverse. Since none of these measurements requires pre-defined grouping, they resolve the problem of binning highlighted above with reference to the Simpson's diversity index.

- *conclude literature with the clear gap*
  - *simpson problematic*
  - *LISA reductionalist*
  - *Power law - distribution often unknown*
  - *IQR simplistic*

7.3.1.1.3 Diversity of continuous variables

- *focus on diversity of measurable characters - continuous variables*
- *diversity as statistical dispersion*
- *three types*
  - *dimensional*
    - *list them and make a short description*
  - *dimensionless*
    - *list them and make a short description*
  - *binned*
    - *list them and make a short description*
    - *add more than simple Simpson's?*
- *robustness of measures*
  - *discuss the issue of robustness*

7.3.1.1.4 Method of comparison

- *select large sample*
- *select easy-to-understand characters*

- *measure diversity of each character using every method*
- *normalise results*
- *visual assessment*
- *correllation*
  - *Pearson*
  - *Spearman?*
  - *Scatter plots*

7.3.1.1.5 Results

- *present raw results*

7.3.1.1.6 Interpretation

- *selection of the most appropriate per each type*
  - *dimensional*
  - *dimensionless*
  - *binned*
- *synthesis of all*
- *discuss diversity in homogeneity and order (many small units..) vs measurable diversity*
- *conclude*

7.3.1.1.7 NOTES ../../00_Core/06_Identification of urban tissues/03_Characters tests/Diversity as statistical dispersion/Diversity as statistical dispersion.md

7.4 MIXED CHAPTER NOTES

~~* Scales~~ ~~* Measurable elements~~ ~~* Characters~~ ~~* Measuring framework~~ ~~* Test~~ ~~* Validation~~

7.5 What needs to be captured 190110

To identify urban tissue within urban fabric we have to capture complex description of each tissue. It has to be inclusive in terms of scales, because city is composed of nested complexities as well as categories of characters. [bullshit]

7.5.0.0.0.1 Strategy for character selection Set of taxonomic characters should:

1a. **Capture structural complexity by covering all categories of characters** - dimensions - shapes - spatial distributions - intensity - connectivity - diversity 1b. **Across all fundamental elements** - building - street - cell 2. **Capture scalar complexity by covering all topological scales** - single/small - medium - large - extralarge

I. To fulfil these conditions, we have to first **define subsets as combinations of all three rules listed above**. II. Then we have to **identify characters for each subset** by following selection process:

1. **Extract all characters found in literature** (477 in total). This database worked as the main source of taxonomic characters. Due to its extent, it is expected that the majority of possible characters is included.
2. **Select those characters using data intended to be used within each subset** . ~~Large portion of characters identified in the first steps is using data sources which are not widely available. Because this research is aiming to be as universal as possible, it is using only two data layers. The characters relying on other inputs were excluded from the selection.~~
3. **Adapt characters to fit the framework** Some characters needs to be adapted to fit the framework, mostly as a translation of alto-based characters into cell-based and metric-based into topology-based.
4. **Eliminate logical correlations** (full correlations as well as partial, where possible). Logically correlated characters should be omitted, otherwise the feature which is causing the correlation would distort the resulting taxonomy and clustering. Fully correlated characters (in terms of causality) have

to be excluded and only one should be kept, partial logical correlations should be limited, but in case of urban form, we are not able to fully avoid them.

5. **Eliminate ineffective characters** (axial maps, topological skeleton). Due to the nature of the analysis, working with large-scale data or even big data in some cases, the process of measuring has to be computationally effective. Some of the characters are not easily measurable, and it has to be evaluated whether the value of the characters would balance the difficulty of coding and / or computational demand.
6. **Add characters where are clear gaps** (diversity, plot-level Voronoi cell). Because I am using Voronoi cells the smallest scale spatial unit in a scope previously unused, there is a range of characters which had to be adapted from original plot-based to cell-based. The database of characters also showed imbalance of different categories and gaps in the measuring of diversity. New taxonomic characters have to be implemented to cover those gaps and provide coherent description of urban form. This part of the research is still ongoing.
7. **Exclude invariant characters**. Some characters might be invariant over the entire sample of OTU's. Those should not be included as they are not bearing any taxonomic value. However, this exclusion is an ongoing process, because it depends on actual measured values.
8. **Limit empirical correlation*** When we have the evidence that more than one factor affects two correlated characters within a study, regardless of whether this evidence comes from within study or from outside, we would include both characters; otherwise we would employ only one. We assume that at least some independent sources of the variation in any empirical correlation, unless we have reason to believe otherwise.
9. **Exclude characters which does not have the ability to capture patterns**. Test capability of each character to capture spatial patterns by measuring spatial autocorrelation. Those without sufficient level of autocorrelation should be excluded as they do not bear any value in the process of identification of DHC.
10. **Balance scalarity and uniqueness of values**. The set of taxonomic

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

(Based on Sneath and Sokal, 1973)

7.5.0.0.0.2 Notes based on Jacob's thesis 190117 He stated "These characters of urban form, at this point, are only hypothesised to be valid. Although certain characters utilised in this Methodology are quite simple and relate to non-abstract concepts, such as area and length, there is no preconception of the relative importance of the characteristics of urban form." - lets make a general test of reliability of characters via spatial autocorrelation. Characters are valid and valuable when they capture some patterns, hence spatial autocorrelation has to be present. It can work as a validation (of capability not completeness) of set of characters.

7.5.0.0.0.3 Scales for categorisation of characters for definition of DHC Scales are selected to capture characters of urban tissues on all relevant levels of complexity. At the same time, it is trying to be as compatible with MoRE as possible, but naturally it is not the same. I am not using Sanctuary Areas or any other qualitatively stated characters (street hierarchy).

Plot Scale of a plot as the smallest element of urban form covers characters deductible from a single plot, building (or buildings) on the plot and their interrelations.

7.5.0.0.0.4 Plot

- In urban morphology, plot is usually seen as one of the fundamental elements of urban form.(Moudon 1997, Kropf) However, its definition and employability in the analysis is far from ideal as Kropf 2018 identified.
- Cadaster is not equal to morphological plots
- Physical boundaries are not always present
- There is no data source which could be used in algorithmically quantitative analysis
- The level of plot is still the important one, so there must be another solution how to capture the smallest scale
- While Schirmer and Axhausen propose what they call 'influence zone' defined by topological skeleton of unbuilt space, Hamaina et al (2012, 2014) use Voronoi tessellation. Both approaches are giving very similar results, but I believe that Voronoi tessellation is algorithmically easier to achieve.
- Voronoi cell can be used as a substitute of plot in some cases like covered area ratio and similar. The hypothesis states, that characters based on Voronoi cells generated from building footprints have similar information values as characters based on plots. However, it is practically impossible to test this hypothesis. (There would be option to use Gorbals and compare, because Ale have plots in Gorbals)

Street (Edge, Frontage) (Plot series?) Minimal set of plots forming coherent structure. Characters can be based on one street edge or a street itself (composed of 2 street edges facing each other). What about the situations where it is complicated to identify street edge? Like modernist structures. This scale expects to have some king of frontage, or plots series present. There could be cases where street edge (plot series) will be the same as the block. How to resolve this situation? Use the same value for both? Might be the solution.

Block Enclosed group of street edges (plots) forms block. Characters are capturing values reflecting attributes of blocks as structural units of urban form. Definition of block is not easy, Conzen, as well as Cannigia, didn't use it. There are some reasons, but Kropf (2014) is proposing some solutions. Definition just

by streets is simplistic, figure ground pattern deny its structural connections to urban tissue.

Vicinity Literature usually use the term neighbourhood for scale larger than block, however the definition of neighbourhood is unclear especially in terms of its physical structure. Therefore, it is usually used fixed euclidean or walking distance to simulate the neighbourhood scale (Song and Knaap, 2007; Schirmer and Axhausen, 2015; Song et al., 2013). In our case, term vicinity is used to avoid misunderstanding. The exact distances (either euclidean or walking) will be investigated later, but it is expected something like 400m, (which also reflects the usual size of SA). Kropf (2017) uses term urban tissue on this scale, but it doesn't make sense to use this term while trying to identify urban tissues :). Moreover, tissues defined as DHC might have quite variable size. *Modifiable Areal Unit Problem* "The buffer definition does not avoid the Modifiable Areal Unit Problem (MAUP) but it enables us to focus on the neighbourhoods around newly developed single-family parcels, which are most consistent with individual behaviour (Handy, 2004), and makes it possible to compute the following unique measures for each of the 6788 single-family houses. There are other limitations of using buffers, particularly for parcels located near the edge of residential subdivision and parcels near natural barriers. The 1/4-mile radius method could potentially include a large amount of surrounding land of different characteristics. However, as the intention is to capture features of each house's immediate neighbourhood, it is thus only claimed that the buffer is a convenient unit of analysis and illustrates the effects the study sought to capture. " (Song and Knaap, 2007)

`400 metres radius (Alterations in scale, Urban Nuclei, Neighbourhood unit- UDSU`
`400 - 600 metres radius (Towards an urban renaissance - Task Force)`

`Schirmer and Axhausen use 300 metres radius and then approx 1km reachable area`

Let's use 400m at the moment and if needed adapt later according to characters. Question is whether use only one radius or whether use more (for different characters, otherwise there should be dependency).

180919 There should be also sub-chapter discussing the use of topology based distance and euclidean. (Kane and Kim 2018, Van Eggermond and Erath 2015)

Topology based method is trying to simulate walkability, e.g. how many destinations (and their characters) can be reached within certain distance. It is a question, whether walkability based area of vicinity is the one we are aiming for. It is rather scale attempting to capture the immediate context of the area no matter the street network, which is measured separately. In case of topology based characters, we might find that the results are heavily affected by the connectivity of network which might skew the results. Moreover, topology based characters capturing 'accessibility of buildings and floorspace' are specificied individually. Topology is useful in modelling a journey from A to B, less for capturing the context of of the element.

Urban area The aim is to identify urban tissues, so the scale of settlement might seem strange, but it should be used for street network analysis (at least). The results of this kind of assessment is granular to the level of street, so the results are effectively usable on smaller scale. Def.: an aggregation or continuous network of urban communities (Merriam-Webster)

There is a question whether measures on sub-metropolitan scale (e.g. district) should be used, probably in the same form of fixed distance as locality. It will be further investigated.

While thinking about those scales, it is important to understand that all values from all characters no matter the scale will be assigned to plot/building and all the clustering etc will be conducted on this smallest scale.

181203 Every element has inherent relation to elements of other type. There is always relation *building-cell-street-block*. As such, characters using building-cell relation are using topological distance 0, because the relation is always present

7.5.0.0.0.5 Cross-scale measuring system / explanation

- Measuring will cover relevant scales, to capture different levels of complexity of urban form

- Values from measuring will be stored at the plot level, no matter the original measuring scale
- Different scales have different methods of measuring affecting the uniqueness of values derived for each plot (illustrated in the figure below - each colour represents unique value and shows how plots share them).

7.5.0.0.0.6 Selection of taxonomic characters

- 'characters should be distributed as widely and evenly as possible over the organism studied. Colles (1969c) advocates a deliberate policy of stratified sampling from various organs' (Sneath and Sokal,p.103)
- 'A systematic survey of all known characters, or the inclusion of all characters the investigator has been able to observe, should prevent bias of this sort.'(SS, 96) [bias in choosing the characters]
- there is no real benchmark assessing the selection of characters. Theory say, that if you include all possible characters, it should be okay. (Kinda weak I must say)
- the aim is to cover relevant scales as well as different aspects of complexity of form
- characters are identified across 5 scales (plot, street, block, vicinity, urban area)
- characters are identified across 6 categories (dimension, shape, distribution, intensity, connectivity, diversity)
- the base for the selection is content analysis s presented in chapter 03. It generated list of almost 500 characters used in previous studies, but naturally a lot of them are overlapping or using different sets of data than I intend to.
- the database of existing research was then used to select appropriate set using only basic elements of form without any information on hierarchy or uses.
- it generated around 260 unique characters
- some of the characters were logically or partially logically correlated, hence were omitted. Also the characters describing most of the simple dimensions

of derived elements were left out and used only as an input for more complex characters.

- further omission of characters was lead by the intention to eliminate possible doubling - if one use same metric for street and then block level, there is clear and relatively strong partial logical correlation, so following the principle of applying more fine grained data the smaller scale will stay and character on higher will go.
- approximately half of the characters were excluded due to some logical correlations or based on existing research assessing the quality of characters (e.g. shape indicators)
- existing database was not well balanced, for example characters capturing diversity were in only few examples

7.5.0.0.0.7 Shape

- I have identified high number of characters assessing the shape of the object, but applying them all would shew the results towards shape itself.
- Some of them are correlated, then only one was applied
- some of them are designed to capture complexity. In that case, selection is based on Basaraner and Centinkaya (2017) who assessed shape complexity characters.
- characters based on topological skeleton and centreline (Schirmer and Axhausen, 2015) are omitted for two reasons - it is expected that their input is already covered by another characters and it is algorithmically complicated (REF)

7.5.0.0.0.8 Diversity

- new characters capturing diversity has to be introduced
- Dibble was using statistical values (min, max, stdev…) to capture diversity within the subsets. But the applicability is limited, therefore it is better to use more complex calculations focusing directly on diversity

- the most frequently used measurement of diversity is Simpson's diversity index, but it has one major drawback. It is necessary to identify all subsets of character before calculation. It means, for example, that to measure Simspon's index of plot size, one first have to come up with categories of those sizes and those will necessarily be arbitrary and specific for each case (as in Bobkova 2017). I argue that while most of the characters are linear values, categorising them means loosing the information and adding unnecessary bias into the dataset.
- I am proposing the use of Gini coefficient as a measurement of difference between equal and real (statistical) distribution. Gini index in urban morphology was used by Thai 2005 and power law distribution (which can be captured by Gini) by Louf and Barthelemy 2014 as well as Salat 2017, Gielen (2017)(index of concentration) and Feliciotti (2018).
- Another approach proposed again Tsai 2005 and later Caruso, Hilal and Thomas (2017). That is use of Local index of spatial autocorrelation (LISA). In my case, Local Moran's I or Geary's C autocorrelation coefficient are in fact capturing homogeneity of distribution of values, capturing *homogeneity in heterogeneity* - if the locality is heterogeneous in its key values, the LISA-based character will remain similar across the area and allow us to identify this heterogeneity as typical description of that locality and therefore identify it as DHC.
- Both Gini and LISA are employed in previously unused range (X and X characters).
- LISA
  - "Moran's index is somewhat equivalent to Geary's coefficient and they can be substituted for one another. However, in practice, Moran's index cannot be replaced by Geary's coefficient and vice versa due to a subtle difference of statistical treatment. Compared with Geary's coefficient, Moran's index is more significant to spatial analysis." (Chen 2013 p. 2)
  - Let's work with Local Moran's I

7.5.0.0.0.9 Selection strategy in points

1. List all characters found in literature (470)
2. Select those using data intended to be used within this research (building footprints w. heights, street network)
3. Eliminate logical correlations (full as well as partial)
4. Eliminate ineffective characters (axial maps, topological skeleton)
5. Add characters where are clear gaps (diversity, plot-level Voronoi cell)
6. Balance scalarity and uniqueness of values
7. Add - meaningless, invariant, empirical correlation

7.5.0.0.0.10 Data preprocessing To minimise errors, it is crucial to run momepy analysis on preprocessed data set. UK OS data (building layer, all of them) are not ready for analysis as they are often split into more elements e.g. one building consists of more shapes and I have no idea what to tell the script to resolve this issue, because different parts of the buildings doesn't even have a common attribute (like Vienna for example) so it is not possible to easily dissolve shapes. In order to prepare this data layer for analysis, some change of topology has to happen, but I have no clue what should be the process. Poor quality of cadastral layer excludes its use (which could be great). To partially overcome this issue, it might be necessary to work with non-uk data (Prague, Vienna…).

7.5.0.0.1 Generating blocks Blocks are generated based on the street network and morphological tessellation. Because the street network obtained from open data portal is capturing car-based network, it sometimes does not connect where it should. This should be fixed.

In the case of Prague, using original street network I have generated 9428 blocks, out of which 1839 were "unusual". (19.5%)

```
bdkSec > 500 or bskCom < 0.2 or bskCon <0.76 or bskERI < 0.7 or bskShI <0.5
```

After that I fixed the street network so it snapped to itself and closed gaps in street network - if the 20m extension of line intersects street network - snap. If the

70 extension of line intersects boundary of built-up area (defined by tessellation), snap.

The result gave me 9800 blocks and 1092 unusual (11%). 10% of unusual blocks are randomly selected and assessed whether they are correct blocks or incorrect. Based on that, the approximate error is estimated.

Out of 109 randomly selected blocks, 76 were marked as correct representation of block, 33 as incorrect. Based on that, the **estimated error is 3.4%**. That includes blocks which were incorrect before the network snapping as well as blocks which were falsely identified by the snapping.

Additionally, there should be a subchapter talking about exceptions which momepy is not able to capture (Krizikova, Karlin, Nabrezi Karluv most).

7.5.0.0.1.1 Problem with blocks in modernist structure As block is defined by street, it is expecting that street is major divider of space and was there first. In modernist structures, street is often designed as a way though the area, in the middle of what we could see as morphological piece (or block). We are effectively trying to define something which does not exist. **WHAT IS THE CONSEQUENCE OF THIS???**

7.5.0.0.2 DHC recognition Instead of straightforward cluster analyisis, we might identify significant clusters based on spatial autocorrelation. DHC could be identified as a specific combination of those significant clusters???

7.5.1 ORIENTATION

Duchêne et al. [5] summarized five measures (Fig. 2), including 'longest edge,' 'weighted bisector,' 'wall average,' 'statistical weighting,' and 'smallest minimum bounding rectangle (SMBR),' and concluded from their experiments that the SMBR is the most appropriate one. (Yan, 2007)

Chapter 8

Taxonomy of urban tissues

- Forming a taxonomy from sample data (chosen UK cities?)

Chapter 9

Synthesis

Appendix 1: Some extra stuff

Add appendix 1 here. Vivamus hendrerit rhoncus interdum. Sed ullamcorper et augue at porta. Suspendisse facilisis imperdiet urna, eu pellentesque purus suscipit in. Integer dignissim mattis ex aliquam blandit. Curabitur lobortis quam varius turpis ultrices egestas.

Citation examples:

`[@almazan2012]` (Almazán & Nakajima 2012) `[see @almazan2012]` (see Almazán & Nakajima 2012) `[@almazan2012, pp.33-34]` (Almazán & Nakajima 2012, pp.33–34) `@almazan2012` Almazán & Nakajima (2012) `@almazan2012 [p.2]` Almazán & Nakajima (2012, p.2) `[-@almazan2012]` (2012)

From Zotero export as Better BibTeX

Chapter 10

References