Parametric Urban Data Structuring and Spatial Query

Advanced data mapping and selection methods for parametric modelling environments

Abdulmalik Abdulmawla [1], Sven Schneider[1], Reinhard Koenig[12], Martin Bielik[1], Ekaterina Fuchkina[1]
[1]Bauhaus-University Weimar [2]Austrian Institute of Technology
[1]{abdulmalik.abdulmawla|sven.schneider\reinhard.koenig \martin.bielik\ekaterina.fuchkina}@uni-weimar.de [1]

This paper presents a method for organising urban data inside the CAD environment into a hierarchical structure, which promotes the ease of transferring information between all available urban elements, from streets to buildings passing by the plots and blocks. This is done using parametric methods that map the urban data using the available CAD and GIS records. Finally, the paper presents a couple of example scenarios where such methods are most needed and how much they could facilitate more detailed and complex data to be accessed, compared, and analysed.

Keywords: *Data Query, Urban Geometry, Spatial Mapping.*

INTRODUCTION

Nowadays, urban designers must operate with more complex data, increasing their tasks' difficulty. The extraction of basic spatial information depends mainly on the individual's technical knowledge. Moreover, as the process that generates the project data is challenging to document, more project data are deprived of any guidance to be understood, reused, or further developed quickly.

Furthermore, Since the advancement of different computational methods and the spread of parametric modelling environments, urban designers have generated many urban data that, once well organised, could increase efficiency, reduce the time of repetitive tasks, and save a lot of resources.

Urban data from GIS to CAD

Geographical Information System (GIS) studies well summarised the nature of urban data. They are divided into two main categories, Geometrical (Locale) and Attribute. Geometrical data is simplified into three main types: Points, Curves, and Polygons. At the same time, Attributes are the features carried inside the geometrical types, like a point carrying information about a tree's location and specifications, as in figure 1.

The structure of information inside the attributes also matters. For example, a curve geometry could represent a street with two names with different lengths. This means that the key "Address" attribute will have its value as a list of two items with the name of each part. And if there is another key for the "Length" of each part, the list of values should also be in an order corresponding to the names of each part.

Using this structure, any type of information could be stored. Even analytical information like the area of visibility at a specific location. Or a result of empirical data gatherings like the number of pedestrians passing at a specific date and time. Moreover, if all the geometry types in a file are numbered, connections could be made between them by storing their number ids inside each other. For example, a Block of an ID=5 could host several Plots with IDs= {110, 111,112}.

Utilising this data structure, generating any GIS information inside the CAD environment is possible. But CAD files are not that easy to organise or maintain. Due to the difference in use for each software. Since the advancement of Computational design and the growth of parametric modelling platforms like Rhino and Grasshopper, the amount r of information generated inside a CAD file has grown exponentially. Since Rhino provides a location dedicated to storing the user information, any geometry stored in Rhino could follow the same structure as standard GIS data.

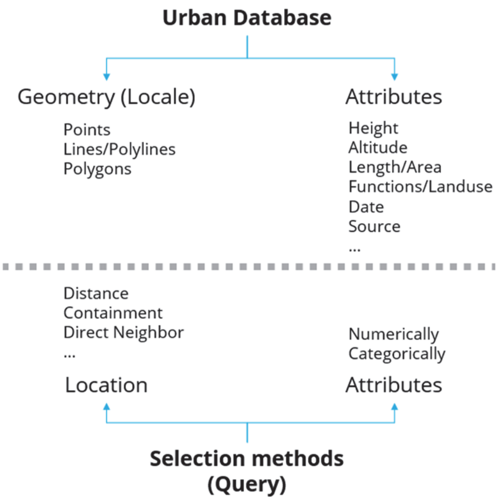


Figure 1
Types of Urban Database (Top)
Types of Selection methods (Bottom)

Data Selection (Spatial Query)

Exporting information from a database is called a query, and since we are dealing with geometric data, we will call the selection of urban data Spatial Query.

The spatial query is more like an algorithm. It is a set of rules that define the process of selecting, manipulating, and extracting the data. This paper considers data mapping from one geometry to another as part of the spatial query.

There are mainly two types of Spatial Query: Attribute-based query and Location or geometrical-based query. Attribute queries, on the one hand, are direct searches for features in a GIS database. For example, extracting a value based on a specific key embedded with a shape.

Location queries are what spatial queries are, is a process of asking a question to be extracted from the spatial geometry data. For example, a simple distance search like the neighbours of a plot. Or finding out the number of shops overlooking a street is a more complex query that requires multiple steps.

Spatial data linkages

With the growth of open geodata, urban designers have easier access to GIS data supplied by many sources. Parametric modelling environments like Grasshopper and a Plugin for the CAD software Rhino are popular tools. Grasshopper has many add-ons that handle shapefiles, the most common form of online GIS data, like reading, mapping, and writing inside Grasshopper.

But with such data inside a parametric environment, many data querying tools are still missing from the GIS software to be integrated inside Grasshopper. Some of the most basic methods for spatial data mapping are about how to understand urban data structuring and Geometrical relations.

Plots as Link

Berghauser Pont et al. (2019), Lim et al. (2015), and Vialard (2012) have outlined the importance of the linkage between the different elements of urban morphology. They highlighted the importance of Blocks, Plots, or both in facilitating the connection between the streets and the buildings.

Bobkova et al. (2019) argue that we must study urban form to develop and sustain life in the urban environment. They consider both Streets and Buildings as the main layers of urban form. Streets facilitate the accessibility to different urban spaces hosted mainly inside the city buildings. In a recent study, Bobkova et al. (2021) added that some properties of the plots, like their frontages, correspond to essential qualities such as the ability to operate as an interface between street and building.

This is also highlighted in other urban morphology studies. Hillier (1996) and Kropf (2009) have also differentiated urban elements into two main categories, the buildings and the spaces between them. And by studying both categories together, one should better understand how they affect life in different urban areas.

Modifiable Areal Unit Problem (MAUP)

Openshaw (1984) and Wong (2004) have pointed to a common problem in any spatial analysis process in Physical Geography. Since different studies could be applied to the same area over time, some problems in the analysis units would need to be addressed. They collected them under the Modifiable Areal Unit Problem (MAUP) using two main problems the zoning effect and the scaling effect.

The zoning effect is when the boundary of the analysis area changes. Like if a study group selected another boundary to measure the density of the same region. The Scaling effect addresses any change in the unit design or type. As if a new plot division was added to an existing dataset.

Both problems are related to changes in the real world and the digital data collected. And since our control over the real world is impossible, we should focus more on how to structure the data so that once any of the effects appear, the data can be easily adapted.

METHODOLOGY

This paper systematically organises urban data to access information needed to solve specific spatial problems quickly. The method consists of a collection of mapping techniques that record the data in a CAD file once and then store them in a location corresponding to the type of information. This is done using parametric modelling inspired by the work of Lee and Hardy (2005). Specifically, the topological relationship among urban geometry

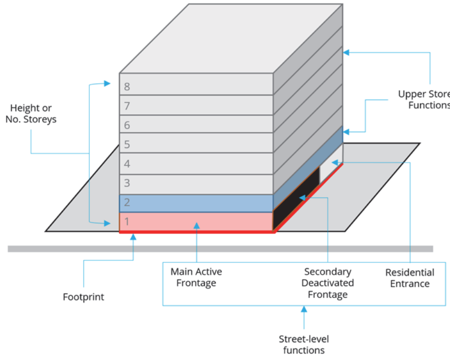


Figure 2
Geometrical features of buildings

features like ensuring shared geometry and considering interfering features.

We create a hierarchical data structure to link urban geometry (from the street, over the block, and plot to the building). These links between the different scales are bi-directional to access the plots that belong to a street and the streets that belong to a plot. This data can include analysis information, like visible buildings to streets, local building densities, active building frontages, etc. We will demonstrate the method in the case of Weimar, using Data from OSM and GeoPortal from Thuringia, Germany (GDI-TH).

All analysis was calculated inside Rhino and Grasshopper using DeCodingSpaces Toolbox by Koenig et al. (2018) as it has a collection of methods for urban data generation and evaluation.

In summary, we propose:

- Mapping urban geometry in a specific hierarchical structure using the parametric environment.
- Recording urban data relations and embedding them in each urban element. This should help the users quickly rebuild the exact spatial hierarchy when the files are reused.
- Spatial data query in the parametric environment while considering MAUP effects.

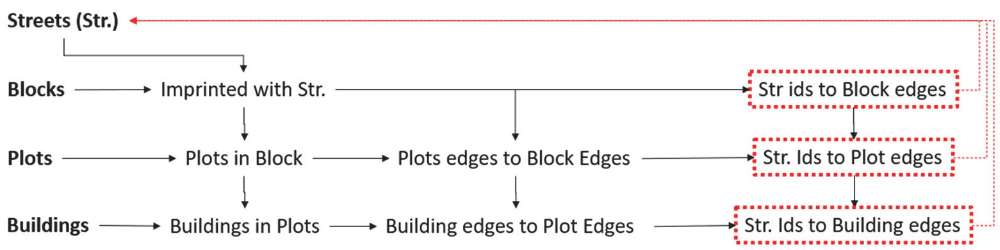


Figure 3
Mapping Hierarchy from streets to buildings passing by plots and blocks

Organising the city geometry

The method we used for dividing the city geometry into layers with simple elements was first adopted by Dennemark et al. (2017). Figure 3 shows our approach to the hierarchy connecting the geometry. We start with the street network, passing data through Blocks that contain the Plots, which contain Buildings with street frontage information.

Streets

In our study, the streets mainly measure accessibility using the nodes and linkages in the space syntax spatial configuration methods (Hillier 1996a). This simplified our focus only on the segments of the streets as a network. Each street segment is carefully placed between the plots and buildings to avoid intersections.

Also, as presented in the case study later, the streets are numbered and named to differentiate them from each other when their information is passed to the block, plot, or buildings. This is used mainly to keep the data structure from being randomised in Grasshopper every time we reload the data. Also, since each street has an ID, it could be used to quickly access their information on any other platform such as Mapbox, as we will show in the case study later.

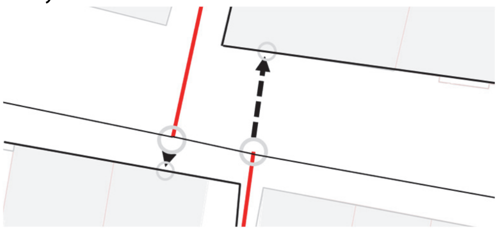


Figure 4
Mapping Street intersection to the Block to facilitate the transfer of multiple streets on one building frontage

Blocks

The block definition is not a straightforward process, and some blocks could include pedestrian paths in them. They are making the borders of the block challenging to define. Moreover, some building designs can include passages like tunnels or bridges under or over their buildings. This adds to the complexity of making a generalisable method for block definition.

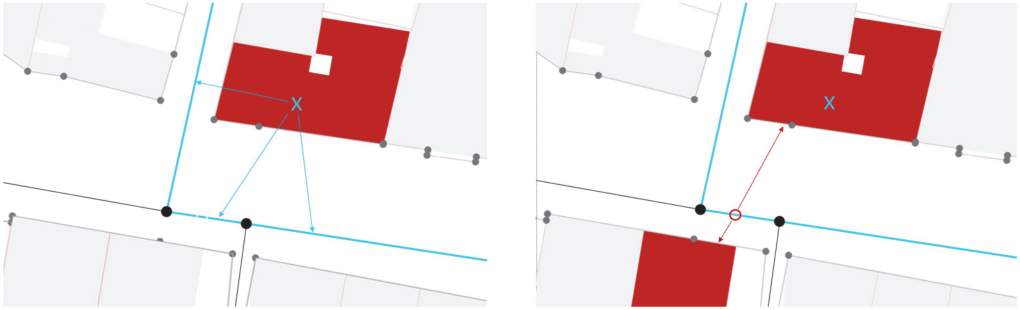
We found the best practice to define the streets first as the public area is the easiest to clarify. Then define the rules for dealing with the semi-private semi-public spaces.

For example, if the study is about the frontage access directly from the street, then the building volume above the tunnel could be excluded in the geometry but kept as an attribute. And if the density of the buildings is more crucial, then the opposite could be done. For solutions where both elements are with the same value, we recommend creating a parallel layer to maintain the original data structure.

The order of the segment lines constructing the Block edges was recorded. This made the transfer of the street frontage information much more straightforward. Moreover, the street intersections were transferred to each Block edge, and the edge which faced this intersection was divided at the mapped point. This made the transfer of all surrounding streets' information much more accessible.

We have also found that extracting the blocks from the plots is easier since they are primarily used in defining land use. And therefore, if we eliminate

Figure 5
Mapping Buildings to their direct Streets and reverse.



all the public street plots, we will have the remaining islands as the plots that are within one block.

Plots

The data from the GDI-TH had all the land in the city covered with plots that define the land use. Even streets and forest areas as well as hiking roads.
Since we can define all public areas, all the streets and plazas were excluded, and the street blocks were generated from the outlines of the rest.

All plots had their ID and which Block they existed in. They also should have street direct front access. Therefore, we look for island plots and merge them with neighbouring street-overlooking plots. This is governed by the buildings that are in them. If multiple plots contain a building, it will be difficult for the algorithm to connect it to all the plots. Therefore, we prefer merging those plots and treating them all as a new larger plot. If no buildings were found, it is merged with the plot with the closest distance to the street.

The edges of the plots were also ordered, and the information on their relation to the Block edges was documented. In other words, each plot edge has the information of all the block edges it touches. And since the Blocks were generated from the plots, any edge that is not touching a block edge could be identified as a not frontage edge and might not have a relation to the streets. This made it easier to differentiate between internal plot edges and frontage edges by looking if plot edges have any recorded information on the corresponding attribute.

Buildings

All three-dimensional building geometries were reduced to their footprint polygon surface to reduce the file size. We found surfaces are more straightforward to transfer than polyline curses. This is mainly due to the problem of transferring information about the open areas inside the building. A building could have many polyline curves inside it, increasing the complexity of attaching information to one single object, as is the case of the actual building data.

All buildings were grouped by containment inside the Plots and Blocks. This made the transfer of information between them straightforward. Therefore, each building has information regarding which street it overlooks, which edge has this specific street ID, and which buildings are directly attached to each edge.

Building functions were recorded inside each building attribute as two types of lists. One list for all the functions with direct individual access from the street level. And another list for the rest of the functions in the upper levels accessible through the same main building entrance.

Frontages

As discussed in the previous section, the information about the frontages is already embedded in each building. Nevertheless, we will discuss some

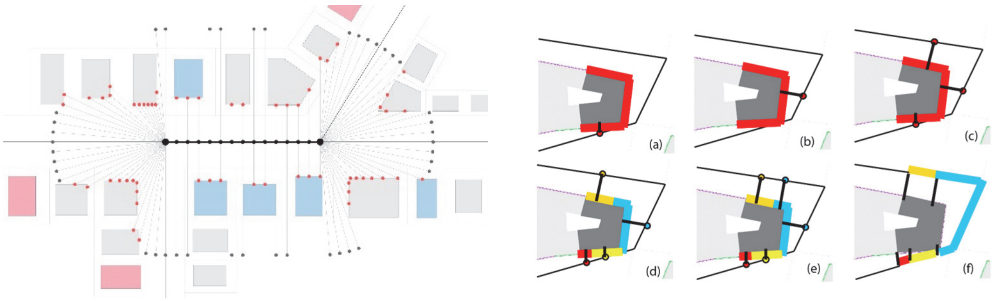


Figure 6
(Left)Mapping visible buildings onto a street (Schneider et al. 2017)
(Right) Mapping building functions through their frontages (adopted from Bielik 2021)

Figure 7
(Left) Mapping frontages to dedicated street segments
(Right) Mapping Frontages to shared street segments.

conditions regarding mapping the frontage information.

Starting with the definition of a Frontage. In this case, we use the term defined by Kickert (2016) as the Active frontages. Specifically, we define how accessible a frontage is to the public and the type of visibility and permeability of this frontage. We will discuss in the second case study how this information is used.

The information on the GDI-TH platform also included the address of the building. Both the name of the street and the number of the building were available. This will be beneficial information for our second case study that will be presented later, as we have observed that the information about its address can define the entrances of each building. Hence the entrance of a building with multiple street frontages is on the address street name.

NOTES ON DATA MAPPING AND QUERY

When querying data inside the parametric environment Grasshopper, some worth mentioning conditions we found.

Information Hierarchy

The direction of mapping information from streets to buildings was most efficient. This method guarantees that the data organisation reduces the complexity of spatial queries.

Moreover, since the inner objects like Buildings and Frontages have all the information about their containment, all the data structure could be reversed so that each Street could have a list of all the information regarding its relation to all the buildings that have its ID.

Visibility and Direct access

When mapping buildings to their streets, some data conflicts might appear. Specifically when choosing to map all visible buildings to the streets. It will produce a significant duplication of street IDs, resulting in difficulty defining their relation to the buildings.

For example, one building could have one direct street frontage and at the same time is visible to 5 other street segments within a 100-meter radius. The only way to specify the streets would be through the research objective.

Nevertheless, we recommend using the direct frontage method due to its simplicity for beginners and wide use for applications that do not include large visibility areas.

MAUP for Building Functions

The problem of mapping building functions is outlined in Bielik (2021). Using our method, we found that some notes must be considered once we introduce more details to the model.

First, different options might affect the analysis result when mapping the functions to the street. But the case is a little different when using Frontages.

As frontages are more defined by their street location, as we mentioned in the case of the building entrance. More methods could be used to map the location of each frontage to a dedicated part of the street.

Another case when is when mapping the number of functions in a building. We found some functions occupying more than one building. Therefore, each function will require a separate identification to avoid duplications in their numbers.

Case study 1

In this case study, we presented a structured set of data in a CAD file for a group of students to design a small neighbourhood and analyse it in the context of the city of Weimar in the seminar on Parametric Urban Design and Analysis (PUDA). The provided data on the city was created using the methods described above. It consisted of 4 main layers for the street network, blocks, plots, and buildings. Each layer has the corresponding geometries with IDs, and embedded attributes that describe each object like the area, the number of storeys, and the number and type of functions inside. The dataset also includes precomputed and analysed information about the performance of each object, like the betweenness centrality of a street segment or the local density of the area surrounding each building.

Using the tool mineR (Abdulmawla et al. 2018), the students could extract information about the city immediately while still being inside the same design and analysis environment of Rhino/Grasshopper. They could also reduce the computational effort by embedding the design information inside the CAD file.

We developed a process for combining various data from the student's group in a shared file. As shown in figure 8, it is possible to upload the data to an online map platform like Mapbox. This allowed a simplified and more general for of presenting the resulting information.

While working with the data, the students required some tutoring to understand the process of extracting and recording information to the CAD model.

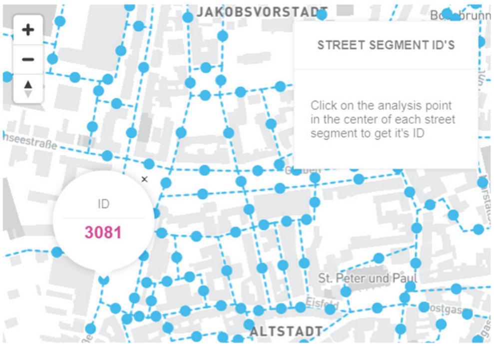


Figure 8
Street information passed from our data model to an online mapping platform.

Case study 2

For a research study, our objective was to find relations between different groups of functions within the active city centre of the city. We used the method to map the active building frontages to their dedicated segment of the streets inside the historical centre of Weimar, Germany.

To analyse the betweenness centrality of each function frontage, we had to find the part of the street segment that each frontage overlooks. This is crucial to capture the through movement that the betweenness centrality indicates. The final model was generated using more information based on the method to connect each frontage ID with its corresponding street segment length.

Based on the processed data, matching the information about the visibility with the accessibility of each frontage information was straightforward using matching the IDs of the frontages in the buildings. We used the direct frontage mapping shown in Figure 9, where we dedicated a frontage for each function type and connected this information to the dedicated street segment.

For further analysis, we used the statistical t-test to find relevant relations between each function. We used sub-groups of functions (Fig 9) of the whole function group inside the city centre. The idea is to compare each sub-group with the rest of the group to show how different the mean of each sub-group is from the rest. Next, we use this information to compare the sub-groups with each other's (Fig. 10) to find functions that share the same location characteristics. We found three different types of functional clusters inside the city centre.

Figure 9
location information for all the seven sub-groups of functions inside Weimar city centre

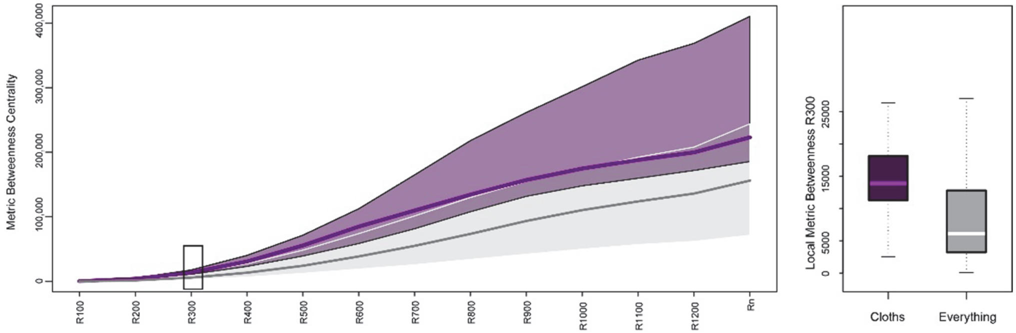


Figure 10
(Left) a plot of the metric betweenness centrality of clothes function sub-group (purple) and the rest of the group (grey)
(Right) Boxplot of the BC at radius R300 to compare the relation between the mean of the cloth to the rest of the group

DISCUSSION

Although the tools and methods we used are well defined, a few notes need to be addressed: First, the complexity of the data required for analysing the connection between geometries requires a large computer memory. And it grows exponentially with increasing the study area. This could be solved by dividing the study area into smaller sub-areas that are easy to join after the analysis.

Second, updating the geometry might require the embedded data to be double-checked for changes. For example, we found that one street block had to be cut into two during the study. This resulted in a new street segment, two new blocks, and new plots and building sides exposed to the street.

Third, the complexity of some geometries was manually modified to run the analysis efficiently. For example, a plot from the GDI-TH had a plot inside another, which created a plot with no accessible street front. Therefore, the inner plot was excluded.

CONCLUSION

Using parametric CAD environment methods can facilitate further advanced spatial data analysis. As shown in this paper, linking the elements of urban geometry results in more possibilities for spatial data analysis.

The presented data structure facilitates easy access to link the available urban geometries and related data (like land uses and local density) for either urban design or analysis-based tasks. It makes it possible to manoeuvre through urban data easily.

Moreover, since the whole process is developed in a parametric CAD environment (Rhino/Grasshopper), any processing step is easily accessible for data extraction or manipulation. This gives advanced users more flexibility to tailor the tools for their needs for any custom investigations.

In the case studies, we exemplified the methods, such as (1) extracting and recording street segment information in the data structure and (2) identifying similarities and differences in location characteristics between functional groups.

This presented method facilitates the ease of exacting information and reduces repetitive computational tasks by recording their results. However, since these applications were not used in cross-disciplinary fields, once shared with the community, we could further develop the tool for broader integration.

REFERENCES

Abdulmawla, A., Schneider, S., Bielik, M. and Koenig, R., 2018. Integrated data analysis for parametric design environment-mineR: a grasshopper plugin based on R.

Bielik, M., 2021. Form-Activity-Movement Interaction Model. Doctoral Dissertation. https://doi.org/10.25643/bauhaus-universitaet.4397

Bobkova, E.; Marcus, L.; Berghauser Pont, M.; Stavroulaki, I.; Bolin, D. (2019) 'Structure of Plot Systems and Economic Activity in Cities: Linking Plot Types to Retail and Food Services in London', Amsterdam and Stockholm. Urban Sci. 2019, 3, 66. https://doi.org/10.3390/urbansci3030066

Bobkova, E., Berghauser Pont, M. and Marcus, L. (2021) 'Towards analytical typologies of plot systems: Quantitative profile of five European cities', Environment and Planning B: Urban Analytics and City Science, 48(4), pp. 604–620. DOI: 10.1177/2399808319880902.

Berghauser Pont, M., Stavroulaki, G., Bobkova, E., Gil, J., Marcus, L., Olsson, J., Sun, K., Serra, M., Hausleitner, B., Dhanani, A., & Legeby, A. (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), 1226–1242. https://doi.org/10.1177/2399808319857450

Berghauser Pont, M., 2018. An Analytical Approach to Urban Form. In Teaching Urban Morphology (pp. 101-119). Springer, Cham.

Dennemark, M., Schneider, S., Koenig, R., Abdulmawla, A. and Donath, D., 2017. Towards a modular design strategy for urban

masterplanning. Sharing of Computable Knowledge!, p.485.
Hillier, B., 1996a. Space Is The Machine: A Configurational Theory Of Architecture.
Hillier, B., 1996b. Cities as movement economies. Urban design international, 1(1), pp.41-60.
Kickert, C.C., 2016. Active centers–interactive edges: The rise and fall of ground floor frontages. Urban Design International, 21(1), pp.55-77.
Koenig, R., Beilik, M., Knecht, K., Abdulmawla, A. and Fuchkina, E., 2018. New Methods for Urban Analysis and Simulation with Grasshopper-Using DeCodingSpaces-Toolbox.
Kropf, K., 2009. Aspects of urban form. Urban morphology, 13(2), p.105.
Lee, D. and Hardy, P., 2005, July. Automating Generalization–Tools and Models. In 22nd ICA Conference Proceedings, A Coruña, Spain.
Lim, L., Yang, T., Vialard, A., Feng, C. and Peponis, J., 2015. Urban morphology and syntactic structure: A discussion of the relationship of block size to street integration in some settlements in the Provence. The Journal of Space Syntax, 6(1), pp.142-169.
Pratt, M., 1999. Asking the right questions. https://www.esri.com/news/arcuser/0309/files/rightqueries.pdf
Vialard, A. (2012) 'Measures of the fit between street network, urban blocks and building footprints', In: Eighth international space syntax symposium, pp.8101:1–8101:16.
Wong, D.W., 2004. The modifiable areal unit problem (MAUP). In WorldMinds: Geographical perspectives on 100 problems (pp. 571-575). Springer, Dordrecht.
Openshaw, S., 1983. The modifiable areal unit problem (Vol. 38). Norwich: Geo Books.
Parametric Urban Design and Analysis – PUDA (2022). Online Learning Platform for the Bauhaus University-Weimar. Available at: https://otp.uni-weimar.de/courses/parametric-urban-design-analysis-puda-open/ (Accessed 15 June 2022)