Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale

Martin Fleischmann[*], Alessandra Feliciotti[1], Ombretta Romice[1], Sergio Porta[1]

University of Strathclyde, James Weir Building, 75 Montrose Street, Glasgow G1 1XJ, United Kingdom

ABSTRACT

Urban Morphometrics (UMM) is an expanding area of urban studies that aims at representing and measuring objectively the physical form of cities to support evidence-based research. An essential step in its development is the identification of a suitable spatial unit of analysis, where suitability is determined by its degree of reliability, universality, accessibility and significance in capturing essential urban form patterns. In Urban Morphology such unit is found in the *plot*, a fundamental component in the morphogenetic of urban settlements. However, the plot is a conceptually and analytically ambiguous concept and a kind of spatial information often unavailable or inconsistently represented across geographies, issues that limit its reliability and universality and hence its suitability for Urban Morphometric applications. This calls for alternative methods of deriving a spatial unit able to convey reliable plot-scale information, possibly comparable with that provided by plots.

This paper presents M*orphological Tessellation* (MT), an objectively and universally applicable method that derives a spatial unit named *Morphological Cell* (MC) from widely available data on building footprint only and tests its informational value as proxy data in capturing plot-scale spatial properties of urban form. Using the city of Zurich (CH) as case study we compare MT to the cadastral layer on a selection of morphometric characters capturing different geometrical and configurational properties of urban form, to test the degree of informational similarity between MT and cadastral plots.

Findings suggest that MT can be considered an efficient informational proxy for cadastral plots for many of the tested morphometric characters, that there are kinds of plot-scale information only plots can provide, as well as kinds only morphological tessellation can provide. Overall, there appears to be clear scope for application of MT as fundamental spatial unit of analysis in Urban Morphometrics, opening the way to large-scale urban morphometric analysis.

1. Introduction

Urban morphometrics (UMM) is an expanding area of Urban Morphology that, through the application of quantitative spatial analysis techniques, aims at capturing the composition and configuration of the urban form of cities and, on this basis, supporting evidence-based comparisons between different urban areas and possible correlations with non-spatial dynamics. To fulfil this aim, the identification of a reliable, significant and universal spatial unit of analysis is of crucial importance.

In traditional Urban Morphology, such unit of analysis is often the *plot*, considered to be the smallest meaningful unit of spatial subdivision and a fundamental component to understand the spatial structure of the ordinary fabric of urban settlements (Panerai, Castex, Depaule, et al., 2004; Porta & Romice, 2014) and their processes of formation and transformation in time (Whitehand, 1981).

However, despite its significance, the plot remains a problematic construct. At *ontological* level, there is no agreement on *what exactly a plot is*: indeed, it has been variously defined as *"a land-use unit defined by boundaries on the ground"* (Conzen, 1969, p. 128), a *module* of the urban tissue constituted by a built-up area and its open pertinent area (Caniggia & Maffei, 1979), a piece of property, subject to subdivision and amalgamation as a result of successive patterns of occupation (Moudon, 1986), or again, according to Bobkova, Marcus, and Berghauser Pont (2017), as *"a basic unit of control"*, *"a fundamental link between spatial and non-spatial medium"*, *"a connection between built space and space of movement"* and *"the framework for building evolution over time"* (p. 47.5). And crucially, more often than not, these definitions may represent very different entities on the ground *"potentially leading to misinterpretation and so a somewhat obscured picture of the dynamics of urban form"* (Kropf, 1997, p. 1).

In addition to this issue, the identification of plots in the urban fabric also poses a series of *analytical* problems: given a map or a satellite image, how to consistently determine plot boundaries (Fig. 1). And, in case of existing datasets, what do they *actually* represent? What definition of plots do they adopt? Are different datasets comparable?

Not all mapping agencies explicitly report plots and, even when they do, not all of them define or represent plots in the same way. In
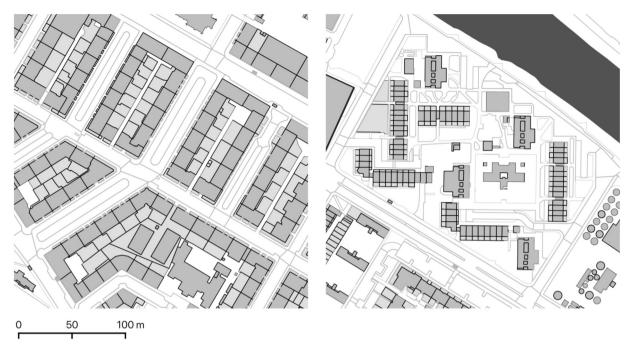
Fig. 1. Comparison of traditional (left) and modernist (right) urban tissues in Glasgow. Plots are clearly better identifiable—even just visually—in the former, where distinction of public and private space is clear-cut, than in a modernist housing estate, where the transition between public and private is blurred. Source: Ordnance Survey MasterMap, January 2019 (EDINA Digimap Service).

some spatial databases, as in the Swiss Katasterwesens, plots are represented as unitary land parcel, whilst in other cases, ownership-based plots can be made of multiple unlinked features, as in the French Cadastre, limiting comparability between different datasets. In other cases, the identification of plots from available sources is inferred by the analyst via resource-intensive manual interpretation. However, that makes the resulting procedure on one hand unsuitable for large scale analysis, and on the other potentially biased, as heavily dependent on both individual interpretation and the often uneven quality of the underlying data. Indeed, while through open-data policies (Huijboom & Van den Broek, 2011) and Voluntary Geographical Information System (VGIS) (Barrington-Leigh & Millard-Ball, 2017) the availability of free-to-use geo-data is growing dramatically, their quality, coverage and resolution is often insufficient to determine individual plots and generally limited to building footprints, street centrelines, natural features and administrative boundaries. All of this reduces considerably the reliability of the analysis and the universality of its results.

Given the aforementioned issues, and despite plots being still widely used in Urban Morphology to capture the "*pattern of human intention and activity*" (Kropf, 1997, p. 5), they are ill-suited as basic unit for UMM applications.

To this end, this paper presents *Morphological Tessellation* (MT), a method of deriving a spatial unit of analysis, the *Morphological Cell* (MC), which is able to convey reliable, universal and meaningful plot-scale information and, at the same time, to minimise manual labour, subjective interpretation and data dependence. After presenting the MT method, MC and plots are compared, to test whether the two entities hold similar informational value and similar capacity to capture meaningful spatial patterns. This is done using Zurich, a city characterised by a relatively heterogeneous urban fabric, as explorative case study, and comparing the MC generated via the MT method with plots derived from the cadastre.

The remainder of the paper is structured as follows:

- Section 2 introduces the Voronoi tessellation and its use in urban form studies;
- Section 3 describes the MT method;
- Section 4 proceeds to the validation of the method in the case of Zurich;
- Section 5 presents and discusses the findings of the analysis;
- Section 6 offers some final conclusions and some reflections on limitations of the method and future research.

2. Voronoi tessellation in urban form: a state of the art

At the core of the proposed MT lies the *Voronoi tessellation* (VT), a method of geometric partitioning that from a planar set of *'seeds'* generates a series of polygons, known as *Voronoi Cells* (VC). Each VC encloses the portion of the plane that is closer to its seed than to any other (Fig. 2), ideally representing its *'influence zone'*.[2]

VT has been already used in relation to urban form, in the context of spatial clustering algorithms (Dogrusoz & Aksoy, 2007) and built-form geometry generalisation techniques (Ai & Zhang, 2007; Basaraner & Selcuk, 2004; Li, Yan, Ai, et al., 2004; Liu, Guo, Sun, et al., 2014), or as input for image-based pattern recognition (Yu, Ai, & Liu, 2017). In recent years, VT was used to study the micro-scale properties of the urban fabric (Hamaina, Leduc, & Moreau, 2012, 2013) in order to produce a reliable method for urban form patterns' recognition, which pioneered the generation of VC from building footprints. Later, Schirmer and Axhausen (2015, 2019) devised a method to define "*influence zones*" around buildings using a "*topological skeleton*" of unbuilt space that is mathematically similar to VT. In parallel, Usui & Asami, 2013, 2019; Usui, 2018) included the street network as an additional input alongside the building footprint to the VT algorithm, to mimic the plot structure of traditional Japanese urban fabrics. Whilst the generated mesh shows remarkable similarity to the plot pattern, its main limitation is the inability to capture the spatial pattern of modernist (post WWII) urban tissues and the highly variable distance between building and street that is typical of such fabrics. On a similar vein,

---

[2] The term Voronoi Tessellation can be used to describe both the process of partitioning space (method) and the geometric mesh it generates (output). In this text, the two meanings are used interchangeably.
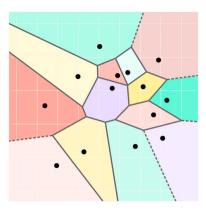
Fig. 2. Voronoi tessellation based on randomised seeds. Each colour represents the area of one tessellation cell (influence zone). Dashed lines end in infinity.

of this line of works, insofar it too utilises the VT procedure. However, unlike previous studies, this paper aims to provide a fully operational and replicable method by examining the details of the tessellation process and its parameters, and testing the similarity of morphometric characters as measured on both MC and plots through direct comparison.

3. Morphological tessellation: the method

Whenever observing a map or a satellite view of a city, the eye of the observer is caught by the existence of a fundamental relationship between buildings – their geometry and spatial configuration – and the plot pattern. This *'intuitive'* relationship is the reason why approaches based on VT appear to *'make sense'* when applied to the urban form of cities: by partitioning the space into cells, they capture the way buildings relate to each other in space and, more precisely, give a spatial
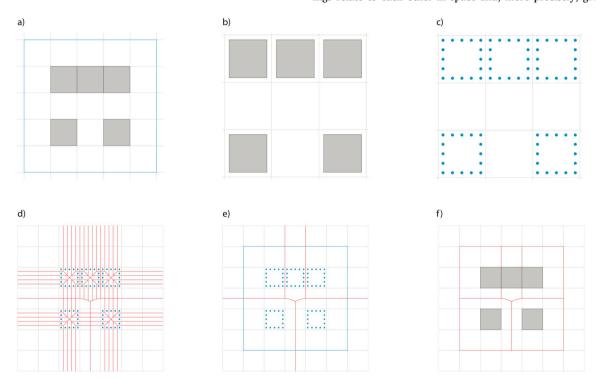

Fig. 3. The proposed MT method. Grey polygons represent building footprints, while red lines show the edges of tessellation at each step. 3a) Building shapes within boundary of study area (blue); 3b) inward offset from building footprint polygon; 3c) discretization of boundaries of polygons into points; 3d) generation of VCs around points: at this stage, the edges of cells (red) tend to infinity; 3e) dissolution of Voronoi cells based on original building ID; 3f) clip of preliminary tessellation by study area. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
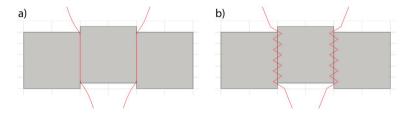

Fig. 4. Illustration of the effect of improper combination of inward offset distance and discretisation interval causing "saw-like" geometry on the boundary bethwen adjacent buildings (4b) compared to ideal combination (4a).

Araldi and Fusco (2017, 2019) developed an approach based on VT and street segments to define a spatial unit based on the pedestrian point of view.

In all these cases, the use of Voronoi tessellation helped to rigorously and reliably cluster components according to their configuration although, as pointed out by Usui and Asami (2019), the relationship between VC and 'conventional' plots has never been directly tested to date. In this sense, the MT approach is to be intended as a continuation

meaning to the "*morphological influence*" that each building exerts on its immediate spatial context (Usui & Asami, 2017). This, in turn, implicitly captures how spatial configuration affects visibility, light penetration, ventilation, movement, etc. around each and every building (Hamaina et al., 2012).

The main advantage of methods based on VT is the capacity to derive objective spatial partitions that are applicable to every type of urban tissue in a way independent from the researcher's subjective
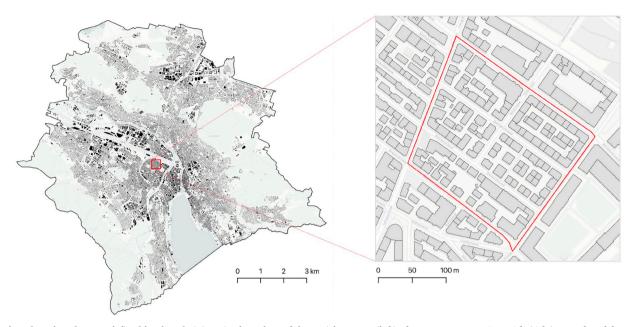
Fig. 5. The selected study area, defined by the administrative boundary of the Zurich Kanton (left); *the Langstrasse area in Zurich (right) was selected for testing the tessellation algorithm parameters: the red-line boundaries follow the street centerlines.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
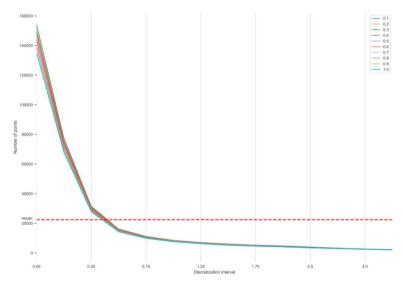

Fig. 6. *Relation of discretization segment length and number of points generated. The red line illustrates the mean value above which the number of points more than doubles at each step.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

interpretation. In addition, most of these methods (Hamaina et al., 2012, 2013) require minimum data input, as they fundamentally rely on the polygon that describes the footprint of a building. Similarly, the proposed MT method only requires a polygon layer representing building footprints (Fig. 3a). From this, MT moves forward in five steps:

1. *Inward offset from building footprint* (Fig. 3b). The offset is necessary to avoid overlaps between boundaries of adjacent buildings and generate a gap between adjacent geometries which will later define the boundaries of the cell.
2. *Discretization of polygons' boundaries into points* (Fig. 3c). As VT can only be generated from point features, the polygonal shape of the building footprint needs to be approximated as series of points to be placed at regular intervals along its boundary, where generated points retain the ID of the building they belong to.

3. *Generation of Voronoi cells* (Fig. 3d). VC are generated around each of the points representing the building footprint. Again, the original ID of the building is preserved in the resulting VC.
4. *Dissolution of Voronoi cells* (Fig. 3e). All VCs sharing the same building ID – and hence generated from the same building – are dissolved in unitary geometries. This step provides a preliminary boundary of the MC.
5. *Clip of preliminary tessellation* (Fig. 3f). As a geometrical construct, VCs tends to infinity as the boundaries of each cell are only defined by proximity with adjacent 'seeds'. However, when applied to the analysis of urban form, for obvious reasons no cell can tend to infinity. To avoid this, it is necessary to limit the maximum spatial extent of the tessellation by setting defined study area boundaries.

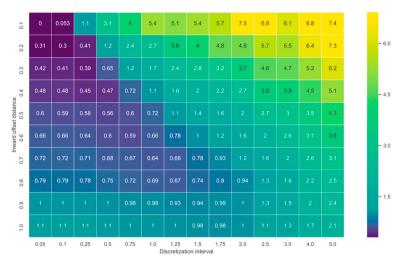Three of the five steps listed above, namely inward offset distance

Fig. 7. *The mean deviation (%) of perimeter of each cell* for each combination of inward offset distance (vertical axis) and discretization interval (horizontal axis).
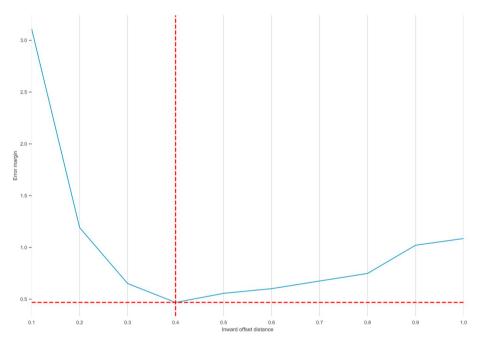

Fig. 8. Relation of inward offset distance and error margin, showing that for 0.4 m, the error margin reaches its minimum.

(step 1), discretization interval (step 2) and clipping method (step 5), require setting parameters that can have a significant effect on the resulting tessellation. As such these need to be evaluated in greater detail. More specifically, in the case of inward offset distance (step 1), the selection of too large values may cause the collapse of narrow parts of building shapes and loss of detail, while too small ones may generate unwanted "*saw-like*" geometries between adjacent buildings. Similarly, a large discretization interval (step 2) may produce the same "*saw-like*" geometry issue, whilst the opposite would increase exponentially computational demand (Fig. 4). Additionally, since the two parameters are interlinked, their individual effect on the shape of each cell is not independent: as such, their combined effect needs to be balanced to generate geometries with insignificant shape deviation and minimum computational burden. Finally, the adoption of a clipping method for the tessellation (step 5) also requires considerations in order to appropriately limit the focus of the analysis to the urbanized footprint and exclude large open un-built spaces while limiting potential MAUP effects (Openshaw, 1984). Due to the importance of correctly setting

these parameters, Section 4.2 will discuss the adopted method for the determination of inward offset distance (step 1), discretization interval (step 2) and method.

The conceptual sequence described in this section, was translated into a Python code, building its key parts on the capability of SciPy (Jones, Oliphant, Peterson, et al., 2001), Shapely (Gillies, 2007) and GeoPandas (Jordahl, Van der Bossche, Wasserman, et al., 2019). Computation was run on Ubuntu Bionic 18.04 running at Amazon Web Services EC2. The resulting Python script is released as part of the open source Python package named "*momepy: Urban Morphology Measuring Toolkit*" (Fleischmann, 2019).

4. Morphological tessellation and plots: data and comparison method

4.1. The dataset

The MT method and the informational value of MCs compared to

Fig. 9. Morphological tessellation cells as generated across four different areas of Zurich; 4a) organic tissue of Niederdorf; 4b) compact tissue of Langstrasse; 4c) detached villas of Hottingen; 4d) mixed post-war development of Friesenberg.

plots were tested within the administrative area of Zurich (Fig. 5). This was chosen for its historically characterised and heterogeneous urban fabric as well as for the availability of the *'Amtliche Vermessung'* dataset,[3] a freely-accessible resource containing high-quality information on cadastral plots and building footprints. Before generating the MT, data was cleaned as follows:

a) From the *cadastral layer,* which covers the 100% of study area, all features not containing buildings (e.g. streets or large open spaces) were removed, as they do not represent built-up form;

b) From the *building layer,* features smaller than 30 m^2 were filtered out, as such smaller objects are likely ancillary structures rather than actual buildings.

4.2. Definition of MT Parameters: inward offset distance, discretization interval and clipping method

To determine the optimal setting for inward offset distance (step 1) and discretization interval (step 2), a test was run on a portion of the Langstrasse area in Zurich (Fig. 5), a heterogeneous fabric predominantly characterised by adjacent buildings (significantly more prone to error than isolated buildings) limited by the street network.[4] We considered several combinations of inward offset (from 0.1 to 1 m) and discretisation interval (from 0.05 to 5 m) and evaluated them against the most precise setting (0.1 / 0.05), which provides the highest-resolution tessellation with minimal effect on the building shape. We then assessed deviation of cell perimeter and area values for each combination, as well as its computational demand: the latter is a function of the number of discretization points, as these directly impact

---

[3] The Amtliche Vermessung dataset is accessible from the municipal GIS open data portal (https://maps.zh.ch). From it can be extracted the cadastral layer (Liegenschaften_Liegenschaft_Area) and the layer of buildings (all features named 'Gebäude').

[4] It would not be possible to run this analysis on the whole area of Zurich due to the computational demands of the high-resolution tessellation (0.1 m offset and 0.05 m discretization intervals).
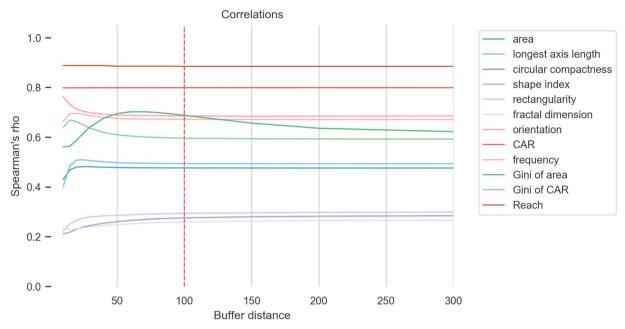
Fig. 10. Spearman's rho rank correlation between cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi- building plots are found in the supplementary materials).

Fig. 11. NRMSD of cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi- building plots are found in the supplementary materials).

Fig. 12. Example of LISA patterns of Frequency measured on the MT.

Fig. 13. LISA accuracy of cadastral values and each of the selected buffers of tessellation based on the whole dataset (figures for single and multi- building plots are found in the supplementary materials).
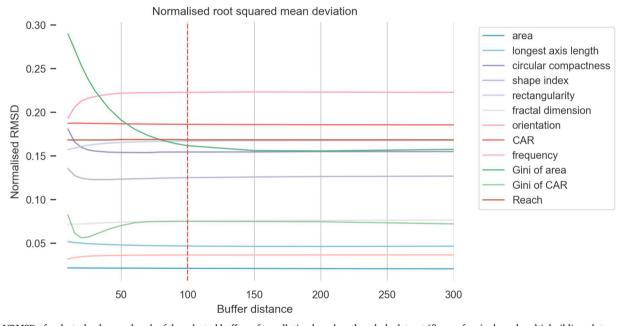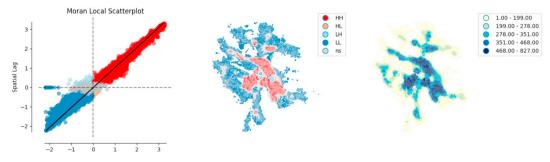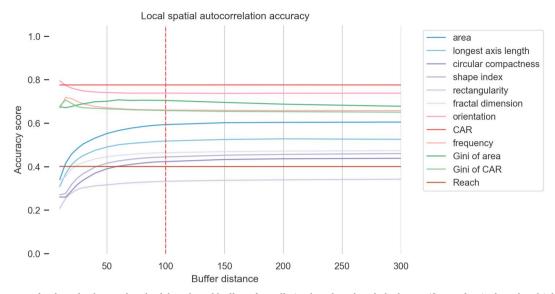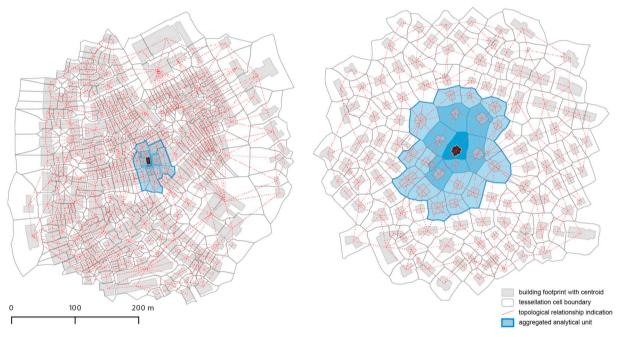

Fig. 14. Relationship between morphological cells of topological distance 2: the red geometry represents the adjacency network of neighbouring elements (buildings, MC) at topological distance 1 (adjacent neighbour of first order), while blue geometry represents the boundary of the aggregated analytical unit of topological distance 2 for each of the highlighted buildings. In the image, a fabric characterised by fewer and sparser buildings (b) generate larger cells and aggregated units compared to a denser and more compact fabric (a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on memory demand. The result of this test is presented in Section 5.1. Based on it, the optimal combination of the two parameters was adopted to generate the MT in later stages.

Finally, in order to clip the tessellation (step 5) we adopted a definition of urban footprint aligned with Angel and Parent (2007); Angel, Arango Franco, Liu, et al. (2018), and limited the study area extent by setting a 100 m buffer from the built-up area. However, to test the robustness and stability of the buffer and avoid arbitrary selection, we also tested 14 other buffers, ranging from 10 to 300 m. The stability of the 15 buffer distances is discussed alongside the comparative analysis in Section 5.2.

4.3. The informational value of MT vs plots: the 12 morphometric characters

To test the informational value of the MCs compared to plots, 12 morphometric characters (Table 1) grouped into the six morphometric categories (dimension, shape, spatial distribution, intensity, connectivity and diversity)[5] proposed by Fleischmann (2017), are selected and measured on both the cadastral layer and MT layer, at the 15 buffer

---

[5] In the current study only a limited number of connectivity-based characters was considered, due to their focus on the street network, which is not particularly significant in the comparison between MCs and plots.

Table 1
Selection of morphometric characters used for comparison.

Category	Character	Formula	Reference
Dimension	Area	$area$	(Song & Knaap, 2003)
	Longest Axis Length (LAL)	$\max\{d_1, d_2, ..., d_n\}$*	
Shape	Circular Compactness	$\dfrac{area}{area\ of\ enclosing\ circle}$	(Dibble, 2016)
	Shape Index	$\dfrac{\sqrt{\frac{area}{\pi}}}{0.5 * \text{LAL}}$	(Steiniger, Lange, & Burghardt, 2008)
	Rectangularity	$\dfrac{area}{area_{MBR}}$**	(Dibble, 2016)
	Fractal Dimension	$\dfrac{2\log\left(\frac{perimeter}{4}\right)}{\log(area)}$	(McGarigal & Marks, 1995)
Spatial Distribution	Orientation	$\begin{cases} azimuth_{MBR} ***, & azimuth_{MBR} < 45° \\ azimuth_{MBR} - 2(azimuth_{MBR} - 45°), & azimuth_{MBR} \geq 45° \end{cases}$	
Intensity	Frequency	$\sum_{dist=1}^{400} element$	(Hallowell & Baran, 2013)
	Coverage Area Ratio (CAR)	$\dfrac{area_{building}}{area}$	(Schirmer & Axhausen, 2015)
Diversity	Gini Index of Area	$G = \dfrac{\sum_{i=1}^{n}(2i - n - 1)area_i}{n\sum_{i=1}^{n} area_i}$	
	Gini Index of CAR	$G = \dfrac{\sum_{i=1}^{n}(2i - n - 1)CAR_i}{n\sum_{i=1}^{n} CAR_i}$	
Connectivity	Reach Centrality****	$R^r[i] = \| \{j \in G - \{i\} : d[i,j] \leq r\} \|$	(Sevtsuk, 2010)

*$d_1 ... d_n$ are diagonals of convex hull of element.
**MBR is minimum bounding rectangle.
***Azimuth is defined as orientation of axis between 1st and 3rd quadrant
****"The reach centrality, $R^r[i]$, of a building i in a graph G describes the number of other buildings in G that are reachable from i at a shortest path distance of at most r" (Sevtsuk, 2010, p. 9).

distances mentioned in section 3.1. With the exclusion of Reach Centrality, which is measured using the *Urban Network Analysis (UNA) Toolbox* (Sevtsuk & Mekonnen, 2012), all characters are computed using Python scripts released as part of the *momepy* package (Fleischmann, 2019).

Once all morphometric characters are calculated for cadastral plots and the 15 MT layers (at each buffer distance), the similarity of the resulting values for the two datasets is evaluated using three methods: 1) Spearman's rank correlation; 2) Normalised root squared mean deviation (NRSMD) and 3) Accuracy of significant patterns defined by local Moran's I indicator of spatial autocorrelation (LISA) (Anselin, 2010).

Spearman's rank correlation *is "a measure of the correlation between ranks, calculated by using the ranks in place of the actual observations in the formula for the correlation coefficient r"* (Kokoska & Zwillinger, 2000, p. 372) (see Eq. 1) and was used due to non-normality of distribution of measured values. It ranges from −1 (negative correlation) to 1 (positive correlation), with values > 0.5 or < −0.5 indicate moderately significant positive or negative correlation (Hinkle, Wiersma, & Jurs, 2003).

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \tag{1}$$

where: $d_i = \text{rg}(X_i) - \text{rg}(Y_i)$ is the difference between the rank of observed and expected value and n is the number of observations (Kokoska & Zwillinger, 2000).

NRSMD is a frequently used measure of *"an estimate of the standard deviation of residuals from the model"* (Alexander, Tropsha, & Winkler, 2015, p. 5) normalised by the range (see Eq. 2), and it is used to measure the difference between the expected and observed values, normalised by the range. As a ratio of deviation, it ranges from 0 to 1, where 0 means no deviation and 1 means deviation equal to the range of values. As the range is sensitive to outliers, NRSMD might not be relevant for characters of *Dimension* category.

$$\text{NRMSD}(y,\widehat{y}) = \frac{\sqrt{\text{MSE}(y,\widehat{y})}}{y_{\max} - y_{\min}} \tag{2}$$

where:

$$\text{MSE}(y,\widehat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \widehat{y_i})^2$$

where $y, \widehat{y}$ are observed and expected values.

Accuracy is *"closeness of computations or estimates to the exact or true values that the statistics were intended to measure"* (OECD, 2006) and is here used to measure the similarity of significant spatial clusters identified from the cadastral layer and those identified from each version of the tessellation (see Eq. 3). Since studies in Urban Morphometrics are more interested in uncovering *recurrent patterns* in urban form rather than *actual values* (Feliciotti, Romice, & Porta, 2018), this method is probably the most relevant of the three. In fact it measures whether corresponding features from both datasets (cadastral plots and morphological cells in this case) significantly fall within the same cluster (i.e. $p \leq 0.05$), with values ranging from 0 (no match) to 1 (perfect match).

$$\text{aLISA} = \frac{SC_{match}}{SC_{max}} \tag{3}$$

where

SC_{match} is the number of the elements belonging to the same significant cluster (HH, HL, LH, LL) in both $y, \widehat{y}$ and SC_{max} is the number of the elements $\widehat{y}$ belonging to any significant cluster.

The adjacency matrix used for LISA represents 200 m Euclidean distance from each building.

It must be noted that, for the statistical comparison of selected morphometric characters across the MT layers and the cadastral layer, these must correspond perfectly. However, whilst there is a 1:1 match between MCs and *buildings*, the same does not apply to MCs and plots, as the latter may contain one building (single-building plots) or more than one (multi-building plot). To resolve this issue, the building layer is used as a *proxy* between tessellation and cadastre and, therefore, all morphometric characters computed on both MCs and plots are associated to the building layer (i.e. each building is linked to the value of its MC and of the plot it sits on). However, to better understand the impact of *'single-building'* and *'multi-building'* plots (79% and 21% of all plots respectively), the three methods described above are applied to the *whole dataset* and, separately, for *single-building* and *multi-building*

plots. In particular, we expected that multi-building plots, although important for their effect on the overall analysis, would hold limited comparative value for most of the assessed morphometric characters (perhaps with the only exclusion of covered area ratio and Gini index of CAR, which capture compatible concepts).

5. Results & discussion

5.1. Determination of optimal parameters of the MT algorithm

The test performed on the selected inward offset ranges (from 0.1 to 1 m) and discretisation intervals (from 0.05 to 5 m) allowed to assess computational demand (i.e. number of discretization points) and deviation of cell perimeter and area for each combination. In terms of computational demand, as shown in Fig. 6, it appears that the discretisation segment length has an exponential effect on the number of generated points. For values below the mean (tail of the distribution), computational demand remains relatively stable, whilst for higher values (head of the distribution) it grows sharply, more than doubling at each step. Discretization intervals ≥ 0.5 m are therefore preferred as more computationally effective.

The effect of inner buffer and discretization interval on the deviation of the MC's area compared to the high-resolution tessellation is insignificant for all tested combinations (0.00 and 0.01%), showing that, no matter the parameters, results are stable. In turn, the same effect on the MC's perimeter is more pronounced (Fig. 7) due to the aforementioned phenomenon of "*saw-like*" geometries (see section 3.1) with percent deviation ranging from 0.05% to 7.4%. Focusing on the 0.5 m discretisation interval, providing the balance between the MC shape detail and computational demands, deviation values range from 0.47% to 3.1% (Fig. 8). This suggests that the combination of 0.5 m metres discretisation interval and 0.4 m inward offset distance provides the optimal balance in terms of effectiveness of computation and minimisation of error. These values are hence adopted as parameters in the computation of the MT in the next stages.

5.2. Comparison between the cadastral layer and morphological tessellation

Having determined the optimal combination of the tessellation parameters (inward offset = 0.4 m and discretisation intervals = 0.5 m), the MT for Zurich is computed using *momepy* " (Fleischmann, 2019). From a first visual inspection of the generated layer, it is already possible to appreciate how the MT is able to nicely capture variations in size, grain and compactness of buildings (Fig. 9), not dissimilar what observed in a typical cadastral layer. We subsequently calculate the 12 morphological characters in Table 1 for cadastral plots and tessellation cells. In the next section, their correlation at each buffer of tessellation is studied.

5.2.1. Spearman's rank correlation

Using Spearman's rank correlation, we find that correlation of measured characters ranges between 0.25 (fractal dimension) to 0.89 (reach), with differences between morphometric categories and between *single-* or *multi-building* plots. Characters in the *Shape* category exhibit the worst performance, with insignificant correlation for the whole sample (~0.27) and multi-building plots (~0.09) and low significance for single-building plots (~0.42). This result was expected, due to the intrinsically different geometry of the two spatial units (MC and plots) and to the existence of multi-building plots.

Dimension characters inherently differ between *multi-* and *single-* building plots, showing only low significance for the former (~0.35, ~0.4) and high significance for the latter (~0.83, ~0.7). Remaining characters show moderate or high significance for all samples, with higher values for single-building plots (Fig. 10 and Table 2). Results for all buffers are quite consistent, with fluctuations observed only at smaller distances (< 50 m), indicating stability of the selected value of 100 m.

Table 2

Spearman's rank correlation of the whole dataset, single-building plots and multi-building plots at 100 m buffer (emphasis reflects significance of correlation).

Category	Character	All	Single-building plots	Multi-building plots
Dimension	Area	*0.4767*	**0.8273**	*0.3583*
	Longest Axis Length (LAL)	*0.4943*	**0.7055**	*0.4073*
Shape	Circular Compactness	0.2758	*0.4203*	0.0864
	Shape Index	0.2758	*0.4203*	0.0864
	Rectangularity	0.2940	*0.4040*	0.1214
	Fractal Dimension	0.2593	*0.4407*	0.0360
Spatial Distribution Intensity	Orientation	**0.6859**	*0.7985*	0.5713
	Frequency	**0.7995**	*0.9103*	**0.7093**
	Coverage Area Ratio (CAR)	0.6721	*0.7649*	0.5567
Diversity	Gini Index of Area	0.6882	*0.7291*	0.6312
	Gini Index of CAR	0.5963	0.6263	0.5551
Connectivity	Reach	**0.8851**	*0.9371*	**0.8282**

Table 3

NRMSD for the whole dataset, single-building plots and multi-building plots at 100 m buffer.

Category	Character	All	Single-building plots	Multi-building plots
Dimension	Area	0.0213	0.0075	0.0326
	Longest Axis Length (LAL)	0.0469	0.0162	0.0645
Shape	Circular Compactness	0.1545	0.1270	0.1788
	Shape Index	0.1252	0.1000	0.1479
	Rectangularity	0.1671	0.1563	0.1773
	Fractal Dimension	0.0754	0.0566	0.0970
Spatial Distribution Intensity	Orientation	0.2229	0.1775	0.2601
	Frequency	0.1862	0.1507	0.2163
	Coverage Area Ratio (CAR)	0.0366	0.0432	0.1224
Diversity	Gini Index of Area	0.1618	0.1509	0.1724
	Gini Index of CAR	0.0752	0.0691	0.0838
Connectivity	Reach	0.1685	0.1528	0.1828

5.2.2. Normalised RMSD

Overall, the RMSD test indicates a high level of similarity between datasets (Fig. 11 and Table 3), excluding *Dimension* characters which, as mentioned in Section 3.2 are heavily skewed by large outliers, hence not comparable with the rest of the data. Apart from *Orientation*, which is the worst-performing character in the set (~0.22 for the whole dataset, ~0.26 for multi-building plots and ~0.18 for single-building plots), all other characters score RMSD values lower than 0.2 (~0.15 for single-building plots and ~0.18 for multi-building plots). This suggests that, even though the spatial coverage of the morphological tessellation is different from plots, this difference is, in terms of information, only minor. Even the poorer performance of *Orientation* depends more on the way this is measured than on dissimilarity between datasets: unlike other metrics, *Orientation* is calculated as a deviation of the orientation of the longest axis of MBR from cardinal directions in degrees and, as such, it ranges from 0 to 45°. Hence, a deviation of 0.2 corresponds to a difference of only 9°. It is worth noting that for smaller buffers (15 to 40 m) results show high instability, where some characters exhibit the highest correlation values and others the lowest: this confirms that smaller buffers are unsuitable as parameters to limit the tessellation. In turn, the 100-m buffer is confirmed as robust and stable across all characters.

Table 4
aLISA for the whole dataset, single-building plots and multi-building plots at 100 m buffer.

Category	Character	All	Single-building plots	Multi-building plots
Dimension	Area	0.5938	0.6448	0.5090
	Longest Axis Length (LAL)	0.5181	0.6138	0.4028
Shape	Circular Compactness	0.4235	0.5061	0.3319
	Shape Index	0.4449	0.5312	0.3475
	Rectangularity	0.3330	0.3930	0.2761
	Fractal Dimension	0.4644	0.5652	0.3489
Spatial Distribution	Orientation	0.7389	0.8055	0.6711
Intensity	Frequency	0.7763	0.8240	0.7318
	Coverage Area Ratio (CAR)	0.6610	0.7313	0.5908
Diversity	Gini Index of Area	0.7050	0.7333	0.6759
	Gini Index of CAR	0.6586	0.6742	0.6423
Connectivity	Reach	0.4007	0.3363	0.4644

5.2.3. Recognition of significant patterns using LISA

The analysis of patterns with LISA (Fig. 12) captures differences across measured characters; however, given the dissimilarity of the datasets due to multi-building plots, the accuracy scores are not expected to reach values close to 1. Highest pattern similarity is recognised for the *Frequency* character (~0.78, corresponding to an almost 80% match) (Fig. 13 and Table 4), followed by *Orientation* and *Diversity* characters (Gini Index of Area and Gini Index of CAR) and *CAR* (0.74–0.66), while *Dimension* characters are around ~0.5 depending on the sample considered (single- or multi- building plots). *Shape* characters are consistently the ones providing lowest accuracy, apart from *Reach Centrality*, due to the single-building vs. multi-building deviation in the datasets. Overall, the difference between samples is relatively consistent, with single-building plots reaching values between 0.1 and 0.2 higher than multi-building plots. While none of the values indicates an equality of both datasets, some are close enough to be considered as proxies of each other. The effect of buffer distance confirms already observed pattern and the stability of the 100 m buffer.

5.3. Discussion of results

Our results suggest that the proposed method contributes to resolve some of the limitations associated with using the plot as unit of UMM analysis. However, the picture resulting from testing the similarity between cadastral plots and MCs is rather complex.

Notably, the significance of similarity between plots and MCs varies considerably depending on the morphometric character selected: this is generally high for all *Intensity* characters (*Frequency, CAR*), whilst *Shape* characters (*Rectangularity, Circular Compactness, Shape Index, and Fractal Dimension*) report a comparatively poorer performance and a higher deviation. This means that if, for several of the morphometric characters assessed, MT is able to retain plot-level information which is comparable to that provided by the cadastral layer, for other characters MCs are less efficient proxies of plots and capture comparatively different information.

It is also evident that the similarity of datasets is higher across all measured characters for single-building plots compared to multi-building plots. Considering that the former ones are predominantly found in pre-industrial urban tissues while the latter are more typical of modern and contemporary development, it is suggested that MT might be a *better proxy* of plots in the context of traditional fabrics than it is in modernist and contemporary ones.

Overall, there appears to be substantial scope for the MC to be utilised as basic unit of UMM analysis, given its ability to capture meaningful *patterns* of urban form at the plot scale, the degree of

reliability and universality of the underlying method and the wide accessibility of the data required to generate it. Indeed, while the recognition of plots can be very troublesome and resource intensive, morphological tessellation is consistent throughout, since it is only based on building footprint information which is equally present in all kinds of urban areas. Moreover, by using morphological tessellation instead of traditional methods relying on buildings, street networks and plots, data dependency is reduced by a third as the tessellation is generated from the building layer alone.

5.4. Limits of the method

Whilst MT and MC appear to be promising new addition to the tools available for UMM applications, a first limit of this work is that the cadastral parcels in Zurich, which we have loosely treated as 'plots', are solely based on land-ownership. That causes discrepancy between the generated tessellation and the cadastral layer which includes multi-building plots. However, as only 21% of plots are affected and results are reported for both groups, we believe that the presented method is robust enough to provide relevant results.

Second, whilst it is true that a MT can be generated directly from a building layer alone, it cannot be created from *any* building layer, as this needs to comply to certain quality requirements. Notably, since the method sees every feature of GIS layer as an individual input for tessellation, it is important not to have buildings composed of multiple features each representing, for example, different heights or different parts of the same (as in the case of British Ordnance Survey). Similarly, it is important not to have different independent buildings collapsed into a single simplified feature (as in the case of vast portions of Open Street Maps).

6. Conclusions

6.1. Summary of key findings

Plots are commonly seen as the ideal spatial division for morphological analysis, but they also have their drawbacks, causing the limited applicability of plot-based methods and, more importantly, the reduced reliability of results obtained by employing them. The aim of this work is to address some of the issues characterising the definition of plot and plot boundaries, the availability and accessibility of plot data and the labour intensiveness of manually extracting reliable plot-level information, aspects that limit the potential of Urban Morphometrics. The need to objectively define a unit of analysis able to capture the smallest and arguably most fundamental level of spatial subdivision, and to develop a reliable and replicable method to generate and measure it, is the rationale behind the *Morphological Cell (MC)* unit and the *Morphological Tessellation* (MT) method.

The universal and algorithmic nature of the proposed MT has the potential to scale up UMM analysis with minimum effort to the large scale,[6] while significantly reducing the interpretative input of the analysts along the process. This latter property of MT appears to be particularly relevant to making large scale UMM analysis viable and take full advantage of big data in the GIS area. The robustness of the proposed method and the validity of the proposed spatial unit of analysis is verified through the assessment of 12 representative morphometric characters and the application of three different *quantitative* comparative methods, Spearman's correlation, NRMSD and accuracy of LISA, aimed at evaluating the similarity of information between MCs and cadastral plots.

---

[6] Prime examples would be French cities, having high-quality building data provided by national mapping agency, while ownership-based plot data of French cadastre do not hold the properties fit for morphometric analysis (Kropf, 2018).

The MT method, as tested and presented in this paper, offers a different approach to spatial division whilst still capturing a level of quality of information on urban form that is similar to that conveyed by the plot. Findings presented in this paper indicate that there is overlap between the information derived from cadastral plots and the one derived from morphological tessellation. The degree of this overlap depends on the category of morphometric characters and the type of urban context, but for certain types of morphologic analysis it is large enough to consider MCs reliable proxies of plots. At the same time, MCs cannot fully *replace* plots in the understanding and analysis of urban form patterns.

By the same token, it should be pointed out that the MCs do offer added values that are relevant on their own, regardless to their similarity to the plots. These have to do with the potential innovations – yet largely unexplored – which are triggered by the very nature of this geometry. For example, unlike other methods of urban form partitioning, the MT covers uniformly the totality of space within the set study area, allowing to capture the topology of contiguous space at the plot-level. Indeed, since all MCs are determined by adjacency, by using MT it is possible to think in terms of topological distance (set number of topological steps between cells) rather than geographic distance (set metric distance around elements, either "as the crow flies" or along the street network). Moreover, thinking in terms of topological distance as opposed to metric, the MT can be used to define new aggregated analytical units that are able to capture the immediate area of influence of a building *on* its surrounding fabric and, at the same time, of the surrounding fabric *on* the building. Indeed, since the size of each MCs depends on the granularity of the urban structure, the spatial representation of a set topological distance would be far smaller for a MC located in a fine-grained built-up area than for the same located in a coarse one (Fig. 14). Crucially, this is a kind of information that would not be possible to access with plots alone, which allows to reframe the very idea of 'proximity' by rethinking the relationship between scale and spatial meaning, thereby enhancing the ability to capture the context in UMM analysis.

6.2. Further research

Keeping in mind the novelty of the method proposed, future work should focus on testing its generalisability beyond the case presented in this paper, as well as over a wider selection of case studies and urban tissues. Another promising development area, as briefly anticipated in Section 6.1, is the exploration of the implications of this new geometry in capturing unique configurational and geometrical properties of different types of urban form and understanding the level of difference and similarity between heterogeneous morphological patterns.

In addition to this, and to increase the value of the proposed method, there is clear scope for expanding the assessment framework adopted by employing new characters, this time selected *uniquely* thinking at how best express the intrinsic potentials of the tessellation *in itself,* rather than merely as a potential *surrogate* of plots. To this regard there is a vast untapped potential in the combination of MT and VGIS (i.e. Open Street Map), which could vastly extend the applicability of UMM models to potentially any human settlement in the world.

Acknowledgements

References

Ai, T., & Zhang, X. (2007). The aggregation of urban building clusters based on the skeleton partitioning of gap space. In J. Jokar Arsanjani,, A. Zipf,, & P. Mooney, (Eds.). *OpenStreetMap in GIScience, Lecture Notes in Geoinformation and Cartography*

(pp. 153–170). Berlin, Heidelberg: Springer Berlin Heidelberg.
Alexander, D. L. J., Tropsha, A., & Winkler, D. A. (2015). Beware of R(2): Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of chemical information and modelling, 55*(7), 1316–1322.
Angel, S., & Parent, J. (2007). Urban sprawl metrics: An analysis of global urban expansion using GIS. *ASPRS* (pp. 2007). Annual Conference: Tampa.
Angel, S., Arango Franco, S., Liu, Y., et al. (2018). *The shape compactness of urban footprints. Progress in Planning.*
Anselin, L. (2010). *Local Indicators of Spatial Association-LISA. Geographical Analysis. 27(2),* John Wiley & Sons, Ltd93–115 10.1111.
Araldi, A., & Fusco, G. (2017). *Decomposing and Recomposing Urban Fabric: The City from the Pedestrian Point of View.* 365–376 (Available from: 10.1007/978–3–319–62407–5).
Araldi, A., & 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), 1243–1263.
Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's open-source street map is more than 80% complete. *PLoS One, 12*(8), 1–20.
Basaraner, M., & Selcuk, M. (2004). *An attempt to automated generalization of buildings and settlement areas in topographic maps.* Istanbul.
Bobkova, E., Marcus, L. H., & Berghauser Pont, M. (2017). *Plot systems and property rights: Morphological, juridical and economic aspects.* Valencia.
Caniggia, G., & Maffei, G. L. (1979). *Compositione architettonica e tipologia edilizia.* Venice, Italy: I Lettura dell'edilizia di base.
Conzen, M. R. G. (1969). *Alnwick, Northumberland: A study in town-plan analysis.* Institute of British Geographers.
Dibble, J. (2016). *Urban Morphometrics: Towards a quantitative science of urban form.* Glasgow: University of Strathclyde.
Dogrusoz, E., & Aksoy, S. (2007). Modeling urban structures using graph-based spatial patterns. *IEEE,* 4826–4829.
Feliciotti, A., Romice, O., & Porta, S. (2018). From system ecology to urban morphology: Towards a theory of urban form resilience. *Paper presented at International Forum on Urbanism.* Spain: Barcelona.
Fleischmann, M. (2017). *A systematisation of attributes for quantitative urban morphology measuring urban form.* (Glasgow).
Fleischmann, M. (2019). momepy: Urban Morphology Measuring Toolkit. *Journal of Open Source Software, 4*(43), 1807. https://doi.org/10.21105/joss.01807.
Gillies, S. (2007). others*Shapely: manipulation and analysis of geometric objects* Toblerity.org. Available from: https://github.com/Toblerity/Shapely.
Hallowell, G. D., & Baran, P. K. (2013). Suburban change: A time series approach to measuring form and spatial configuration. *The Journal of Space Syntax, 4*(1), 74–91.
Hamaina, R., Leduc, T., & Moreau, G. (2012). Towards urban fabrics characterization based on buildings footprints. *Bridging the Geographic Information Sciences, Lecture Notes in Geoinformation and Cartography* (pp. 327–346). Berlin, Heidelberg: Springer Berlin Heidelberg (Available from: https://link.springer.com/chapter/10.1007/978-3-642-29063-3_18).
Hamaina, R., Leduc, T., & Moreau, G. (2013). A new method to characterize density adapted to a Coarse City model. In J. Jokar Arsanjani,, A. Zipf,, & P. Mooney, (Eds.). *OpenStreetMap in GIScience, Lecture Notes in Geoinformation and Cartography* (pp. 249–263). Berlin, Heidelberg: Springer Berlin Heidelberg.
Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioural sciences.* Houghton Mifflin College Division.
Huijboom, N., & Van den Broek, T. (2011). Open data: An international comparison of strategies. *European Journal of ePractice. 12*(1), 4–16.
Jones, E., Oliphant, T., Peterson, P., et al. (2001) SciPy. Available from: http://www.scipy.org/.
Jordahl, K., Van der Bossche, J., Wasserman, J. et al. (2019) geopandas/geopandas: v0.5.1. Zenodo.
Kokoska, S., & Zwillinger, D. (2000). *CRC standard probability and statistics tables and formulae.* Hoboken, NJ: Taylor and Francis.
Kropf, K. (1997). *When is a plot not a plot: Problems in representation and interpretation.* Birmingham.
Kropf, K. (2018). Plots, property and behaviour. *Urban Morphology, 22*(1), 5–14.
Li, Z., Yan, H., Ai, T., et al. (2004). Automated building generalization based on urban morphology and Gestalt theory. *International Journal of Geographical Information Science, 18*(5), 513–534.
Liu, Y., Guo, Q., Sun, Y., et al. (2014). In R. Ji (Vol. Ed.), *A Combined Approach to Cartographic Displacement for Buildings Based on Skeleton and Improved Elastic Beam Algorithm. 9(12)*PLoS ONE (e113953–27).
McGarigal, K., & Marks, B. (1995). *FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure.*
Moudon, A. V. (1986). *Built for change: Neighborhood architecture in.* San Francisco: MIT Press.
OECD (2006) OECD Glossary of Statistical Terms. Available from: https://stats.oecd.org/glossary/detail.asp?ID=21 (accessed 15 February 2019).
Openshaw, S. (1984). *The Modifiable Areal Unit Problem.*
Panerai, P., Castex, J., Depaule, J.-C., et al. (2004). *Urban forms: The death and life of the urban block.* (Routledge).
Porta, S., & Romice, O. (2014). Plot-based urbanism: Towards time-consciousness in place-making. In C. Mäckler, & W. Sonne (Eds.). *Dortmunder Vorträge zur Stadtbaukunst [Dortmunder Lectures on Civic Art]* (pp. 82–111). Niggli: Sulgen, DE.
Schirmer, P. M., & Axhausen, K. W. (2015). A multiscale classification of urban morphology. *Journal of Transport and Land Use, 9*(1), 101–130.
Schirmer, P. M., & Axhausen, K. W. (2019). A multiscale clustering of the urban morphology for use in quantitative models. In L. D'Acci (Ed.). *The Mathematics of Urban Morphology* (pp. 355–382). Cham: Springer International Publishing.

Sevtsuk, A. (2010). *Path and place: A study of urban geometry and retail activity in Cambridge and Somerville.* MA. Massachusetts Institute of Technology.

Sevtsuk, A., & Mekonnen, M. (2012). Urban network analysis. A new toolbox for ArcGIS. *Revue internationale de géomatique, 22*(2), 287–305.

Song, Y., & Knaap, G.-J. (2003). New urbanism and housing values: A disaggregate assessment. *Journal of Urban Economics, 54*(2), 218–238.

Steiniger, S., Lange, T., Burghardt, D., et al. (2008). *An Approach for the Classification of Urban Building Structures Based on Discriminant Analysis Techniques. Transactions in GIS. 12(1)*, Wiley/Blackwell (10.1111)31–59.

Usui, H. (2018). Statistical distribution of building lot frontage: application for Tokyo downtown districts. *Journal of Geographical Systems, Springer Berlin Heidelberg, 20*(3), 295–316.

Usui, H., & Asami, Y. (2013). Estimation of mean lot depth and its accuracy. *Journal of the City Planning Institute of Japan, 48*(3), 357–362.

Usui, H., & Asami, Y. (2017). Size distribution of urban blocks in the Tokyo Metropolitan Region: Estimation by urban block density and road width on the basis of normative plane tessellation. *International Journal of Geographical Information Science, Taylor & Francis, 32*(1), 120–139.

Usui, H., & Asami, Y. (2019). Size distribution of building lots and density of buildings and road networks: Theoretical derivation based on Gibrat's law and empirical study of downtown districts in Tokyo. *International Regional Science Review, 10* (016001761982627–25).

Whitehand, J. W. R. (1981). Background to the urban morphogenetic tradition. In J. W. R. Whitehand (Ed.). *The urban landscape: Historical development and management* (pp. 1–24). London: Academic Press.

Yu, W., Ai. T., Liu, P., et al. (2017) The analysis and measurement of building patterns using texton co-occurrence matrices. International Journal of Geographical Information Science, Taylor & Francis 31(6): 1079–1100.