XXVIII International Seminar on Urban Form
ISUF2021: URBAN FORM AND THE SUSTAINABLE AND PROSPEROUS CITIES
29th June – 3rd July 2021, Glasgow

Multi-scale gridded urban morphometrics for settlement classification and population mapping

Dr. Warren C. Jochem[1], Professor Andrew J. Tatem[2]

[1] WorldPop, School of Geography and Environmental Science, Faculty of Environmental and Life Sciences, University of Southampton, UK

[2] WorldPop, School of Geography and Environmental Science, Faculty of Environmental and Life Sciences, University of Southampton, UK

Abstract

Urban areas are expanding rapidly around the world, and much of this growth is expected in low- and middle-income countries. Policy makers, researchers, and those implementing development projects need up-to-date and consistent information on cities in order to plan and track progress towards Sustainable Development Goals. Yet in many places experiencing rapid growth, information on urban areas and their population is lacking, outdated or incomplete. In recent years, increasing availability of very high spatial resolution imagery (<1 m resolution) and computing power is enabling sets of building footprint polygons to be automatically extracted from the imagery and mapped for whole countries. These building footprint datasets provide a unique resource to study urban morphometrics in places which may lack other local data. This paper demonstrates the use of a spatial grid to classify urban fabric into settlement types. This unit of analysis is in contrast to plots or parcels which are more commonly used in urban morphology studies, and a case study in Southampton, UK is used to explore the sensitivity of the results to varying the parameters used to define the size of the grid. These initial results suggest that multiple scales of observation windows can be combined to identify key patterns across space and that multiple grid resolutions can give relatively consistent classification results. Future work is needed to explore the use of grids to study urban form in other settings.

Keyword: *morphometrics, settlement classification, building footprint, population*

Introduction

Urban areas are rapidly growing around the world and many of these changes are expected to occur in low- and middle-income countries (UN Habitat, 2016; UN DESA, 2019). Amidst the opportunities of developing cities, the UN has highlighted in the Sustainable Development Goals and the New Urban Agenda the need to ensure that these areas are healthy, safe, and sustainable for the environment and the population (UN General Assembly, 2017). Monitoring changing urban environments and providing up-to-date and spatially-detailed information on settlements and urban populations for local policy makers and planners to meet these sustainability goals is, thus, becoming more important. Remote sensing from satellite and airborne sensors has long been used to monitor changing urban land cover with consistent, repeated measurements over large areas (Herold et al., 2002). More recent advances in higher resolution imagery and computer processing power are now extracting and mapping complete sets of building footprint polygons for whole countries (Sirko et al., 2021). Such datasets provide a new opportunity to study urban form in places which are undergoing rapid urban transitions and which often lack other sources of parcel or cadastral data.

This paper contributes to ongoing discussions in urban morphometric analysis to quantify the patterns in building footprints and to classify these patterns and differentiate settlement types. Specifically, we explore the use of a spatially-defined grid as the unit of analysis for classifying urban form. This choice of unit is in contrast to most urban morphology studies and raises significant questions about its design. The next section provides background on the units of analysis and a more detailed overview of the features of grids and considerations of grid-based morphometric analyses. A case study in Southampton, UK then explores the sensitivity of quantifying urban form and settlement classification to changes in the design of the grid. Finally, we close with a brief discussion about the choice of analysis units in urban morphology.

Background

The plot has often been the preferred unit of analysis for urban morphologists. It represents a foundational element of urban form, yet its definition is challenging and often context-specific (Kropf, 2017). In recent years there has been a growing interest in quantitative methods to study differences in urban form using spatial data layers and plots become difficult to operationalise and use in such morphometric analyses (Fleischmann et al. 2020). A key advancement for morphometrics has been to use Voronoi tessellation (VT) to define spatial units as a substitute for plots. VT completely partitions a planar space into a set of discrete cells around a set of seed points. Hamaina et al. (2012) used Voronoi diagrams constructed around building footprints as consistent units to study patterns in the urban fabric. Other refinements have included the street network to better approximate plots (Usui and Asami 2019) or to offer a pedestrian’s perspective of the built environment (Araldi and Fusco 2019). Fleischmann et al. (2020) examined the details of the tessellation procedure and found similarities in the information content of tessellation cells and cadastral plots, suggesting the suitability of a VT-based approach. Fleischmann et al. (2020) argue that Voronoi cells can describe the “area of influence” (12) around a structure. However, the outer boundary limiting the extent of the tessellation, must be pre-defined by the researcher. Moreover, in sparsely settled areas, such as on the urban fringe or in rural areas, it is not clear that widely separated buildings should have such large areas of influence assigned to them, as assumed in a VT. In contrast to both plots and Voronoi tessellations, we propose a grid-based representation for morphology metrics.

Gridded morphometrics

By a *grid*, we mean an evenly-sized lattice of square cells defined in 2D space over a study area. These types of data structures are common are often referred to as *rasters*. Grids are a very common data structure in remote sensing studies of urban form and land cover (Benza et al. 2016), and they are also used in cellular automata models of cities (Batty et al. 1999). Grids have been used previously to analyse urban form when integrated with climatological models (Long and Kergomard, 2005).

Grids provide both a pragmatic and a principled option for urban morphometrics. From a theoretical perspective, grids are appropriate for representing spatial fields (Couclelis, 1992). While buildings are discrete features, their locations and patterns can be viewed as indicators of local political-economic decisions and historical conditions. Examining how these factors vary across space through changes in morphology, supports a landscape perspective of the built environment, which is more consistent with a continuous spatial field representation. Importantly, grids are defined independently of structures. Regularly-spaced grid cells ensure that empty lots, easements, and rural areas are observed equally compared with densely settled areas. Additionally, examining changes in the urban fabric over time are more feasible since a consistent measurement grid can be maintained. Finally, no additional information on boundaries, plots, or road networks are needed to define the grid, making them well-suited for analyses in places without extensive spatial databases, such as many low- and middle-income countries.

In a gridded morphometric analysis, each cell is a sample unit in which measurements are taken to describe the local landscape of built features. There are three key attributes to define for grids in this case: resolution, scale, and inclusion. *Resolution* refers to the size of the grid cell (also referred to as a pixel), and, effectively, the sampling rate for the calculations. A grid cell which is too small may not capture a broader trend while creating a larger volume of measurements. A grid cell which is too large and important local-scale variation may be missed by the coarse resolution. *Scale* in our approach refers to a window around each grid cell in which building features are observed. This observation window could be restricted to only include features within the grid cell, but a wider window will include additional features which may generate a more complete picture of the local context. These two attributes interact. When small grid cells are used with a large scale window, each window will overlap multiple grid cell locations providing a spatial smoothing effect to the morphometrics. *Inclusion* refers to how to define which features contribute to an observation window. All buildings which touch, are fully contained, or whose centroids are contained are potential options. Other minor attributes include the map projection, alignment, and origin of the grid.

While grids offer certain advantages for morphometric analyses, they require more choices by the researcher. The options to define the grids and observation windows makes them subject to the modifiable areal unit problem (MAUP) (Openshaw, 1984). In the following section we introduce a case study to examine the sensitivity of choice in the main attributes of the grid and potential techniques to address MAUP.

Methodology

This case study focuses on the city of Southampton, UK using building footprints from the Ordnance Survey's OpenMap (Ordnance Survey, 2020). This study area (Figure 1) includes a mix of settlement types with historic areas, large port infrastructure, and terraced and apartment residential areas. Jochem and Tatem (2021) recently presented a set of gridded morphology measurements and unsupervised classification of settlement types for England, Scotland, and Wales. However, they did not report on the details of the choice of grid resolution and analysis windows. This case study expands on that previous analysis.

Figure 1: *Study area of Southampton showing building footprints from Ordnance Survey's OpenMap Local (Contains Ordnance Survey data © Crown copyright and database right 2020).*

First, grids with spatial resolutions of 25, 50, 100, 250, and 500 meters are defined over the study area and are used to calculate average area and count of building footprints intersecting each grid cell. We then apply a range of scales to widen the observation window used to calculate these metrics. We define the observation window as a circle centred on each grid cell's centroid. The radius of this circle is varied from 100 m up to 1000 m in 100 m increments. For the 100 m, 250 m and 500 m resolution grids, the observation scales range from only 200 m, 300 m and 600 m up to 1000 m, respectively. Initial test results showed limited impact from different inclusion criteria. Therefore we only report results where the inclusion is any buildings which intersect (touch) the observation window and focus on the impacts of changing scale and resolution.

Rather than attempting to select a single, best scale for the observation window, we assume each scale can contribute additional information about the urban fabric within a local context defined by the circular buffer. However, using multiple scales of the observation window generates additional output layers and these may include redundant information. We use singular value decomposition (SVD) to extract key patterns in morphology across the scale dimension. The SVD of a matrix **X** is: $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T$, where **U** and **V** are orthogonal vectors and **D** is a diagonal matrix of positive scaled values. In this case, **X** is a data frame of grid cell locations with morphology measurements for each spatial scale. We apply SVD to the output layers of each morphology metric (area and count) and grid resolution separately. We then examine the principal components visually by plotting the first three components of the right singular vector (**V**) as well as by mapping the rotated and transformed principal components of the morphology metrics.

Finally, we use the first three principal components of average building area and building count in an unsupervised classification algorithm to identify areas of similar morphology. We apply the k-means algorithm to each set of layers from the different grid resolutions and compare the sensitivity in the five output maps. We determine the preferred number of groups for each classification by examining the within-sum-of-squares plot, seeking to minimise the total within cluster variation. We recognise that these two metrics are likely insufficient on their own to provide a detailed settlement typology. Limiting the input data for the classification makes the outputs more readily interpretable for this case study and the goal is not to generate the most accurate cluster maps. All analyses are conducted in R (R Core Team, 2020), using the 'foot' package (WorldPop, 2020) which was introduced by Jochem and Tatem (2021).

Results and Discussions

The study area contains 31286 building footprint polygons. Calculations with the 'foot' package generated 94 output layers with different grid resolutions and scales. In general, increasing the resolution of the grid obscures some of the local-scale variations in morphology but still captures the broader trends, while increasing the size of the observation window creates a spatial smoothing effect. The first three principal components of the morphology metrics across scales are shown in Figure 2, for average building area (Panel A) and count (Panel B). The first principal component represents the baseline trend in morphology. The second and third components can be interpreted as deviations from the baseline trends, such as clustering of more structures at particular spatial scales. In general, the grids with resolutions of 25 to 100 m tend to capture similar patterns. Coarser grids (i.e. 250 m and 500 m) stand out as missing some of the variation present in small scale observation windows.

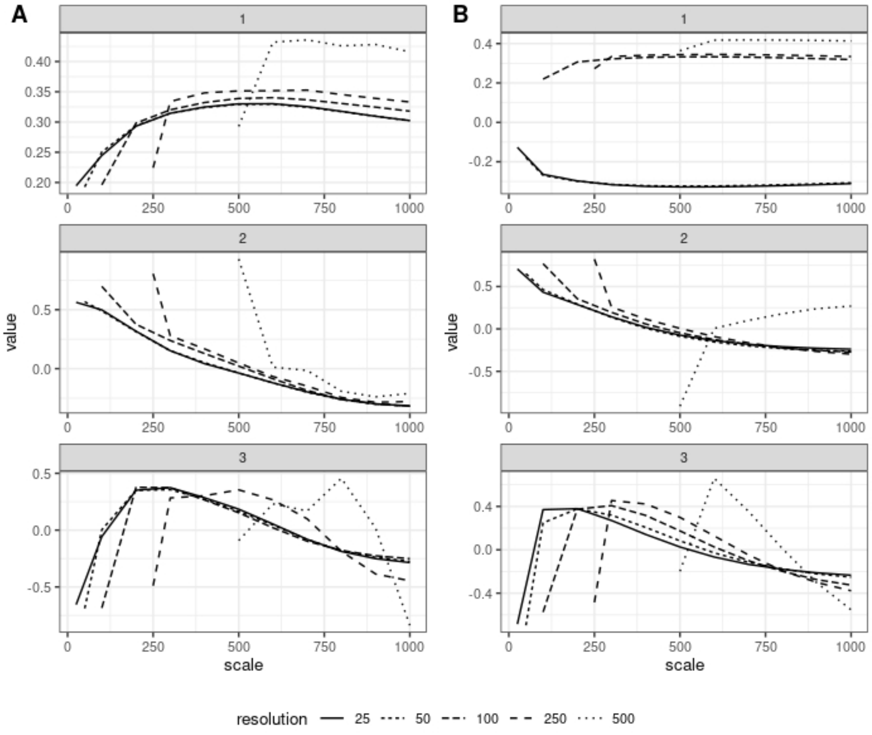

Figure 2: *First three principal components of average building area (Panel A) and building counts (Panel B) across varying scales of observation windows for five different grid resolutions. Scales and resolutions are in meters.*

Minimising the total within sum of squares of the k-means clusters suggested between 5 and 6 classes for all resolutions. To make the outputs more comparable we set the number of classes for all outputs to be 5. The resulting cluster maps are shown in Figure 3. Note that the order of the numbering (and therefore colouring) of the clusters in these maps is arbitrary. These maps highlight the areas within Southampton with similar patterns of the number and size of building footprints. The port and central commercial area, located in southern central part of the study area, are consistently classified distinctly from the main residential areas in the east and west of the city. There is also a notable edge effect in most of the results driven by the study area boundary artificially reducing the number of structures observed along the northern edge of the city.

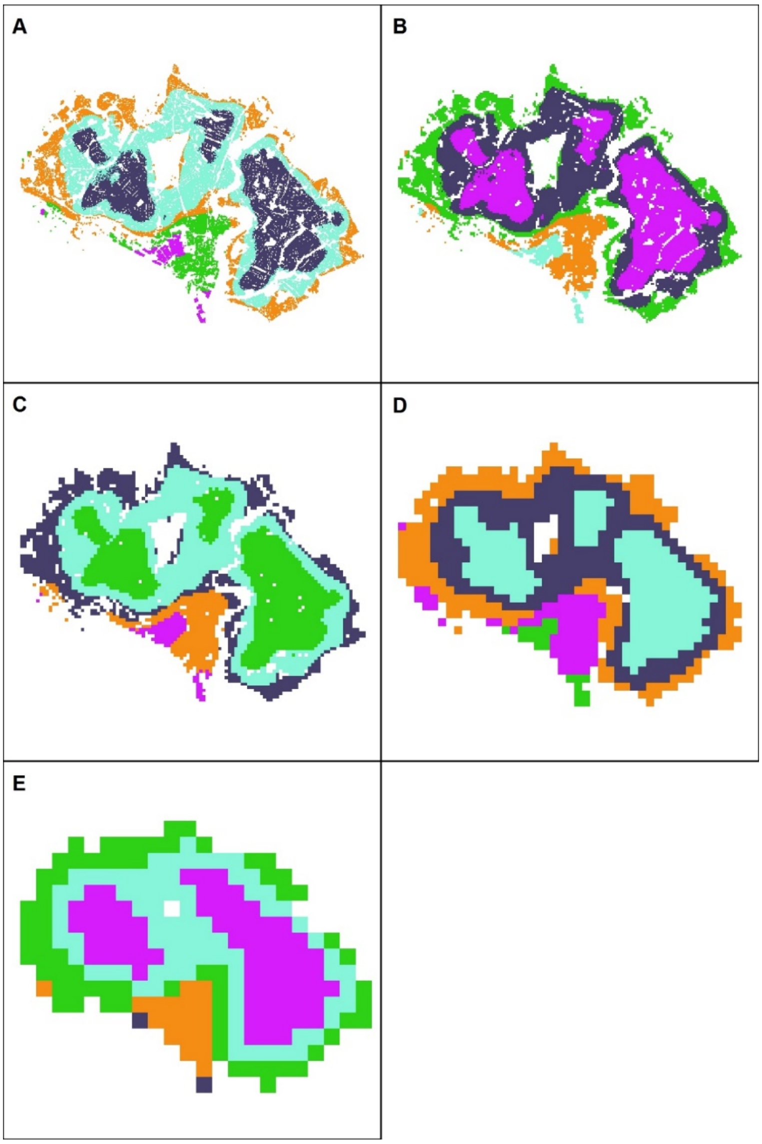

Figure 3: *Results of applying k-means clustering with 5 classes to first three principal components of building area and count for different grid resolutions: 25 m (A), 50 m (B), 100 m (C), 250 m (D) and 500 m (E).*

These initial case study results demonstrate an approach to creating and using gridded morphometric layers of building footprint polygons to create settlement type classifications. In order to use grids, the cell resolution and the scale of the observation window has to be set. These decisions introduce potential issues of MAUP where the findings may be sensitive to the aggregation and configuration of spatial units (Openshaw, 1984). However, we view the decision on the scale of the observation window as an opportunity to derive additional information to define settlement types. We utilised multiple spatial scales and SVD to derive the principal components in the data before clustering. The question of grid resolution, however, remained. In the case study of Southampton, similar trends in morphological patterns across spatial scales were detected in the principal components for all grid resolutions except 250 and 500 m. Despite the varying resolutions, the cluster maps derived using k-means appeared surprisingly consistent in the spatial location of settlement types. Coarser resolution grids did produce less-precise and cleanly defined boundaries between classes, but they identified similar spatial clusters compared with the clustering results using finer resolution grids. When setting the resolution for this type of analysis, other factors, such as the overall size of study area or computational limitations from fine resolution grids, should also be considered. The results suggest that grids can be used in urban morphometric studies; however, future work is still needed to test the sensitivity of these settings in other settings and different urban forms.

Conclusions

Plots, parcels, street-blocks and similar spatial units have typically been used for urban morphology studies. We have proposed an alternative approach using a spatial grid as the unit of analysis to calculate urban morphometrics and classify settlement types. Grids provide a consistent measurement frame across space and time. They offer advantages particularly in settings which lack detailed maps of plots or parcels, and they lend themselves toward a continuous landscape perspective of the urban fabric. Grids are common in other spatial analysis domains such as Geography and in remote sensing, and our experiences in these areas makes it easier to integrate the results of gridded morphology into studies of spatial demographics and other applications. Ultimately, there is not a single, perfect unit for analysing urban form in all cases. The choice of spatial unit should be guided by the research question, the particular application in a specific location.

Acknowledgements

Funding support for this work comes from the Bill and Melinda Gates Foundation and the United Kingdom Foreign, Commonwealth & Development Office as part of the Geo-Referenced Infrastructure and Demographic Data for Development project (GRID3) (OPP1182425).

References

1. Araldi, A. and Fusco, G. (2019) ‘From the street to the metropolitan region: Pedestrian perspective in urban fabric analysis’, *Environment and Planning B: Urban Analytics and City Science* 46, 1243-1263.
2. Batty, M., Xie, Y. and Sun Z. (1999) ‘Modeling urban dynamics through GIS-based cellular automata’, *Computers, Environment and Urban Systems* 23, 205-233.
3. Benza M., Weeks J. R., Stow D. A., Lopez-Carr, D. and Clarke, K. (2016) ‘A pattern-based definition of urban context using remote sensing and GIS’, *Remote Sensing of Environment* 183, 250–264.
4. Couclelis, H. (1992) ‘People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS’, In: Frank A.U., Campari I. and Formentini U. (Eds.) *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes in Computer Science* (Springer, Berlin).

5. Fleischmann, M., Feliciotti, A., Romice, O. and Porta, S. (2020) 'Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale', *Computers, environment and Urban Systems* 80, 101441.
6. Hamaina, R., Leduc, T. and Moreau, G. (2012) 'Towards urban fabrics characterization based on building footprints', In Gensel, J. (Ed.) *Bridging the Geographic Information Sciences: Lecture Notes in Geoinformation and Cartography* (Springer-Verlag, Berlin).
7. Herold, M., Scepan, J. and Clarke, K. C. (2002). 'The use of remote sensing and landscape metrics to describe structures and changes in urban land uses.' *Environment and Planning A: Economy and Space* 34 (8), 1443-1458.
8. Jochem, W. C. and Tatem, A. J. (2021) 'Tools for mapping multi-scale settlement patterns of building footprints: An introduction to the R package *foot*', *PLOS ONE* 16, e0247535.
9. Kropf, K. (2017) 'Plots, property and behaviour', *Urban Morphology* 22, 1-10.
10. Long N. and Kergomard C. (2005) 'Classification morphologique du tissu urbain pour des applications climatologiques', *Revue Internationale de Géomatique* 15, 487–512.
11. Openshaw, S. (1984) 'The modifiable areal unit problem', *Concepts and Techniques in Modern Geography* 38, 3-41.
12. Ordnance Survey, OS OpenMap – Local 2020. Available at: https://osdatahub.os.uk/downloads/open/OpenMapLocal.
13. R Core Team (2020). *R: A language and environment for statistical computing. R Foundation for Statitsical Computing* (Vienna, Austria).
14. Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M. and Quin, J. A. (2021). 'Continental-scale building detection from high resolution satellite imagery', arXiv:2107.1283.s
15. UN Department of Economic and Social Affairs [UN DESA] (2019) *World Population Prospects: Highlights*. Available at: https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf.
16. UN General Assembly (2019) *New Urban Agenda: Res: 71/256* (United Nations, New York).
17. UN Habitat (2016) *Urbanization and Development: Emerging Futures* (UN Habitat, Nairobi, Kenya).
18. Usui, H. and Asami, Y. (2019) 'Size distribution of building lots and density of buildings and road networks: Theoretical derviation based on Gibrat's Law and empirical study of downtown districts in Tokyo', *International Regional Science Review* 43, 229-253.
19. WorldPop (2020). 'foot: An R package for processing building footprint morphometrics'. Available at: https://wpgp.github.io/foot/.