# Geographical Characterisation of British Urban Form and Function using the Spatial Signatures Framework

**Martin Fleischmann**[1,*] **and Daniel Arribas-Bel**[1]

[1]Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool, Roxby Building , 74 Bedford St S , Liverpool , L69 7ZT, United Kingdom
[*]corresponding author(s): Martin Fleischmann (m.fleischmann@liverpool.ac.uk)

## ABSTRACT

The spatial arrangement of the building blocks that make up cities matters to understand the rules directing their dynamics. Our study outlines the development of the national open-source classification of space according to its form and function into a single typology. We create a bespoke granular spatial unit, the enclosed tessellation, and measure characters capturing its form and function within a relevant spatial context. Using two-level K-Means clustering of individual enclosed tessellation cells, we generate a classification of space for the whole of Great Britain. Contiguous enclosed tessellation cells belonging to the same class are merged forming spatial signature geometries and their typology. We identify 16 distinct types of spatial signatures stretching from wild countryside, through various kinds of suburbia to types denoting urban centres according to their regional importance. The open data product presented here has the potential to serve as boundary delineation for other researchers interested in urban environments and policymakers looking for a unique perspective on cities and their structure.

## Background & Summary

How the building blocks that make up cities are spatially arranged is worth quantifying and understanding. By "building blocks", we mean both the activities and agents that inhabit cities, as well as the (infra)structure that supports them. The former can be conceptualised as *urban function*, while the latter falls under the study of *urban form*. Understanding urban form and function is important for two main reasons. First, the combination of both *encodes* rich information about the history, character and evolution of cities. For example, the shape and properties of the street network encode the technology of the time (e.g., automobile); while the degree of mix in land uses can reflect cultural values. Second, the spatial pattern of urban form and function also acts as a frame that *influences* a variety of outcomes, from economic productivity to socio-economic cohesion to environmental sustainability.

In this paper, we use the Spatial Signatures framework[1,2], which develops a “characterisation of space based on form and function designed to understand urban environments”[1]. Spatial Signatures are theory-informed, data-driven computable classes that describe the form and function of a consistent patch of geography. Figure 1 presents an overview of the development of a spatial signature classification. We build a series of enclosures that we combine with building footprints to further subdivide geographical space into what we call enclosed tessellation cells (ETCs). We then attach form and function characters to each of these subdivisions, and use those to group them into consistent and differentiated classes we call signatures. Each phase is expanded in detail in the next section.

We introduce an open data product (ODP[3]) containing a classification of spatial signatures for Great Britain (illustrated in a figure 2). In doing so, we provide an analysis-ready layer that brings urban form and function consistently, in detail, and at national scale. To the best of our knowledge, this is the first dataset capturing urban form and function published both with a degree of detail and scale as ours. Our results are based on the analysis of more than 14 million of ETCs, to each of which we attach more than 300 characters capturing a wide range of aspects relating to urban form and function. We provide access to both granular geographical boundaries of the delineated spatial signatures as well as measurements for each character at the signature level. The ODP also includes a web map that allows exploration without any technical requirement other than a web browser, and we have open sourced all the code, including details on the computational backend. The uniqueness of our ODP makes it challenging to set up a technical validation as a comparison with existing datasets. Nevertheless, we relate our signatures to a few well-established data products that capture each a subset of the form and function dimensions we consider. Our results are encouraging in that they show broad agreement in expected areas, but also highlight aspects that can only be discovered when considering form and function in tandem.

The approach and outputs presented bring several benefits to a range of stakeholders interested in cities. This spatial signatures ODP provides insight generated from detailed, comprehensive and computationally intensive data analysis and

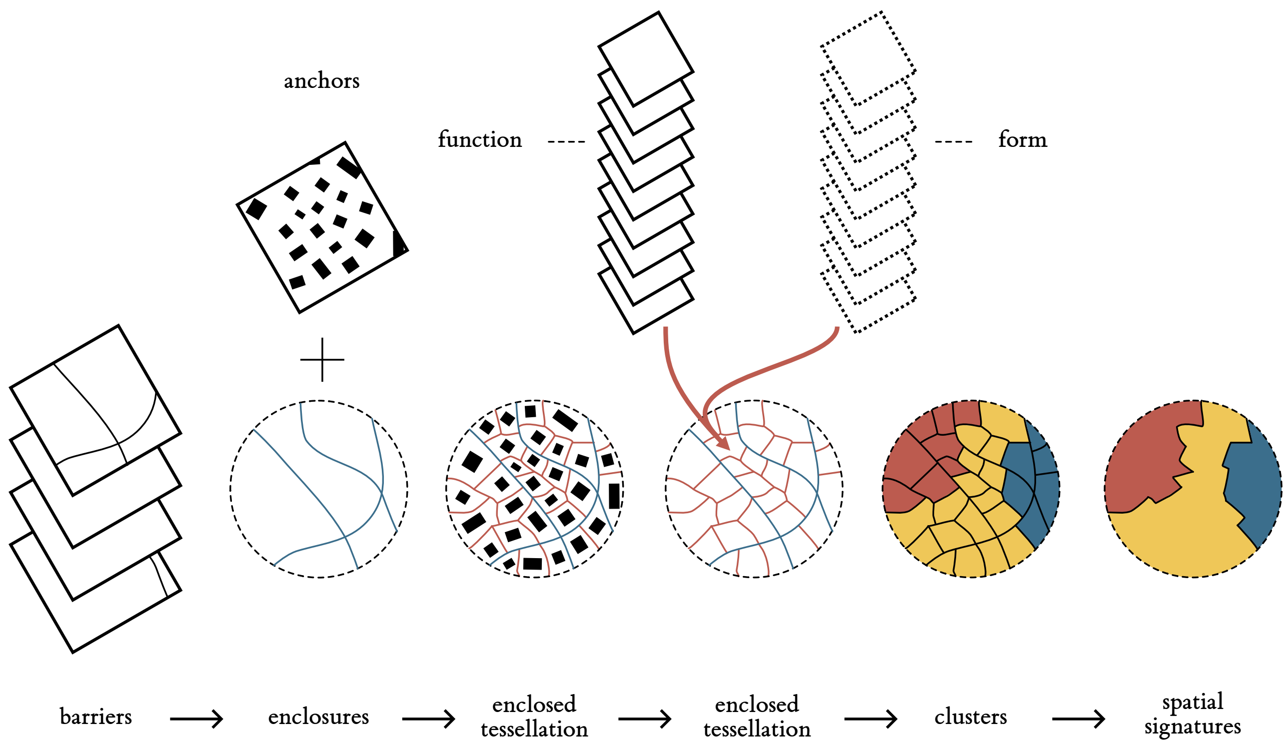


**Figure 1.** Diagram illustrating the sequential steps leading to the delineation of spatial signatures. From a series of enclosing components, to enclosures, enclosed tessellation (ET), the addition of form and function characters to ET cells, and the development of spatial signatures.

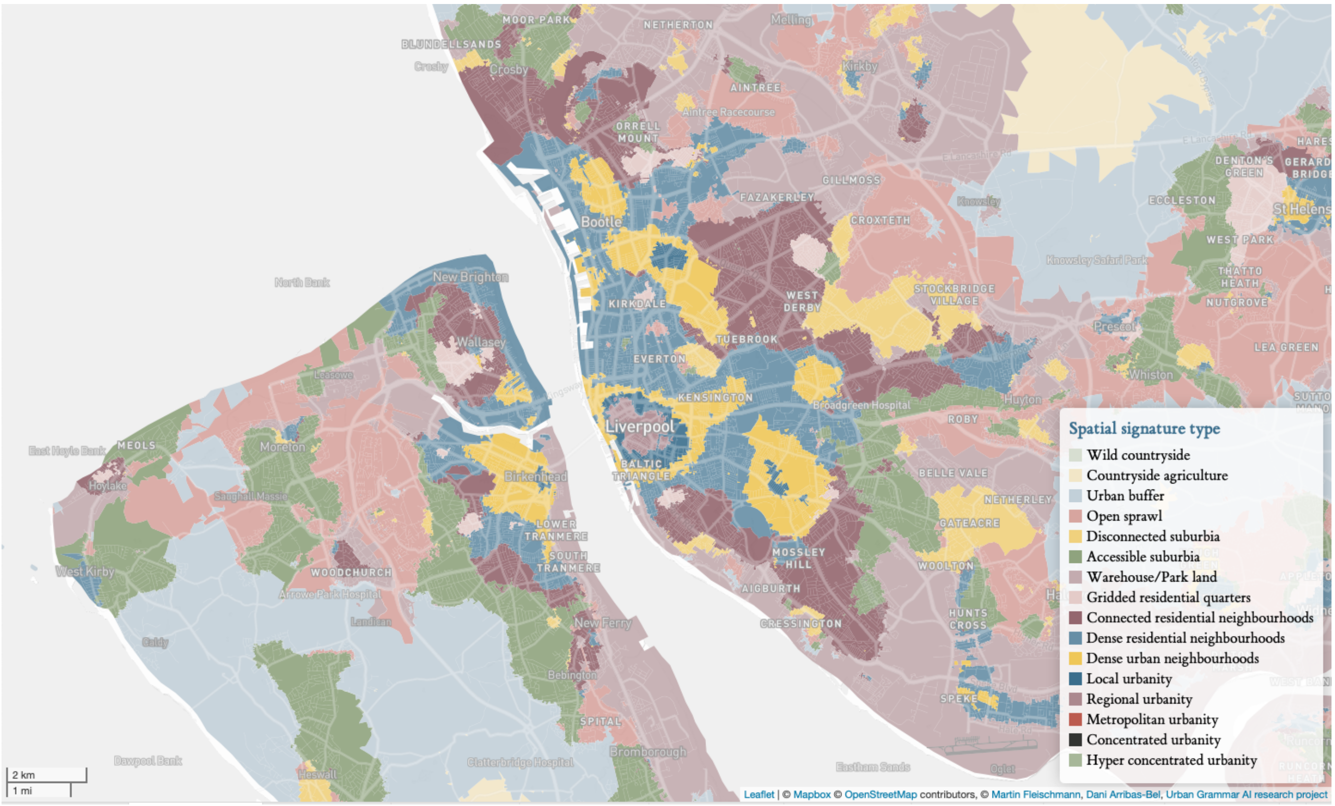


**Figure 2.** Illustration of a classification of spatial signatures in Liverpool and Birkenhead area, in the north west of England.

presents it in a way that is easy to access, work with and integrate into larger projects. Together with the importance of form and function discussed above, we anticipate the output will be relevant to both academic researchers as well as policymakers and practitioners. As a conceptual framework, the spatial signatures provide a flexible yet generalisable approach to understand, characterise and quantify urban form and function. One way to understand our results is as an implementation of a more general way of thinking about the spatial dimension of cities. In this context, it can be useful to researchers and practitioners who, even if not specifically interested in Great Britain, would like to implement a similar approach. In this respect, we hope the present paper serves not only to document our own work but to inspire future efforts aimed at urban form and function.

## Methods

The method of identification of spatial signatures consists of three top-level steps. First, we delineate a spatial unit of analysis that reflects the structure of urban phenomena on a very granular level. Then we characterise each of them according to form and function, capturing the nature of each unit and its spatial context. Finally, we use cluster analysis to derive a typology of our spatial units that, once combined into contiguous areas, forms a typology of spatial signatures.

### Spatial unit

The first major methodological decision relates to the definition of the spatial unit. An ideal candidate needs to reflect space in a granular manner, and we argue it should fulfil three conditions. First, it should be *indivisible*, meaning that any subdivision would result in a unit that is incapable of capturing the nature of urban form and function. Second, it needs to be *internally consistent* - it should always reflect only a single signature type. Last, it should be geographically *exhaustive*, covering the entirety of the study area.

Spatial units used in literature can be split into three groups. One is using administrative boundaries like city regions[4], wards or census output areas[5], that are convenient to obtain and can be easily linked to auxiliary data. However, those rarely reflect the morphological composition of urban space and, in some cases, may even “obscure morphologic reality”[6]. At the same time, most of them are divisible, and larger units are not always internally consistent. Another group is based on arbitrary uniform grids linked either to spatial indexing methods like H3[7] or Ordnance Survey National Grid, or to ancillary data of remote sensing or other origins like a WorldPop grid[8]. Grids however cannot be considered internally consistent as they do not

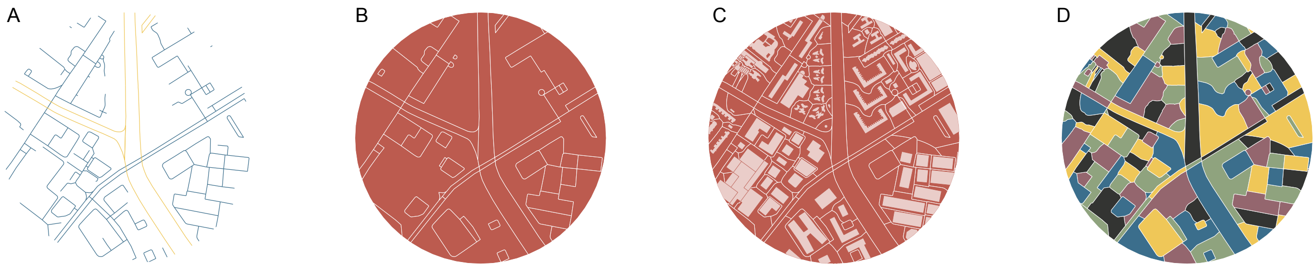


**Figure 3.** Diagram illustrating the sequential steps leading to the delineation of enclosed tessellation. From a series of enclosing components, where blue are streets and yellow river banks (A), to enclosures (B), incorporation of buildings as anchors (C) to final tessellation cells (D).

consider the underlying structure of the landscape. Finally, urban morphology studies tend to use morphological elements as street segments[9], blocks[10], buildings[11] or plots[12] as units of analysis. Some of those could be seen as indivisible and internally consistent, but since they are largely based on built-up fabric, they are not exhaustive. For example, in areas without any building or street, there is no spatial unit to work with. Plots could be theoretically considered as exhaustive, consistent and indivisible, but there is no accepted conceptual definition and unified geometric representation[13].

We are, therefore, proposing an application of an alternative spatial unit called *enclosed tessellation cell* (ETC), defined as "the portion of space that results from growing a morphological tessellation within an enclosure delineated by a series of natural or built barriers identified from the literature on urban form, function and perception"[1]. ETCs follow the morphological tradition in that it is based on the physical elements of an environment but overcome the drawbacks of conventionally used units. Its geometry is generated in the three steps illustrated on a Figure 3. First, a set of features representing physical barriers subdividing space, in our case composed of the street network, railways, rivers and a coastline, is combined, generating a layer of boundaries (3 A). These then partition space into smaller enclosed geometries called *enclosures* (3 B), which can be very granular or very coarse depending on the geographic context. In dense city centres where a single enclosure represents a single block is a high frequency of small enclosures. At the same time, in the countryside, this approach leads to very few large enclosures as their delimiters are far away from each other. Enclosures are then combined with building footprints (3 B), which act as anchors in space and potentially subdivide enclosures into enclosed tessellation cells using the morphological tessellation algorithm[14] (3 D), a polygon-based adaptation of Voronoi tessellation. The resulting geometries are indivisible as they contain, at most, a single anchor building, internally consistent due to their granularity and link to morphological elements composing urban fabric, and geographically exhaustive as they cover an entire area limited by specified boundaries.

In our ODP for Great Britain, street networks are extracted from OS Open Roads datasets[15] representing simplified road centrelines cleaned of underground road segments. Railways are retrieved from OS OpenMap - Local[16] ("RailwayTrack" layer) which captures surface railway tracks. Rivers are extracted from OS OpenRivers[17] representing river network of GB as centrelines, and a coastline is retrieved from OS Strategi®[18], capturing coastline as a continuous line geometry. Building geometry is extracted, again, from OS OpenMap - Local ("Building" layer) and represents generalised building footprint polygons.[1]

## Characterisation of space

Spatial signatures capture the character of the built and unbuilt environment based on two components - form and function. Each of them is quantified at the level of individual ETCs using methods appropriate for each specific dataset. While form is described using urban morphometrics (i.e. quantitative analysis of urban form)[19], function is a composite of a variety of data inputs. We outline each component with a bit more detail below.

### *Form*

Morphometric characterisation of urban form is based on the numerical description of four elements capturing the built environment - buildings, streets, ETCs, and enclosures - and reflects their patterns based on six categories of characters: dimensions, shapes, spatial distribution, intensity, connectivity and diversity[20]. Each element is considered across different scales, from the measurement of individual geometries, to relations of neighbouring geometries, to a graph-based analysis of the street network. The combination of elements, categories and scales results in a set of 59 individual morphometric characters listed in the online table 1.

[1]Note that the dataset does not distinguish between individual buildings when they are adjacent (e.g. perimeter block composed of multiple buildings is represented by a single polygon).

However, measuring individual characters is not enough to understand the predominant spatial patterns. For some types of urban form, high heterogeneity is not uncommon. This means that using, for example, areas of building footprints would, in most cases, result in largely discontinuous clusters that do not capture the aspect of an area. Therefore, we represent each of the morphometric characters using three summary variables reflecting statistical distributions of measured data within a spatial context of each ETC. Context is defined as tenth order of contiguity computed across the mesh composed of contiguous ETCs. Furthermore, each value is weighted by the inverse distance between so-called poles of inaccessibility (defined as a centre of a maximum inscribed circle) of each ETC. Three proxy variables then capture the first, the second and the third quartile of the resulting weighted distribution. Such a characterisation can capture the contextual tendency of each morphometric character and hence identify contiguous clusters in both homogenous and heterogeneous urban tissues.

### *Function*

Characterisation of the function component uses a different approach. While data describing urban form are not generally available in a processed format, forcing us to employ morphometric approaches, different aspects of function are often available as open data products. Therefore, the main goal of our characterisation of ETCs based on function is to develop appropriate transfer methods to link data published as grids or linked to administrative boundaries to ETCs.

In this work, we are using five different transfer methods: Areal interpolation, Building-based dasymetric areal interpolation[21], Network-constrained accessibility, Euclidean accessibility, and Zonal statistics. Areal interpolation is used when the functional data covers the entirety of space in the form of polygon geometry and when there is no assumption that the phenomena it captures are linked directly to the human population, such as land cover data. When there is an assumption of relation to the population, building-based dasymetric areal interpolation is used instead. The main difference is that instead of ETC polygons, building footprint polygons linked to individual ETCs are used as a target of interpolation. That ensures that data like population estimates are linked to ETCs proportionally to their ability to house population rather than by their area. Network-constrained accessibility is used when the input data represent points of interest like locations of supermarkets. Points are then snapped to the nearest node on the street network and linked to the ETCs through the count of observations accessible from the cell within 15 minutes of walk (1200m on the street network) and a distance to the nearest point. In some cases, Euclidean (as-crow-flies) accessibility is measured instead to accommodate for phenomena that are often outside the reach of a drivable network like water bodies. Zonal statistics are used to transfer data originally stored in a raster format to ETCs as the mean value of raster pixels intersecting each polygon geometry. Finally, characters based on interpolation and zonal statistics are expressed using their contextual versions following the method used for form characters to, again, reflect the contextual pattern of measured values. The selection of datasets and the chosen transfer method are listed in the table 1.

## Cluster analysis

When combined, contextual summaries of form and function characters (or characters themselves when they are reflecting the context by definition) compose a dataset describing each ETC by 331 variables (177 for form and 154 for function.) Assigning equal weight to each variable, we standardize them applying Z-score normalization, and use them as input for K-Means cluster analysis.

Due to the nature of the selected K-Means clustering, the step preceding the final analysis is the selection of an optimal number of clusters. We use the clustergram exploratory method[22], reflecting the behaviour of different options, the relationship between clustering solutions regarding the allocation of individual observations to classes, and the separation between the clusters within each tested solution. Clustergram is further accompanied by measures of internal validation measures - the Silhouette score diagram, Calinski-Harabasz index[23] and Davies-Bouldin index[24]. The optimal number of classes is selected based on the interpretation of clustergram supported by additional measures aiming at a balance between cluster separation and an appropriate detail of resulting classification.

The results of the clustering capture the first group of a national signature classification composed of ten clusters. However, since the classified ETCs cover entirety of space from vast natural open spaces to dense city centres, it may result in only a few classes representing urban areas. While that is caused by the variable heterogeneity of our dataset in combination with K-Means clustering, the measured characters have the ability to further distinguish classes of already identified clusters. As spatial signatures are focused on the urban environment, we further subdivide those clusters covering a substantial portion of urban areas using another iteration of K-Means clustering (one class into nine and another into three clusters). The resulting classification then provide classification capturing the typology of spatial signatures with a detailed focus on urban development.

Finally, individual spatial signature geometries are generated as a combination of adjacent ETCs belonging to the same signature class.

| data | source | input geometry | transfer method |
|---|---|---|---|
| Population estimates | ONS Census Output Area population estimates, Statistics.gov.scot | Vector (output area polygon) | Building-based dasymetric areal interpolation |
| Retail POIs (supermarkets) | Geolytix | Vector (point) | Network-constrained accessibility |
| Water bodies | OS OpenMap Local | Vector (water body polygon) | Euclidean accessibility |
| Listed Buildings | Historic England, Historic Environment Scotland, Lle Geo-Portal for Wales | Vector (point) | Network-constrained accessibility |
| Night Lights | VIIRS DNB Nighttime Lights | Raster (500m) | Zonal statistics |
| Food Hygiene Rating Scheme Ratings | CDRC.ac.uk | Vector (point) | Network-constrained accessibility |
| Workplace population | ONS Census Workplace population, Scotland's census Workplace population | Vector (output area polygon) | Building-based dasymetric areal interpolation |
| Culture (theatres, cinemas) | OpenStreetMap | Vector (point) | Network-constrained accessibility |
| Corine land cover | Copernicus Land Monitoring Service | Vector (land cover zone polygon) | Areal interpolation |
| NDVI | GHS-composite-S2 R2020A | Raster (10m) | Zonal statistics |
| Retail centres | CDRC.ac.uk | Vector (retail centre polygon) | Euclidean accessibility |

**Table 1.** Functional characters used to describe the function component of spatial signatures. For details of the implementation, refer to the reproducible Jupyter notebooks available at urbangrammarai.github.io.

# Data Records

The data product described in this article is available through the Consumer Data Research Centre Open Data repository available at https://data.cdrc.ac.uk/dataset/xxx-xxx-xxx under the Open Government Licence v3.0 license. The dataset stored in the repository contains a GeoPackage with a signature geometry (OSGB36 / British National Grid (`EPSG:27700`) CRS) and related signature type, plain-text pen portraits describing individual signature types, a series of CSV files describing individual signatures and signature types, and a CSV files linking signature types to the Output Area and Lower Super Output Area geometry. An online interactive map of spatial signatures for the whole of Great Britain is available on the project website (https://urbangrammarai.github.io/great-britain).

# Technical Validation

Spatial signatures are unique as a classification method, limiting the potential validation approaches. In this context, we focus on indirect methods that use ancillary datasets capturing conceptually similar aspects of the environment. We compare the signatures with three of such datasets, each focusing on a different classification perspective, but all related to our classification to a degree when we can assume there will be a measurable level of association between the two:

- WorldPop settlement patterns of building footprints (2021)[8]
- Classification of Multidimensional Open Data of Urban Morphology (MODUM) (2015)[5]
- Copernicus Urban Atlas (2018)[25]

## Validation approach

All datasets, spatial signatures and those selected as validation contain a categorical classification of space linked to their unique geometry. The first requirement to be able to compare data products is to transfer their information to the same geometry. We take two approaches for this step, depending on the dataset we are comparing the signatures with: an interpolation of one set of polygon-based data to another (input to ETCs); or the conversion of spatial signatures to the raster representation matching an input raster, which is computationally more efficient when one of the layers is already a raster. The second step is a statistical

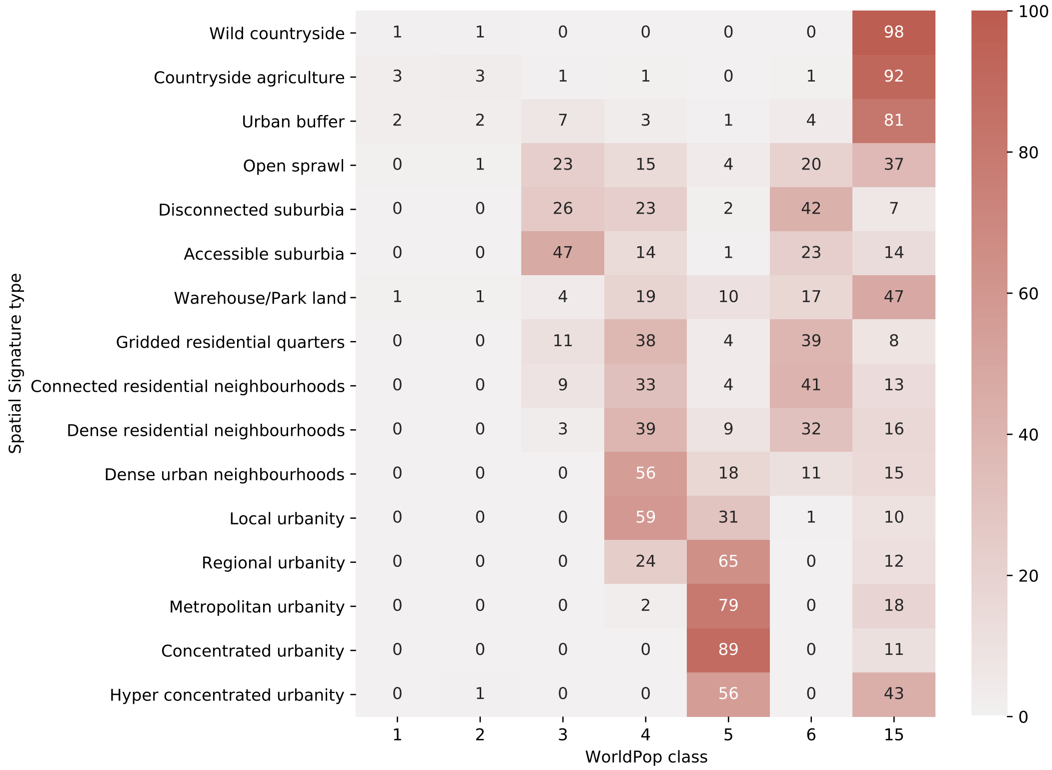


**Figure 4.** Contingency table showing frequencies (in %) of WorldPop classes within signature types.

comparison of two sets of classification labels, one representing spatial signature typology and the other validation classes. We use contingency tables and Pearson's $\chi^2$ test to determine whether the frequencies of observed (signature types) and expected (validation types) labels significantly differ in one or more categories. Furthermore, we use Cramér's $V$ statistics[26] to assess the strength of the association.

## WorldPop settlement patterns of building footprints

WorldPop settlement patterns of building footprints aim to derive a typology of morphological patterns based on a gridded approach with cells of 100x100m, and building footprints. Authors measure six morphometric characters linked to the grid cells and use them as input for an unsupervised clustering algorithm leading to a six-class typology. As the classification is dependent on building footprints, grid cells that do not contain any information on the building-based pattern are treated as missing in the final data product. For the validation of spatial signatures, this *missing* category is treated as a single class. It is assumed that the top-level large scale patterns detected by the WorldPop method and spatial signatures will provide similar results. However, there will be differences caused by the inclusion of function in spatial signatures, higher granularity of both initial spatial units and the resulting classification (6 vs 19 classes).

Signature typology is rasterized and linked to the WorldPop grid. The resulting contingency table is shown in Figure 4. There is a significant relationship between two typologies, $\chi^2(114, N = 22993921) = 13341832, p < .001$. The strength of association measured as Cramér's $V$ is 0.311, indicating moderate association. The contingency table shows that WorldPop classes tend to be linked to groups of signature types of a similarly degree of urbanity. A WorldPop class 15 is "undefined" due to the lack of building footprints in the area, therefore overlapping a large portion of signatures.

## MODUM

Multidimensional Open Data Urban Morphology (MODUM) classification describes a typology of neighbourhoods derived from 18 indicators capturing built environment as streets, railways or parks, linked to the Census Output Area geometry. The classification identifies 8 types of neighbourhoods. Compared to the WorldPop classification, MODUM takes into account more features of the built environment than building footprints, which makes it conceptually closer to the spatial signatures. However, it is still focusing predominantly on the form component, although there are some indicators that would be classified as function within the signatures framework (e.g. population). The MODUM method uses a different way of capturing context compared to the signatures, which leads to some classes being determined predominantly by a single character. For example,

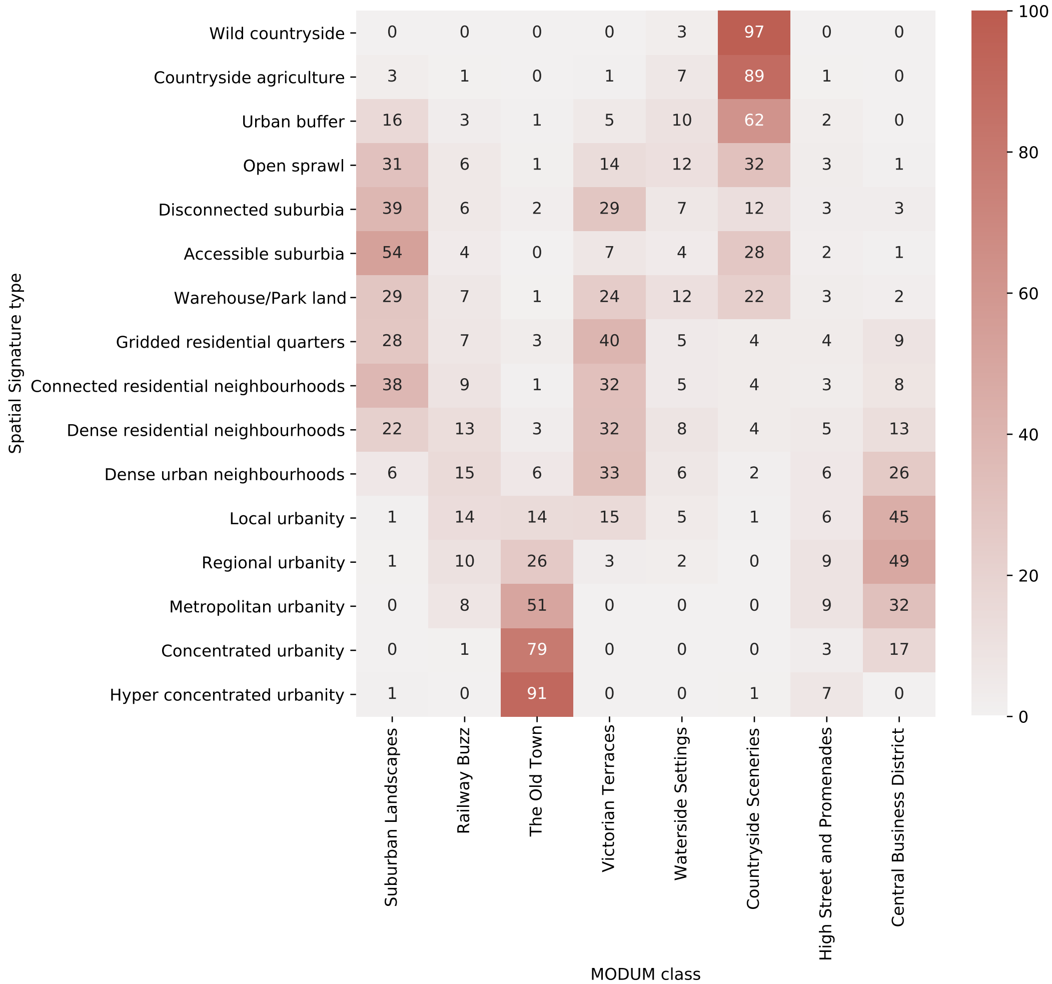


**Figure 5.** Contingency table showing frequencies (in %) of MODUM classes within signature types.

the *Railway Buzz* type forms a narrow strip around the railway network, which is an effect signatures avoid. MODUM typology is available only for England and Wales. Therefore the validation takes into account only ETCs covering the same area. The classification is linked to the ETC geometry is based on the proportion (the type covering the largest portion of ETC is assigned). The resulting contingency table is shown in Figure 5. There is a significant relationship between two typologies, $\chi^2(152, N = 13067584) = 13938867, p < .001$. The strength of association measured as Cramér's $V$ is 0.300, indicating moderate association of very similar levels we have seen above. The contingency table indicates similar relationships, where a single MODUM class overlaps a group of signature types. However, the groups tend to be well defined and formed based on the similarity of types.

## Copernicus Urban Atlas

Copernicus Urban Atlas is the least similar of the validation datasets. It is a high-resolution land use classification of functional urban areas derived primarily from Earth Observation data enriched by other reference data as OpenStreetMap or topographic maps. Its smallest spatial unit in urban areas is 0.25 ha and 1 ha in rural areas, defined primarily by physical barriers. The Urban Atlas classification, which identifies 27 classes predefined classes using the supervised method. The majority of urban areas is classified as urban fabric further distinguished based on continuity and density resulting in six classes of the urban fabric. The classification does not consider the type of the pattern or any other aspect. Furthermore, it does not take into account what signatures call *context* as each spatial unit is classified independently, which in some cases leads to the high heterogeneity of classification within a small portion of land. Signatures take a different approach. Consequently, it is expected that the similarity between the two will be limited. Urban Atlas is available only for functional urban areas (FUA), leaving rural areas unclassified. Validation then applies to FUAs only. The classification is linked to the ETC geometry based on the proportion (the type covering the largest portion of ETC is assigned). The resulting contingency table is shown in Figure 6. There is a significant relationship between two typologies, $\chi^2(450, N = 8396642) = 5229900, p < .001$. The strength of association measured as Cramér's $V$ is 0.186, indicating a weak association. The contingency table shows the difference in the aim of spatial signatures and that of Urban Atlas with a majority of signatures being linked to a few of Urban Atlas classes. Within relevant classes, we see a tendency of signature types to cluster within Urban Atlas classes based on the level of urbanity, albeit not as strong as in the previous two cases.

### Summary

None of the comparisons shows more than a moderate association, but since none of the validation datasets is aiming to capture the same conceptualization of space as spatial signatures do, such a result is expected. The moderate association with both WorldPop settlements patterns and MODUM is reassuring as both are conceptually closer to signatures than the Urban Atlas (especially in their unsupervised design). Urban Atlas, though very different in its aims and methods, still shows a measurable association, which we interpret as sign that the key structural aspects forming cities are captured by both. The validation exercise suggests that general patterns forming cities are shared among signatures and existing typologies. Signature types tend to form groups when we look at their relation to validation classes and it is not uncommon that a single signature type is present in multiple groups linked to different classes. However, all these groups tend to be formed based on the similarity and illustrate the granularity of the presented classification compared to existing datasets, allowing us to distinguish, for example, five types of signature types forming town an city centres.

## Usage Notes

The released data product follows widespread standards for geographic data storage and should be easy to integrate with other data and methods by researchers wanting to reuse it. However, due to the density of signature geometry (resulting from the detailed ETCs), it may be needed to simplify the geometry for a smoother interactive experience on machines with limited resources.

Replication of the analysis optimally requires at least a single computational node with a large amount of RAM (+100GB) due to the size of the input data and detail on which signature characterization is computed. It is also recommended to revisit the state of the development of related software packages, notably `momepy`[27], `libpysal`[28], `tobler`[21] and `dask-geopandas` as they may soon offer more efficient drop-in replacements of the custom code used to produce this dataset.

## Code availability

The source code used to produce this dataset is openly available in a GitHub repository at https://github.com/urbangrammarai/spatial_signatures and in the form of a website on https://urbangrammarai.github.io. Code is organized in a series of Jupyter notebooks and have been executed within the `darribas:gds_dev:6.1`[29] Docker container, unless specified otherwise in the individual notebooks.

## References


1. Arribas-Bel, D. & Fleischmann, M. Spatial Signatures - Understanding (urban) spaces through form and function (2021). Mimeo.
2. Fleischmann, M. & Arribas-Bel, D. Classifying urban form at national scale - The British morphosignatures (2021). ISUF Proceedings.
3. Arribas-Bel, D., Green, M., Rowe, F. & Singleton, A. Open Data Products: a framework for creating valuable analysis-ready data. *J. Geogr. Syst.* (*forthcoming*).
4. Angel, S., Arango Franco, S., Liu, Y. & Blei, A. M. The shape compactness of urban footprints. *Prog. Plan.* **139**, 100429, 10/gg638j (2020).
5. Alexiou, A., Singleton, A. & Longley, P. A. A Classification of Multidimensional Open Data for Urban Morphology. *Built Environ.* **42**, 382–395, 10/gddwsn (2016).
6. Taubenböck, H. *et al.* A new ranking of the world's largest cities—do administrative units obscure morphological realities? *Remote. Sens. Environ.* **232**, 111353 (2019).
7. Brodsky, I. H3: Uber's hexagonal hierarchical spatial index. *Available from Uber Eng. website: https://eng. uber. com/h3/[22 June 2019]* (2018).
8. Jochem, W. C. & Tatem, A. J. Tools for mapping multi-scale settlement patterns of building footprints: An introduction to the R package foot. *PloS one* **16**, e0247535, 10/gh7sjr (2021).

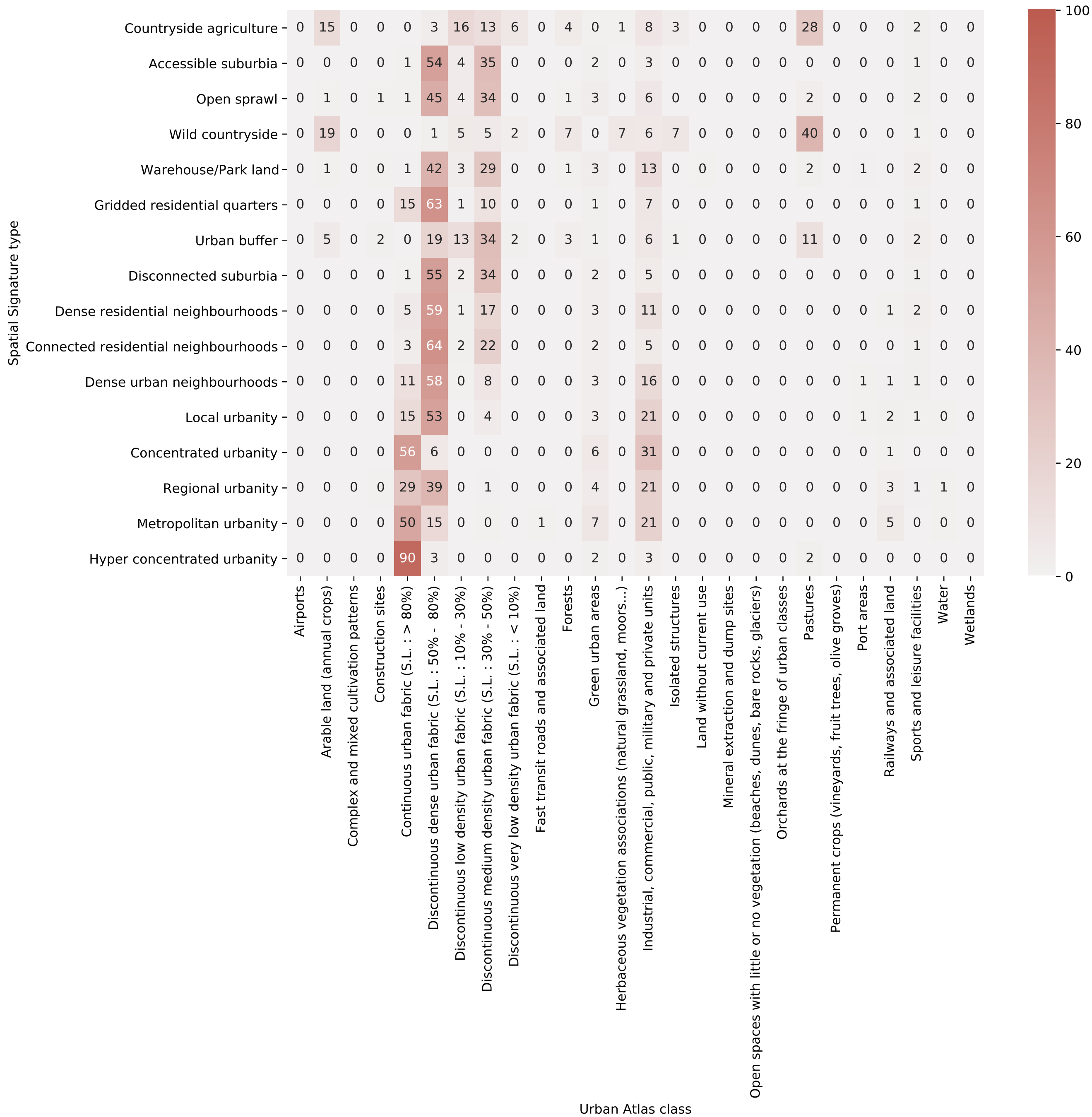


**Figure 6.** Contingency table showing frequencies (in %) of Urban Atlas classes within signature types.

9. Araldi, A. & Fusco, G. From the street to the metropolitan region: Pedestrian perspective in urban fabric analysis:. *Environ. Plan. B: Urban Anal. City Sci.* **46**, 1243–1263, 10.1177/2399808319832612 (2019).
10. Gil, J., Montenegro, N., Beirão, J. N. & Duarte, J. P. On the Discovery of Urban Typologies: Data Mining the Multi-dimensional Character of Neighbourhoods. *Urban Morphol.* **16**, 27–40 (2012).
11. Hamaina, R., Leduc, T. & Moreau, G. Towards Urban Fabrics Characterization Based on Buildings Footprints. In *Bridging the Geographic Information Sciences*, vol. 2, 327–346, 10.1007/978-3-642-29063-3_18 (Springer, Berlin, Heidelberg, Berlin, Heidelberg, 2012).
12. Bobkova, E., Berghauser Pont, M. & Marcus, L. Towards analytical typologies of plot systems: Quantitative profile of five European cities. *Environ. Plan. B: Urban Anal. City Sci.* 239980831988090, 10/ggbgsm (2019).
13. Kropf, K. Plots, property and behaviour. *Urban Morphol.* **22**, 5–14 (2018).
14. Fleischmann, M., Feliciotti, A., Romice, O. & Porta, S. Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. *Comput. Environ. Urban Syst.* **80**, 101441, 10.1016/j.compenvurbsys.2019.101441 (2020).
15. Ordnance Survey. OS Open Roads (2020).
16. Ordnance Survey. OS OpenMap - Local (2020).
17. Ordnance Survey. OS Open Rivers (2020).
18. Ordnance Survey. Strategi (2016).
19. Dibble, J. *et al.* On the origin of spaces: Morphometric foundations of urban form evolution. *Environ. Plan. B: Urban Anal. City Sci.* **46**, 707–730 (2019).
20. Fleischmann, M., Romice, O. & Porta, S. Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environ. Plan. B: Urban Anal. City Sci.* 239980832091044, 10/ggngw6 (2020).
21. eli knaap *et al.* pysal/tobler: Release v0.8.2, 10.5281/zenodo.5047613 (2021).
22. Schonlau, M. The clustergram: A graph for visualizing hierarchical and nonhierarchical cluster analyses. *The Stata J.* **2**, 391–402, 10/ghh97z (2002).
23. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **3**, 1–27, 10.1080/03610927408827101 (1974). https://www.tandfonline.com/doi/pdf/10.1080/03610927408827101.
24. Davies, D. L. & Bouldin, D. W. A cluster separation measure. *IEEE transactions on pattern analysis machine intelligence* 224–227 (1979).
25. European environment agency (EEA). Urban Atlas (2018).
26. Cramér, H. *Mathematical Methods of Statistics (PMS-9), Volume 9* (Princeton university press, 2016).
27. Fleischmann, M. momepy: Urban morphology measuring toolkit. *J. Open Source Softw.* **4**, 1807, 10.21105/joss.01807 (2019).
28. Rey, S. J. *et al.* The pysal ecosystem: Philosophy and implementation. *Geogr. Analysis* (2021).
29. Arribas-Bel, D. `gds_env`: A containerised platform for geographic data science, 10.5281/zenodo.4642516.



## Acknowledgements

(not compulsory)

M.F. and D.A. kindly acknowledge funding by the UK's Economic and Social Research Council through the project "Learning an urban grammar from satellite data through AI", project reference `ES/T005238/1`.


## Author contributions statement

M.F. and D.A. designed the method, M.F. conducted the experiments, M.F. and D.A. analysed the results. M.F. and D.A. wrote and reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Figures & Tables

xxxTODOxxx move figures and tables here