

Global Building Morphology Indicators

Filip Biljecki^{a,b,*}, Yoong Shin Chow^a

^a Department of Architecture, National University of Singapore, Singapore

^b Department of Real Estate, National University of Singapore, Singapore

ARTICLE INFO

Keywords:

Urban planning
GIScience
Morphometrics
OpenStreetMap
GeoAI
Spatial analysis

ABSTRACT

Characterising and analysing urban morphology is a continuous task in urban data science, environmental analyses, and many other domains. As the availability and quality of data on them have been increasing, buildings have gained more attention. However, tools and data facilitating large-scale studies, together with an interdisciplinary consensus on metrics, remain scarce and often inadequate. We present Global Building Morphology Indicators (GBMI) — a three-pronged contribution addressing such shortcomings: (i) a comprehensive list of hundreds of building form multi-scale measures derived through a systematic literature review; (ii) a methodology and tool for the computation of these metrics in a database suited for big data and comparative studies, and release the code freely and open-source; and (iii) we carry out the computations using high performance computing, generating a public repository with data quantifying the form of selected urban areas around the world, and demonstrate their value with novel analyses comparing morphological parameters across cities. GBMI introduces a formalised, structured, modular, and extensible method to compute, manage, and disseminate urban indicators at a large scale and high resolution, while the precomputed dataset facilitates comparative studies. The theory and implementation traverse multiple scales: at the building level, both individual and contextual ones based on encircling buildings by multiple buffers, and aggregations at several hierarchical administrative levels and at multiple grids. Our open dataset, comprising billions of records on a growing scope of urban areas worldwide, is the most comprehensive instance of morphological data parametrising the individual building stock, supporting studies in urban analytics and a range of disciplines.

1. Introduction

Urban form has been studied for decades and a large number of publications has focused on characterising it with numerical metrics, revealing patterns of the built-up fabric and underpinning investigations on their interplay with a multitude of phenomena (Hermosilla, Palomar-Vázquez, Balaguer-Beser, Balsa-Barreiro, & Ruiz, 2014; Sharifi, 2019; Silva, Oliveira, & Silva, 2014; Wang, Madden, & Liu, 2017). As buildings dominate the urban landscape, they are often an indispensable component of such studies, particularly in recent years with the growing volume of building data and they are being increasingly interwoven with multifaceted aspects such as energy, wind, solar radiation, thermal comfort, noise, transportation, carbon emissions, health, urban vibrancy, bikeability, population dynamics, and urban planning (Botta & Gutiérrez-Roig, 2021; Byrne, Hewitt, Griffiths, & MacArtain, 2021; Chen et al., 2022; Chen, Han, & Vries, 2019; Heris, Middel, & Muller, 2020; Ito & Biljecki, 2021; Milojevic-Dupont & Creutzig, 2020; Paiva, Santos, & Rossetti, 2018; Ryu, Park, Chun, & Chang, 2017; Stojanovski,

2019; Szarka & Biljecki, 2022; Tian & Yao, 2021; Tong & Kang, 2021; Usui, 2020; Wang, Cot, Adolphe, Geoffroy, & Sun, 2017; Wang & Deb- bage, 2021; Wong, Tan, Kolokotsa, & Takebayashi, 2021; Wu, Lu, Gao, & Wang, 2022; Yu, Chen, & Wong, 2020; Yuan et al., 2020; Zhang et al., 2019; Zhang, Wu, Zhu, & Liu, 2019).

We identify three major research gaps in this domain, which we tackle in this cohesive and integrated paper. First, even though great strides have been made in the field and there is a mountain of research papers focused on parametrising buildings under the umbrella of urban form, to the extent of our knowledge, there is no review or inventory, nor a consensus on them. In this paper, we conduct a systematic review to identify commonly used indicators and shape a framework, and present the results together with observations from the review. With scalability and multidisciplinary in mind, we focus on indicators that do not require data that are not widely available (i.e. we use building footprints) and those that do not require complex simulation software that is exclusive to a particular domain and not scalable. While we encounter a heterogeneous landscape of metrics and terminological

* Corresponding author at: Department of Architecture, National University of Singapore, Singapore.

E-mail address: filip@nus.edu.sg (F. Biljecki).

disparities, we identify common patterns and derive a list of indicators that may serve a kaleidoscopic set of topics, and introduce new derivative parameters, furthering possibilities in this domain.

Second, tools to compute such building form metrics are scarce, and they may not be suited to handle large-scale needs. As the scale of morphological analyses, in particular on street networks, is growing and sometimes including nation-wide studies and beyond (Boeing, 2020; Liu, Chen, Li, & Chen, 2020; Zhou, Lin, & Bao, 2021), we believe that work supporting wide ranging analyses using building data is necessary and timely. Considering the growth of building data, e.g. open government data on buildings and large-scale data released by companies (Heris, Foks, Bagstad, Troy, & Ancona, 2020; Sirko et al., 2021) and academia (Zhang et al., 2022), studies using buildings will only grow in importance in the future. Our paper presents the development of a tool to implement the computation of hundreds of indicators structured following the review and the creation a large-scale database, which is suited for big data management and analyses, as we demonstrate with several examples. Implementing the work as a database rather than as desktop tool as most of related work (Section 2), has many advantages that will be elaborated later. The developed code is free and open-source, it is extensible (e.g. allowing the addition of new indicators if necessary), customisable (e.g. supporting different forms of input and output data), and flexible (e.g. it can be used to compute the indicators anywhere at nearly any scale and resolution).

Third, while the computation of such indicators may be laborious, consuming time of researchers, especially in studies that are increasingly including multiple locations (Kraff, Wurm, & Taubenböck, 2020), there are no publicly available datasets with precomputed indicators that can be readily used. Having a ready-to-use dataset may save substantial effort to researchers, proliferate urban morphology among less computationally inclined researchers, and enable quick comparative studies involving multiple cities, which is uncommon (based on our review presented in Section 3). We engage the developed method and high performance computing to calculate indicators of buildings in many cities around the world where such data are available (mostly thanks to OpenStreetMap (OSM)), and aggregate them according to zones defined by both a global administrative and a global gridded population dataset at fine scale. The precomputed datasets are accommodated in a public repository, a continuous and growing effort. This part also affirms the large-scale feasibility of our work and confirms the advantages of the database approach, as the generated dataset is of unprecedented size, covering many urban areas in the world that are well mapped in OSM and containing billions of records. At the same time, the database maintains simplicity and efficiency, and provides ease of extracting data. Further, while we focus on OSM, the developed software architecture is versatile and the approach is dataset-agnostic, as different and multiple datasets can be used to derive the metrics, allowing customisation and suiting one's needs. We release this dataset as open data under a liberal license so that anyone can use it without restrictions. To the extent of our knowledge, this dataset presents the most comprehensive one in this domain and, together with the established methodology, we hope that it will contribute to the field in multiple ways. Our work by design enables computation at any scale and at any location, and given the worldwide focus as demonstrated with the datasets and novel analyses that we showcase in this paper, we title our contribution as Global Building Morphology Indicators (GBMI).

The paper is organised as follows. Section 2 expands the introduction by affirming the relevance of building indicators in urban morphology studies and discusses related work. Section 3 presents the compendium of indicators characterising building morphology, which are derived from a systematic literature review. In Section 4, we present the implementation portion of GBMI: an open-source software that generates a structured and easily accessible database, and a large-scale dataset that we release as open data. An insight in the generated data is given with several analyses in Section 5 suggesting urban textures and fingerprints of cities, and asserting the importance of the work for both

studying patterns within single cities and comparative analyses, another scientific contribution. Section 6 discusses the work revealing challenges, limitations, and opportunities, while Section 7 concludes the paper.

2. Background and related work

2.1. Examples of urban form studies relying on data of buildings

There are many examples of using building data to parametrise the urban form. Some are given in this section to provide an understanding of the applications across various disciplines.

Urban morphology has been an integral component of the work by Chen et al. (2020) on microclimate simulations. They zero in on several indicators, such as building coverage and total exterior wall area, mainly derived from a climate study showing their association to air temperature (Jin, Cui, Wong, & Ignatius, 2018). Further examples of microclimate studies relying on quantitative urban form metrics are many (Ng, Yuan, Chen, Ren, & Fung, 2011; Tong et al., 2018).

The paper of Zhu et al. (2020) is another environmental work, describing the effect of urban morphology on the solar capacity of ten cities and discussing what urban form is most desirable to effectively harness solar energy. Similarly to others, they take into account the average height of buildings in an area, but also the standard deviation of the values to indicate their variation in a study area. In the same domain, Morganti, Salvati, Coch, and Cecere (2017) determine seven indicators and investigate their association with the solar availability on façades of buildings. The analysis has been carried out at the neighbourhood level in two cities, and it reveals that the two most useful indicators for this purpose are the ratio of the built-up area to the urban site area (named as gross space index) and the ratio of the area of the walls of buildings to the urban site area (named as façade-to-site ratio). There are scores of studies in this domain, using a different set of indicators and datasets (de Lemos Martins, Adolphe, Bastos, & de Lemos Martins, 2016), suggesting a lack of consensus or the variability of the importance of different metrics across different geographies.

2.2. Related work

There has been some work on developing inventories of indicators, implementing their computation, and generating open datasets. Nonetheless, most of such work is rather in the street network department. For example, Boeing (2021) presents an open dataset of street network indicators with worldwide coverage, generated using a free and open-source software (Boeing, 2017).

Lemoine-Rodríguez, Inostroza, and Zepp (2020) assess the spatial structure of 194 cities over 25 years, based on a set of landscape metrics, identifying clusters of similar cities and various patterns. Their dataset is released openly (Lemoine-Rodríguez, Inostroza, & Zepp, 2020). The method relies on a coarse land use dataset and it is rather focused on the city level. A related effort is the dataset by Demuzere, Bechtel, Middel, and Mills (2019). Their work is focused on classifying local climate zones (LCZ) based on indicators at the block-level. Our approach regards individual buildings and it provides indicators at a fine spatial resolution and at multiple scales, and it could be used as an input to compute LCZs.

When it comes to data, the work that is perhaps the most related to ours is the one of Heris, Foks, et al. (2020). They use a nation-wide building footprints dataset of the United States and rasterise it at high resolution (30 m), computing six metrics for each cell. This dataset has been released as open data. However, the number of indicators is limited, the resolution is fixed, the work does not regard administrative units, and the dataset is focused on a single country. A related effort is presented by Li, Liu, Zhang, Xue, and Li (2021), focusing on dozens of cities in China, with similar shortcomings and the dataset does not appear to be released openly.

There has been previous exploratory work that is focused on

indicators that quantitatively measure the urban form. [Fleischmann, Romice, and Porta \(2020\)](#) examine the state of the art of quantitative analysis of the urban form. Among other aspects of the review such as delineating applications and purposes of the quantitative studies, the paper formalises a framework of metrics to characterise the urban fabric. However, the study is rather general, as it is not focused on buildings only. For a related work, see also ([Dibble et al., 2017](#)). In our review, we focus primarily on understanding the inventory of indicators and the data that was used to compute them. We have encountered and devised some unlisted indicators that we implement in our work.

[Basaraner and Cetinkaya \(2017\)](#) overview 20 indicators to characterise the shape complexity of building footprints — they investigate their usefulness and the computational complexity they entail. The study indicates that only a handful are sufficiently distinct (e.g. several pairs exhibit strong correlations), suggesting that it is of no use to have many different indicators characterising the same aspect. We learn from their work, but at the same time, we design our approach as modular, so additional indicators can be added easily if necessary.

Tool-wise, to the extent of our knowledge, arguably the work that is most akin to ours is the recent one by [Jochem and Tatem \(2021\)](#). They have developed an R package for calculating 2D morphology metrics from building footprints. Such development affirms the importance of deriving indicators based on building datasets and having easily available solutions to do so. Our work presents the following distinctions and contributions. First, we encompass a substantially larger suite of indicators derived from a systematic literature review. Second, our work includes the additional dimension of vertical indicators where building height data are available. While we are aware that 3D datasets are not available as widely as those in 2D, and thus, there are many areas for which such indicators cannot be computed for the time being, the share of metrics calculated from the vertical extent of the built environment in literature cannot be discounted. Third, a pillar of our work is an open dataset, bypassing employing software and running the computations. Fourth, our architecture is a structured database, which may be more suitable for handling big data and large-scale analyses, as demonstrated by later sections. As confirmed by a recent review of [Fleischmann, Feliciotti, and Kerr \(2021\)](#), a database tool has not yet been developed in this field, a gap we seek to bridge in this paper.

3. Catalogue of indicators

3.1. Review methodology

The first part of our triad (inventory, software, and data) is intended to shape the structure and list of indicators, starting with identifying those that are used commonly in the field. For that, we follow the typical approach of systematic literature reviews: we select relevant keywords and search for a set of papers, after which we filter papers relevant for our study, and then extract relevant information from them.

We have searched Scopus for papers that contain the following terms in the title, abstract or keywords: ‘urban’, ‘building’, and ‘morphology’. The query, which was performed in early January 2021, returned 1566 results. Considering the large number of papers and that a very comprehensive literature review is not the main purpose of this paper, it is not feasible to comb through all of them. As a way to reduce the pool of papers, we decided to focus on a subset containing the latest published papers. An advantage of this approach is that the survey will reflect the state of the art and the current status of the related studies. Therefore, we have selected the latest 100 published papers. In the screening phase, we have examined the title and abstract of each paper in the initial pool. We have proceeded to include a paper in our review only if it is a quantitative paper focusing on indexes derived from building data. For example, we have excluded qualitative and descriptive studies and those in which buildings merely play a minor role. After applying such inclusion criteria, filtering left us with 43 papers for a closer review. For each paper identified as relevant, we have extracted

the list of building indicators, and secondary information such as resolution at which the indicators have been computed, type of building data, and data source.

3.2. Summary, observations, and considerations

The literature review has been important in guiding the development of the inventory of indicators that we present in the next section, it highlighted the breadth of building form indicators, and it affirmed that urban form studies are largely disconnected, using a non-standardised set of metrics, even within the same research lines. There are several further relevant observations we unpack in this section.

A seemingly straightforward and unambiguous task such as recording the list of building-related indicators found in publications, together with auxiliary information such as understanding the data source(s), was hampered by a few issues.

First, we have observed that a number of papers is not forthcoming in listing and explaining the metrics used in the analysis. In certain cases, for some listed indicators, it is not clear what they represent and how are they computed. Furthermore, it is not always evident what data source and type was used to derive the metrics. Such issues hinder the assessment how feasible is the implementation of some indicators.

Second, as indicated by [Fleischmann, Romice, and Porta \(2020\)](#), morphological studies exhibit terminological inconsistencies, which we can attest to based on our exploration of papers. Besides multiple terms indicating the same concept appear regularly in publications (e.g. metrics, indicators, parameters, factors), we notice that disciplines collide in terminology, using different terms for the same indicator.

Third, the granularity and definition of indicators may be subject to different interpretations. For example, footprint area, height, and volume of a building are three archetypal measures across multiple domains. However, it is common practice to compute the last of the three by multiplying the first two ([Asadi, Arefi, & Fathipour, 2020](#)). A question is raised whether the volume should be considered as a distinct indicator, or simply a derivative indicator from the footprint area and height as more fundamental indicators, and thus, not counted as a unique indicator. We have decided to include such multiplicity if encountered more than once. Hence, in our work, volume is considered as an indicator, and the same goes for other derivative indicators such as the ratio of the building height to its footprint area ([Hu, Dai, & Guldmann, 2020](#)). Further examples of such derivative indicators include the area not covered by buildings (i.e. total area minus building coverage area) ([Li, Wu, Lin, Li, & Du, 2020](#)), ratio of the width and length of the footprint ([Tikhonova & Beirão, 2020](#)), façade to site ratio (i.e. sum of the product between the perimeter and the average building height in the area) ([Litardo et al., 2020](#)), surface to volume ratio ([Othman & Alshboul, 2020](#)), total area of urban envelopes divided by the corresponding flat area ([Zhu et al., 2020](#)), and absolute compactness (i.e. measuring the compactness of settlements by dividing the building volume by the zone area) ([Mohabat Doost, Buffa, Brunetta, Salata, & Mutani, 2020](#)). Accounting for all such combinations would be futile. Nonetheless, we have broken down such indicators to make sure we include the lowest common indicator used to compute them. Therefore, while we do not list them all explicitly, all these can be computed with our tool if necessary, given its modular nature and since they are usually computed from existing fundamental indicators that we include in our list.

Moving forward to the observations, in the majority of cases (three quarters), the studies focus on aggregating indicators of buildings at the level of a *zone*. The zones, which depending on the context, data, and size, are also called areas, plots and sites, may be represented as regular grids or irregular delineations such as administrative or census areas. There are different instances of irregular zones, such as administrative areas, land use parcels of variable size, and clusters of urban form ([Li, Zhou, Gong, Seto, & Clinton, 2020](#); [Song, Leng, Xu, Guo, & Zhao, 2020](#)). On the other hand, gridded zones are by definition regular and specified by a resolution, and are common. Their sizes vary, and some researchers

opt for multiple levels of aggregation in the same study, e.g. grids varying from 100 to 1000 m (Ribeiro, Martilli, Falls, Zonato, & Villalba, 2021). The indicators that are aggregated are based on the common descriptive statistic measures to indicate central tendency and dispersion of a distribution, such as mean and standard deviation, and are sometimes combined with another value (Li, Schubert, Kropp, & Rybski, 2020; Liao, Hong, & Heo, 2021). For example, such indicators include the coverage of building footprints (i.e. sum of footprint areas of all buildings in a zone divided by the area of the zone, obtaining the site coverage in percentage) and the average building footprint area.

Between the indicators derived independently for each building and the subsequent aggregated indicators derived at the level of irregular or regular zones, there are contextual indicators, i.e. zonal indicators at the building level, that are computed for each building from its surroundings, most often from circular buffers around each building (Milojevic-Dupont et al., 2020; Song, Lu, & Xing, 2020). They are computed for each building, but are dependent on the surrounding buildings, unlike independent indicators such as the complexity of a footprint of a building. An example of such indicator derived from the neighbouring context is the mean distance from the building in focus to all other buildings in a buffer. Further, these contextual building-level indicators may be aggregated at the level of a zone, e.g. mean of the mean distances between buildings, resulting in hierarchical summary statistics and indicating the interrelated and sequential nature of metrics. Consequently, in our catalogue, we outline three groups of indicators: building-level independent, building-level buffer, and zone-level indicators. In our implementation and dataset, we maintain multi-scalability in mind. Thus, the building-level buffer and zone-level indicators are computed at different scales and resolutions, suiting the diverse needs of various disciplines.

Data-wise, most studies rely on a single source of data, but overall, a variety of datasets is observed, e.g. vector building footprints occasionally containing also relevant attributes (most common), point clouds, semantic 3D city models, satellite imagery, street view imagery, and digital surface models (Aktas et al., 2020; Chen, Qiu, et al., 2020; Zhu et al., 2020). In spite of the different provenance of data, many of the indicators, e.g. footprint area, are common and can be computed from different datasets.

In the studies, not all indicators are computed from geometric data, and attributes are in some instances also considered as morphological parameters (Carlucci, Zambon, & Salvati, 2019). For example, it is common to come across studies estimating the building height from the number of storeys multiplied by an assumed fixed floor height (Ku & Tsai, 2020; Li, Koks, Taubenböck, & van Vliet, 2020; Liu, Xu, Zhang, & Shu, 2020; Peng et al., 2020). Such computations imply relaxed data quality requirements, as such approach may be prone to errors.

Most of the studies focus on a neighbourhood or a city, but there are exceptions, e.g. comparative studies including dozens of cities (Kraff et al., 2020; Liu, Wang, Qiang, Wu, & Wang, 2020).

The median number of building indicators used in the included papers is 4, with nearly all studies having a single-digit amount of indicators, barring a few instances, e.g. the one by Milojevic-Dupont et al. (2020) including approx. 80 metrics thanks to different scales (e.g. multiple buffers), derivative indicators, and summary statistics.

This exploration was instrumental in developing the catalogue of indicators presented next (Section 3.3). Still, we have not moved forward with all the indicators we have identified. First, some indicators are highly localised and categorical. For example, some papers categorise buildings into a few height classes (Cao et al., 2020; Ribeiro et al., 2021; Xia & Li, 2021), and count the share of these classes in an area (e.g. percentage of buildings that are taller than 20 m). These indicators are variations of existing indicators (e.g. building height) and they can be readily computed from the data that our tool generates and the dataset that we have released, if necessary. In fact, such an approach would also give researchers greater flexibility to define their own classes and thresholds, as they are not universal across studies.

Second, there are a few indicators that we have not included as we focus on those that can be computed from widely available datasets (i.e. building footprints) and at a large-scale, not requiring additional and/or scarce datasets, and not using specialised software. These are mostly intricate indicators in the microclimate domain (Sadeghi, Wood, Samali, & de Dear, 2020; Yuan et al., 2019), such as frontal area density, that is, “the area of building surface that approach the dominant wind direction in the area” (Ma & Chen, 2020). Another example is the sky view factor, which is influenced by buildings but also by other urban features, thus requiring also vegetation and other data to compute it (Palme, Privitera, & Rosa, 2020; Xia, Yabuki, & Fukuda, 2021). Further, such indicators are computationally complex to compute (Yuan & Chen, 2011), so scaling them would be computationally impossible, and some studies suggest that they may not always add much in comparison to other building form indicators (Gao, Zhan, Yang, & Liu, 2020).

3.3. Inventory of indicators of the urban form pertaining to buildings

We have structured the metrics into three groups: those that are computed for each building based on its characteristics or its surroundings (presented together in Section 3.3.1) and those that are calculated at a spatial unit such as a plot by aggregating building-level indicators gathered from buildings in that area (Section 3.3.2).

In spite of the clear differentiation between these building-level and aggregated indicators, they are intertwined. Computing some building-level indicators first requires computing a set of building level indicators, aggregating them at higher level, and then computing them back at the building level (e.g. rank in the size of each building in comparison to other buildings in the area).

3.3.1. Building-level indicators

Metrics computed at the building level are the essence of virtually all morphometric studies. However, only a minority of studies we have reviewed focuses on such indicators and rather concentrates on their aggregations at a higher level zone. Still, these indicators are essential in computing the aggregated counterparts, thus they should be given proper attention.

The building-level indicators are listed in Table 1. Some of the indicators are illustrated in Fig. 1, while Fig. 2 extends the exemplification by depicting building-level indicators that have been derived from the buffer encircled around the building.

We identify 17 indicators that are computed from the building characteristics alone, in which the surrounding context plays no role; and 26 indicators that are computed from buildings surrounding it. All these indicators may be used to calculate derivative indicators encountered in the literature review. Further, for each indicator, it is possible to compute the *rank* as the percentile of the metric in the zone in which the building is located. For example, the percentile of the footprint area with respect to a particular zone (in our implementation, we include multiple such values since we regard multiple zones). This metric is also an example of how different levels are linked, as it cannot be computed solely from individual buildings, requiring aggregation at the zone.

The indicators are mostly self-explanatory, with a couple requiring additional explanation. There are three indicators that are computed from the minimum bounding box (MBR) of the building footprint: length, width, and area. Further, there are four indicators that parametrise the shape of the footprint: shape complexity, shape compactness, equivalent rectangular index, and number of vertices (Angel, Parent, & Civco, 2010; Basaraner & Cetinkaya, 2017).

The contextual metrics — calculated for each building from buildings that are in the buffer around it — are: number of neighbours, distances, area of footprints, and ratio neighbour height to distance. The last three are a list, which is encapsulated using eight statistical measures: minimum, median, mean, maximum, sum, standard deviation, index of dispersion (D; also known as variance-to-mean ratio), and

Table 1

List of indicators at the building level (both the independent and contextual instances, together with their ranks at the corresponding zone).

Indicator	Data type	Unit
Footprint area	Decimal	m ²
Perimeter	Decimal	m
Height	Decimal	m
Height to footprint area ratio	Decimal	m ⁻¹
Volume	Decimal	m ³
Wall area	Decimal	m ²
Envelope area	Decimal	m ²
Number of vertices	Integer	
Complexity	Decimal	
Compactness	Decimal	
Equivalent rectangular index	Decimal	
MBR ^a Length	Decimal	m
MBR ^a Width	Decimal	m
MBR ^a Area	Decimal	m ²
Orientation (azimuth)	Decimal	degree
Number of storeys	Integer	
Floor area	Decimal	m ²
Number of neighbours ^b	Integer	
Site coverage in the buffer ^b	Integer	
Distance to neighbours ^b		
– Minimum	Decimal	m
– Median	Decimal	m
– Mean	Decimal	m
– Maximum	Decimal	m
– Sum	Decimal	m
– Standard deviation (SD)	Decimal	m
– Index of dispersion (D)	Decimal	m
– Coefficient of variation (CV)	Decimal	
Neighbour footprint area ^b		
– Same 8 descriptive statistics as above		
Ratio neighbour height to distance ^{b,c}		
– Same 8 descriptive statistics as above		

^a MBR – Minimum Bounding Rectangle.

^b The size of the buffer varies in literature. In our implementation, we create three buffers using the following values: 25, 50, and 100 m.

^c The last contextual indicator measures the ratio between the average height of buildings in a buffer and the distance among them.

coefficient of variation (CV; also known as relative standard deviation).

In total, there are 43 indicators at the building-level, which are doubled to 86 if considering their ranks in the corresponding zone. However, in various ways implementing them, such number may actually multiply. For example, in our implementation (Section 4), we calculate buffers of three fixed size (25, 50, and 100 m), increasing the number of metrics. Further, in our calculations, we use multiple zones (e.g. grid, district, city, province), thus, the ranks of each metric are then computed for each hierarchical zone, multiplying the number of indicators. This example suggests why it might be difficult to create a list of indicators, and how the addition of an indicator traverses multiple levels, compounding complexity despite the seemingly straightforward nature of building indicators.

3.3.2. Aggregated indicators

The urban form is most often studied in an aggregated manner, at irregular or regular zones such as a plot, administrative unit, and grid cell (tile). The aggregated measures stem from the building-level metrics (Section 3.3.1) of all buildings in the zone and are summarised with descriptive statistics, which we have listed in the previous section. Such indicators are simply permutations of descriptive statistics and the introduced metrics, e.g. mean shape complexity of all buildings in the zone. This outlook supports many studies, including those that focus on understanding the entropy of an indicator, e.g. variation of heights of buildings in a study area, which is important for climate change mitigation (Adelia, Yuan, Liu, & Shan, 2019; Usui, 2020). Such variation may be measured by one of the calculated dispersion metrics (e.g. index of dispersion) or by understanding their range (i.e. maximum minus

minimum). Further, we have encountered studies that use the minimum and maximum value of an indicator in a zone as an aggregated measure (Cao, Luan, Liu, & Wang, 2021), which we support as well. Permuting the combinations should allow computing most, if not all, such derivative indicators that are calculated as an integration of two or more metrics.

We have noticed that while summary statistics of certain indicators are common, their permutations (all summary statistics for all indicators) are not all covered in literature. Therefore, contributing to the field, our approach expands the identified indicators, potentially revealing the usability of new indicators, which is advantageous for supporting derivative indicators that we have encountered in literature, e.g. a combination between using the mean building area with the building height in an area (Cao et al., 2020). In our implementation, we have excluded computing summary statistics where it is not meaningful to do so (e.g. sum of number of floors of all buildings). Thus, in Table 2 we denote which of the indicators we have included in our implementation. Some of them illustrated in Fig. 3.

In theory, the catalogue has 177 indicators at the zone level. When ranks are considered, the list doubles to 354. Further, note that each of the contextual indicators that is summarised in eight ways (see Table 1) may be expanded to have these statistics aggregated further according to the same descriptive statistics (e.g. coefficient of variation of the mean distance to buffered neighbours in a zone), and with ranks added, that results in 690 indicators at this level.

In practice, it is not sensible to compute them all, especially when having multiple levels of aggregation as in our implementation. Therefore, in the tool, we do not include the ‘double nested’ summary indicators, and consider only the mean values when computing the ranks.

We believe that the presented framework and list of urban form indicators is the most exhaustive one presented in literature hitherto, to the extent of our knowledge, and with the addition of new ones, presents a contribution and novelty in the field.

4. Implementation: Software and open dataset

4.1. Overview

Our implementation includes the development of a software to compute urban form indicators using building footprint data and zone data, and a large-scale high-resolution dataset that was generated using the tool. Both are described together in this section to ease the understanding of the process.

In our approach, we decide to rely on a database management system for all the steps (ingesting and processing data, computation of indicators, and their storage), which has not been investigated before. We do that for multiple reasons: efficiency, organisation, robustness, modularity, scalability, and extraction. A database also allows upgrade with new datasets, and easy querying and extraction of only the data that are required, meaning that the generated datasets can be used in a wide range of statistical and geospatial software, not being confined to one software tool or programming language. Further, this approach means that our solution is cross-platform.

The tool that we develop is a sequential set of Python and SQL scripts that are used to set the scene and process input data (building footprints and defined zones) to build a PostgreSQL database spatially enabled with PostGIS. The Python scripts allow a high degree of customisation as a number of indicators can be toggled on or off. Thereafter, a series of hundreds of statements is run to compute the indicators and store them in a structured hierarchy of tables. Further, export scripts are prepared as well to allow exporting data. Having a database approach does not exclude user-friendly desktop software — these scripts enable export of the database in many different geospatial formats that can be plugged in directly in nearly all GIS software and programming environments, leveraging the best of both worlds. The architecture of GBMI is illustrated in Fig. 4.

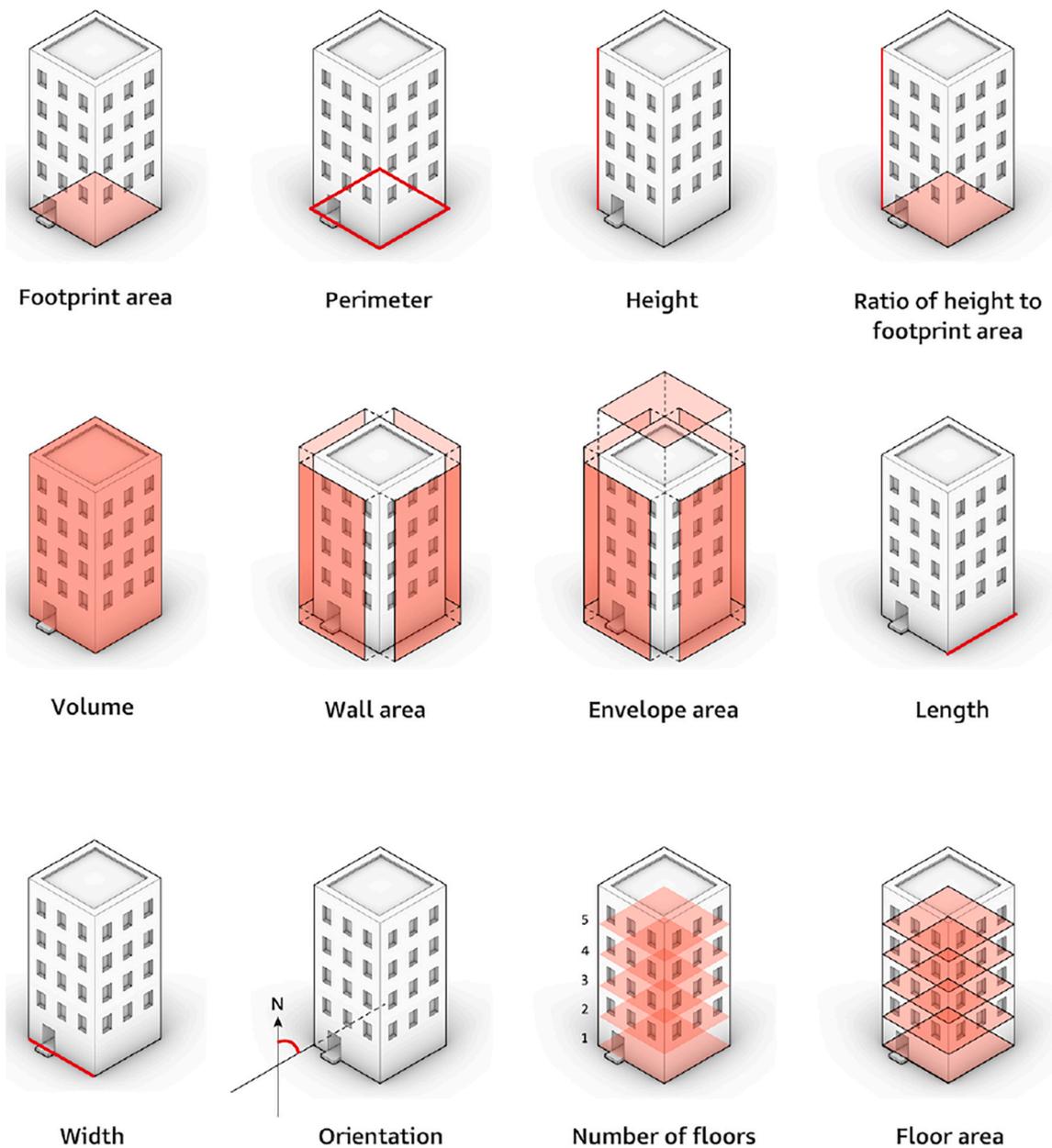
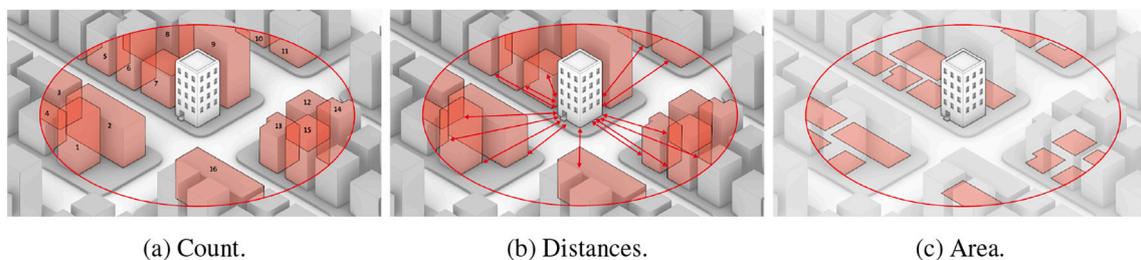


Fig. 1. Illustration of independent indicators at the building level.



(a) Count. (b) Distances. (c) Area.
 Fig. 2. Illustration of contextual aspects computed at the building level based on its surroundings.

Our implementation takes primarily OpenStreetMap as the source of building data, and two particular multi-scale zonal datasets (Section 4.2). However, the tool can be used with other types of building and zone data (e.g. if researchers have a more suitable building and/or zonal dataset, they may use our method with their data).

Both the dataset and the code used to generate it have been released as open data and open-sourced, under the Creative Commons license. As much as possible, we have designed our work to adhere to the principles laid out by the growing initiatives for openness and reproducibility in GIScience (Nüst et al., 2018; Wilson et al., 2021). The code and dataset

Table 2
Urban form measures at the aggregated level, derived from the indicators of the buildings in the corresponding area.

Building-level indicator	Summary statistics								
	Count	Min	Med	Mean	Max	Sum	SD	D	CV
Buildings	●								
Footprint area		●	●	●	●	●	●	●	●
Perimeter		●	●	●	●	●	●	●	●
Height		●	●	●	●	●	●	●	●
H/F ^a		●	●	●	●	●	●	●	●
Volume		●	●	●	●	●	●	●	●
Wall area		●	●	●	●	●	●	●	●
Envelope area		●	●	●	●	●	●	●	●
No. of vertices		●	●	●	●	●	●	●	●
Complexity		●	●	●	●	●	●	●	●
Compactness		●	●	●	●	●	●	●	●
Equiv. rectangular index		●	●	●	●	●	●	●	●
MBR ^b Length		●	●	●	●	●	●	●	●
MBR ^b Width		●	●	●	●	●	●	●	●
MBR ^b Area		●	●	●	●	●	●	●	●
Orientation		●	●	●	●	●	●	●	●
No. of storeys		●	●	●	●	●	●	●	●
Floor area		●	●	●	●	●	●	●	●
No. of neighbours		●	●	●	●	●	●	●	●
Site coverage in the buffer		●	●	●	●	●	●	●	●
Dist. to neighbours		●	●	●	●	●	●	●	●
Neighbour footprint areas		●	●	●	●	●	●	●	●
Neighbour H/D ^c		●	●	●	●	●	●	●	●

^a H/F — height-to-footprint area ratio.

^b MBR – Minimum Bounding Rectangle.

^c H/D indicates the ratio height-to-distance.

are released openly. Both contributions are supported with extensive documentation.

4.2. Input data

4.2.1. Building footprints (OpenStreetMap)

OpenStreetMap (OSM) is the leading instance of volunteered geographic information globally (Yan et al., 2020). Its use to characterise the urban form is extensive and it is often used in urban morphology studies (Barrington-Leigh & Millard-Ball, 2019; Boeing, 2020; Crooks et al., 2015; Jehling & Hecht, 2021; Valdes et al., 2020; Zhou et al., 2021), thanks to the increasing quality of data and contributors including also corporate stakeholders (Anderson, Sarkar, & Palen, 2019; Barrington-Leigh & Millard-Ball, 2017; Li et al., 2020). With the increasing quality (primarily completeness) and attention on buildings in OpenStreetMap (Hacar, 2022; Nievas et al., 2022), they have been used more often, including in some papers we have identified in our literature review (Section 3.3), e.g. (Esch et al., 2020; Liu et al., 2020). The trend of relying on OpenStreetMap as a reliable source of building data is also evident from recent papers focusing on a variety of use cases, including generating 3D building data (Braun, Padsala, Malmir, Mohammadi, & Eicker, 2021; Feldmeyer, Nowak, Jamshed, & Birkmann, 2021; Keil, Edler, Schmitt, & Dickmann, 2021; León et al., 2021; Mocnik, 2021; Palliwal, Song, Tan, & Biljecki, 2021). Despite focusing on OSM for scalability and global applicability, our tool supports other data sources too.

4.2.2. Zones

The tool supports both irregular (e.g. plots, vectors) and regular (e.g. grids, rasters) datasets for zones. Either of the two is required, but we have computed both for the sake of completeness and demonstration of both variants of zones. An advantage of using an administrative dataset is that we can enrich a building dataset with additional information from the dataset, such as the name of the district in which it is located, facilitating querying (e.g. when one needs to examine the data of a particular place) and linking to other data enabling studies across multiple disciplines.

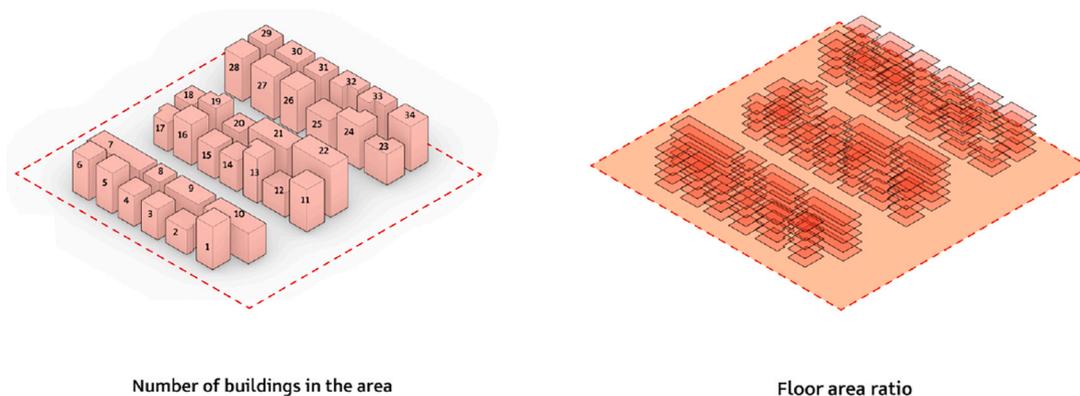
Administrative boundaries (GADM). The global administrative boundaries is obtained from the Database of Global Administrative Areas (GADM),¹ an open dataset that provides administrative hierarchical divisions starting from level 0, the national level, up to 6 levels, e.g. districts. We have loaded all the zones at all levels, constructing a multi-scale database. A hint of these boundaries is given in Section 5 in which we provide maps of indicators aggregated at administrative levels.

Grid (WorldPop). For the grid, we have used a multi-scale global raster released by WorldPop (Lloyd et al., 2019; Lloyd, Sorichetta, & Tatem, 2017; Tatem, 2017; WorldPop, 2018), at a resolution of 30 arcseconds (approx. 1 km at the equator). The raster contains 220 million cells, with an average size of 0.6 km², with a population estimate for each one, based on dasymetric redistribution (Stevens, Gaughan, Linard, & Tatem, 2015). We also include a finer variant of the raster, at 100 m resolution, providing flexibility to users. Together with the 6 levels from the administrative dataset, we have 8 zones, thus, we have 8 layers of aggregated metrics (Section 3.3.2). While we could have created our own grid, we decided to use an existing one by an established project such as WorldPop, easing linking to their data and further analyses. As the tool is flexible when it comes to the input vector data for zones, it allows other input data such as smaller plots and morphological cells when more appropriate for certain studies (Fleischmann, 2019; Fleischmann, Feliciotti, Romice, & Porta, 2020).

4.3. Computing the indicators

The indicators are computed with SQL statements and dozens of tables are generated for each hierarchy. Beyond the indicators that we have listed in Section 3.3 (barring the denoted exceptions) and despite semantic building data rarely discussed in literature, our tool regards attribute data and preserves it should a need to use it arise. For example, a dataset on the age of each building is used by Li, Koks, et al. (2020) to calculate the degree of mixing of building ages arguing that such metric

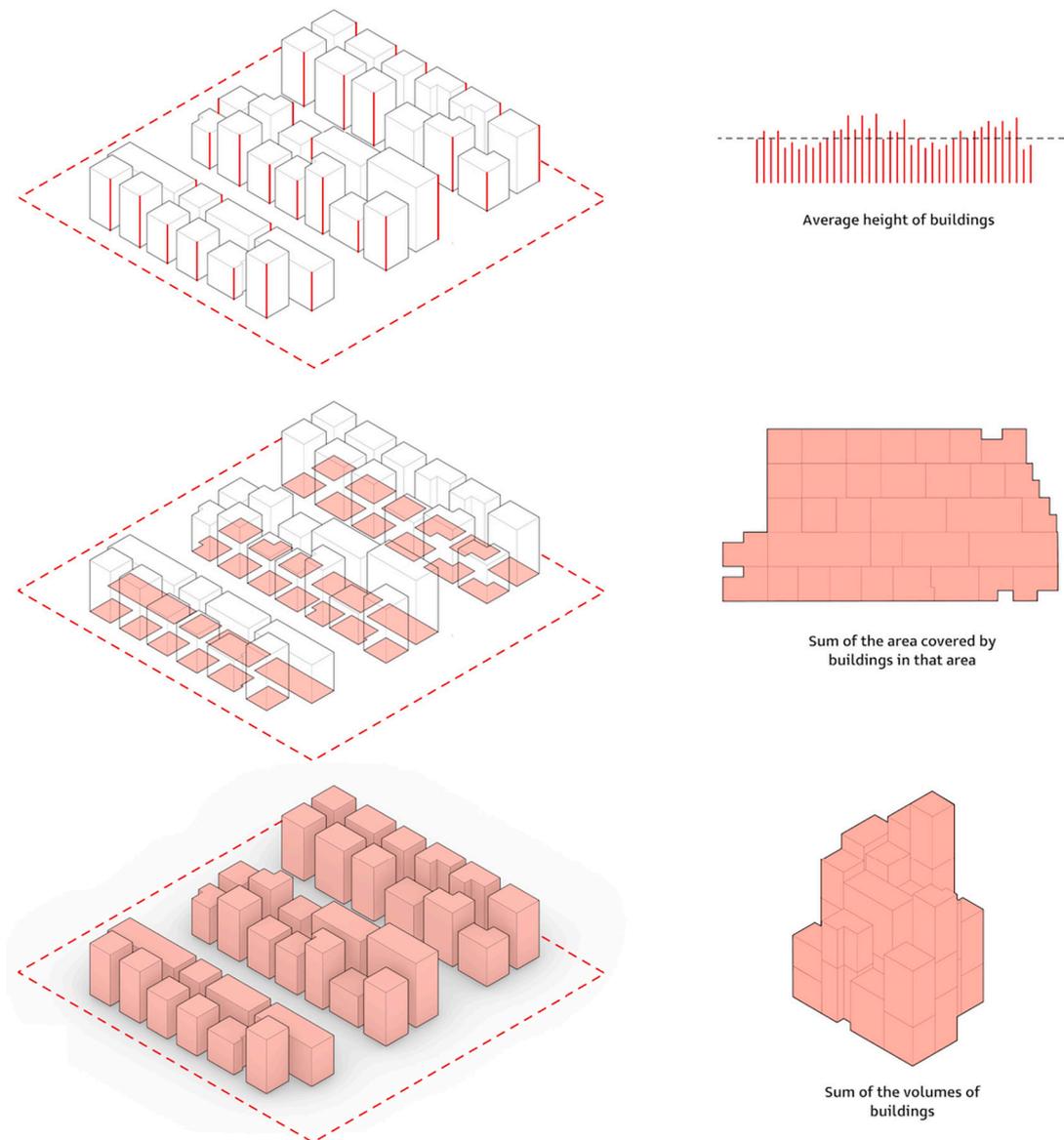
¹ <http://gadm.org/>.



Number of buildings in the area

Floor area ratio

(a) Aggregated indicators that are calculated directly based on one or more indicators at the building-level.



Average height of buildings

Sum of the area covered by buildings in that area

Sum of the volumes of buildings

(b) Examples of aggregated indicators that are based on summary statistics from an array of values such as building heights. Each of these indicators has several counterparts pertaining to the same array of values, such as minimum value and standard deviation.

Fig. 3. Examples of aggregated indicators, divided in two groups.

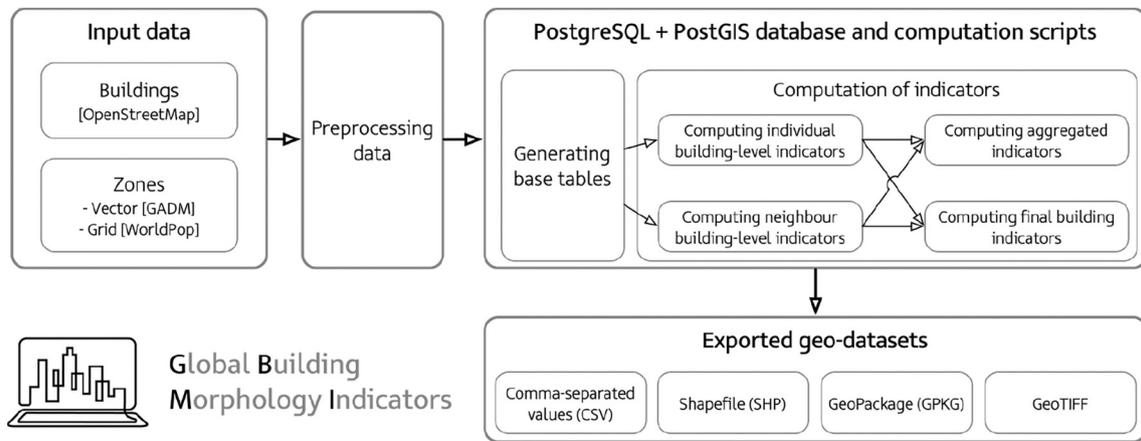


Fig. 4. Flow of the system. The datasets in the square brackets are those that we have used for generating the dataset, while the software is agnostic supporting other formats and datasets as well. All the computations are performed and kept in the database, which allows easy query and extraction of data. Our export scripts support multiple geospatial formats.

indicates urban vibrancy.

By implementing three sizes of buffers (25, 50, and 100 m), triplicate metrics of the same indicator have been computed (e.g. the mean distance between buildings in a buffer is available thrice), supporting studies with variable buffer zones (Milojevic-Dupont et al., 2020; Zhang, Cui, & Song, 2020). All these parameters are flexible and easy to modify.

4.4. Exporting the data and repository

The computed indicators may be straightforwardly queried in the database. The indicators can be exported at different levels and in multiple formats: CSV, shapefile, GeoPackage, and GeoTIFF. The documentation contains examples of several scripts for querying and exporting data into different data formats, both tabular and geospatial.

To facilitate data distribution and release a ready-to-use dataset, we have computed a dataset for a number of cities, and released it as open data in a repository. The database covers cities and countries which we attested to be well mapped in OSM, and the list is continuously growing with the addition of new locations, as the repository is intended to be a continuous development. In the Section 5, we give examples of a subset of the repository.

5. Examples of data and analyses

This section provides an insight in the implementation described in

Section 4 by showcasing the data for more than a dozen cities around the world with diverse cultural, geographical, and morphological signatures, and we also included a few countries to affirm the large-scale nature of the work. These examples serve also as an insight in their interlinking, as in them, we will demonstrate breaking down the nationwide data at multiple administrative levels further down in the hierarchy (regions, cities, neighbourhoods), and some advantages of managing the data in a database, e.g. querying to extract only the data we need for use in other software.

With these analyses, we seek to investigate the following research questions: what is the spatial distribution of the urban form in a selected set of cities? How does it vary among urban areas within the same country and around the world? How are the indicators associated among themselves? Can we link the computed metrics with various socio-economic variables and what are their relationships?

Starting from the basics, Fig. 5 presents the computed data from London across multiple levels of indicators. First, the compactness of building footprints in the study area is visualised. This metric is an example of an indicator that is computed for each building in isolation. In contrast, in the second map, we visualise an indicator that requires information on other buildings, i.e. the surrounding context — number of neighbours in 100 m buffer. Such indicators are thereafter aggregated at corresponding grid (cell) and administrative areas together with all other buildings in the same zone. The two maps on the right are examples of these: site coverage (share of area of a zone covered by

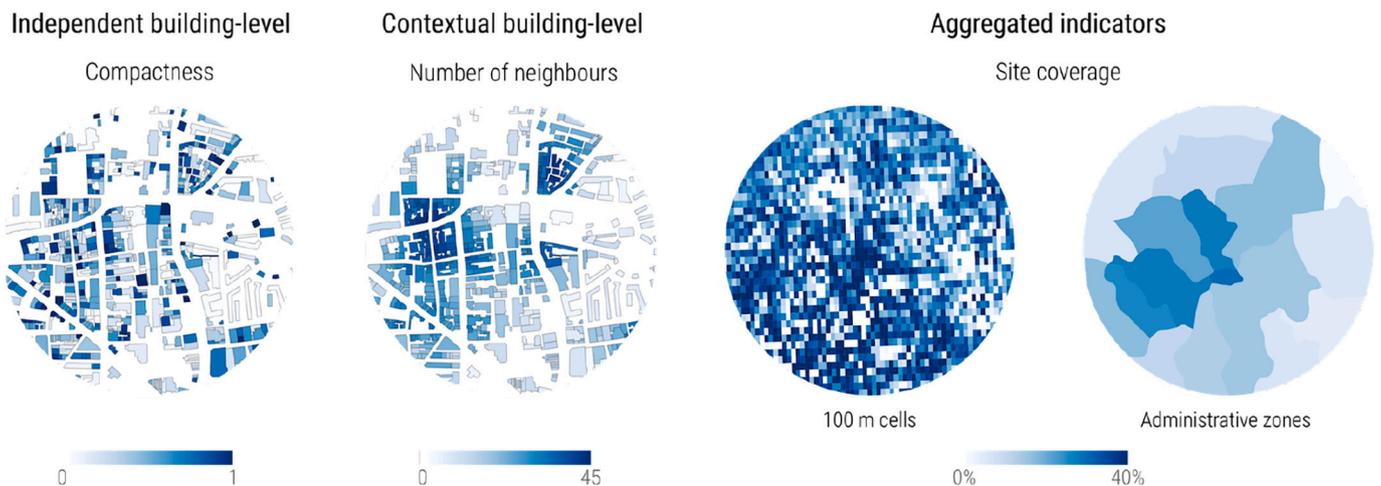


Fig. 5. Example of computed indicators at different levels for a part of London. Sources of building and administrative data: (c) OpenStreetMap contributors, GADM.

building footprints), both in a gridded aggregation (100 m) cells and administrative zones.

Moving on to cross-city analyses, in Table 3, we present a few summary statistics for a selected set of cities. Note that two indicators (the 10th and 90th percentile of the footprint area) are not standard part of GBMI. They double as an example of the flexibility of our work, enabling easy computation of new indicators if necessary.

Another advantage of the database approach is querying — we can find particular examples of zones or buildings we are interested in based on the series of metrics that are computed, e.g. a building in a city that is around both the median footprint area and median complexity (i.e. average building in a city), or finding extreme cases of buildings and zones.

In Fig. 6, we visualise the distribution of site coverage for cities, revealing characteristic urban fingerprints and asserting the advantage of our structured and hierarchical approach for conducting comparative analyses.

Fig. 7 presents a nation-wide dataset and the hierarchical structure of the work, with a visualisation of an indicator, exposing differences in the urban form at a national scale. The administrative dataset matches the subdivisions of Switzerland: cantons (level 1), districts (level 2), and sub-districts (level 3).

Urban form indicators and other proxies are often used for studying economic vibrancy and establishing the relationship with demographic, real estate, political, and other characteristics of urban areas (Botta & Gutiérrez-Roig, 2021; Chen, Zhang, & Zheng, 2021; Gebru et al., 2017; Kim, 2020; Lindenthal, 2017; Xia, Yeh, & Zhang, 2020). The administrative records in our dataset allow associating it with a variety of datasets that are available at different levels of administrative divisions to support such research. Associating the same nation-wide dataset with corresponding socio-economic indicators obtained from the federal government reveals their associations (Fig. 8). For example, the percentage of employment in an area appears to be moderately correlated with multiple disparate indicators.

Next, we show another hint at comparative analyses among cities (Fig. 9). We extract several metrics of three cities using another nation-wide dataset (New Zealand), by querying the database, and visualise them in a radar chart. The cities are compared across several metrics that are computed at cells within each city. The values have been normalised to enable comparison, thanks to which we can observe similarities and differences among cities.

Fig. 10 gives an insight in a contextual building-level indicator aggregated to the grid: the distance of a building to its nearest neighbour in its 25 m buffer, with the indicator aggregated at 100 m cells in Singapore. The map reveals distinct patterns of the urban form, which are corroborated with a cursory glance at the texture of the city-state — the western part (Clementi, top left example), is a typical high-rise setting in which buildings are well separated from each other. On the

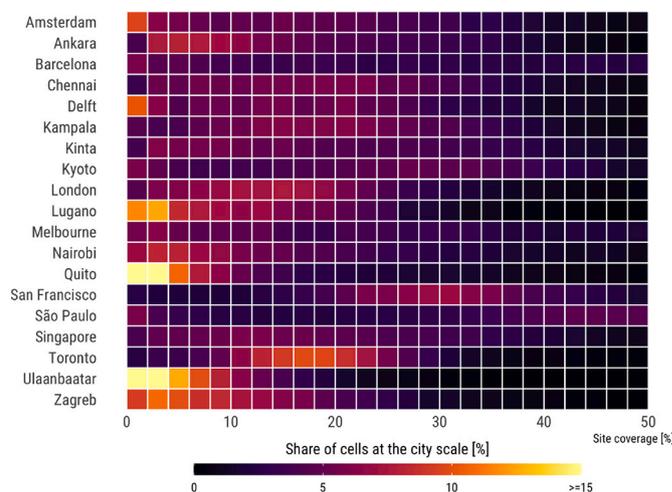


Fig. 6. Site coverage distribution across selected cities around the world.

other hand, Katong (at the bottom right) is known for its densely packed shophouses, terraced houses, and detached properties. Such indicators are crucial in microclimate simulations.

Shifting our attention to shapes, Fig. 11 visualises the distribution of complexities of buildings by city. It reveals distinct patterns of building morphology across the cities, which is interesting to associate with local cultural and architectural characteristics. Ulaanbaatar stands out, likely due to its large number of *gers*, circular dwellings used by nomads in Mongolia. The two Dutch cities (Delft and Amsterdam) are next to each other in the plot, suggesting that a similar architecture predominates nationally. Further, these two cities exhibit a bimodal distribution, possibly due to the mix of new and old architecture, i.e. the famous canal houses (*grachtenpand*), whose architecture has been shaped by taxation, presenting interesting examples how the effect of regulations may be reflected in the quantitative urban form indicators that we compute.

So far, we have presented examples of single indicators. Studying the multiplex relationships among indicators is common (Basaraner & Cetinkaya, 2017; Gisbert, Mart, & Gielen, 2017; Schwarz, 2010). As an example, we focus on three sets of association between various indicators.

First, in Fig. 12 we present scatter plots revealing the association of footprint complexity and the density of buildings in several cities, derived at the 100 m level. The plots suggests a common property of the urban form that the more buildings there are in a zone, the less complex they are.

Second, Fig. 13 indicates the correlation of dispersion metrics of the same indicator. Standard deviation (SD), index of dispersion (D), and

Table 3

Summary statistics per city — footprint area (10th percentile, mean, 90th percentile; in square metres), mean compactness, mean length (metres), and standard deviation of complexity. Two new indicators have been computed additionally to demonstrate the extensibility of the work.

City	FPA 10th	FPA mean	FPA 90th	Cpct mean	Length mean	Cplx st.dev	Buildings [1000]
Amsterdam	6.0	128.2	171.7	0.822	13.3	5.790	185
Kyoto	17.1	80.5	131.1	0.827	10.9	3.792	473
Singapore	95.4	788.2	1695.3	0.758	36.5	12.460	109
Toronto	84.0	247.2	270.7	0.801	20.2	4.705	403
Ulaanbaatar	19.3	103.9	156.5	0.882	11.6	4.456	326
Zagreb	60.1	221.0	321.8	0.845	18.2	5.793	83

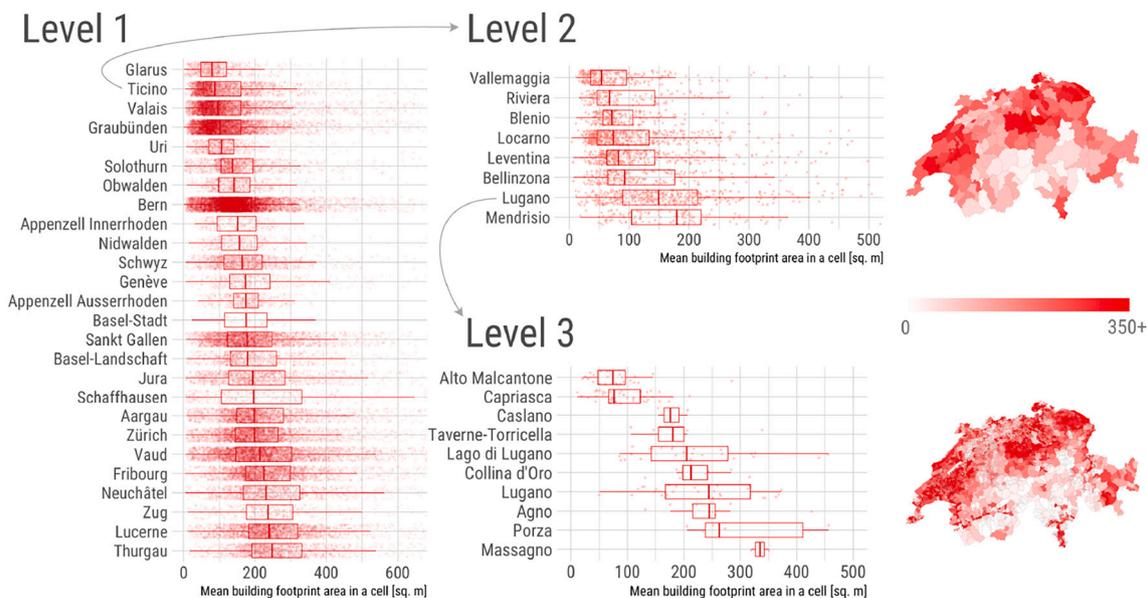


Fig. 7. Hierarchical and structured integration of data. These plots and maps were derived from footprint areas of all buildings in Switzerland, and aggregated at multiple levels. Note that for space constraints the plot does not indicate all levels, and in the lowest level we omitted some zones.



Fig. 8. Correlation between a selected set of indicators computed at the municipality level in Switzerland (level 3 in Fig. 7) and a set of demographic data available as open data by the federal government. Source of statistical data: Swiss Federal Statistical Office — Regional portraits 2021: key data of all communes.

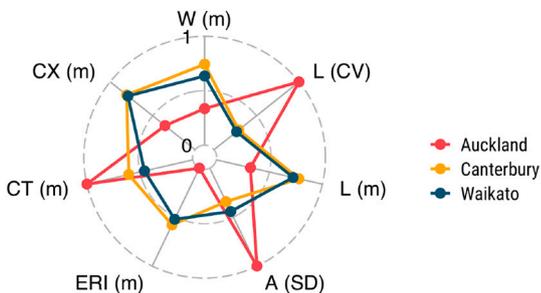


Fig. 9. A visualisation of multivariate data extracted from the dataset of New Zealand. Legend of the indicators (clockwise from top): width of the minimum bounding rectangle (mean), length of the minimum bounding rectangle (coefficient of variation and mean), azimuth (mean), equivalent rectangular index (mean), compactness (mean), and complexity (mean).

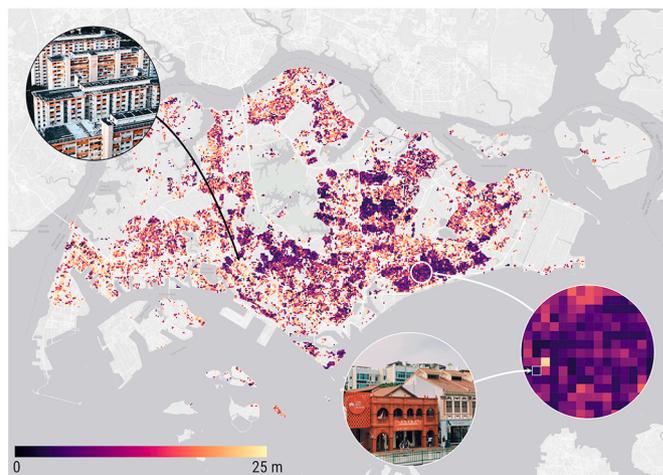


Fig. 10. A map visualising the spatial pattern of the minimum distance among buildings in high-resolution tiles across Singapore. The photographs are courtesy of Unsplash contributors.

coefficient of variation (CV) are not necessarily always perfectly correlated, affirming the importance of computing all three of them. The importance of such multiplicity is also asserted by others, e.g. Wang, Geoffroy, and Bonhomme (2021) use multiple descriptive statistics of the same indicator.

Third, plotting the measures of dispersion of two different indicators together enables us to understand their relation. Fig. 14 provides the relationship of the variation of the size of a building (index of dispersion of footprint areas) and their shape (index of dispersion of footprint complexities) in a 100 m cell. The values are aggregated at the city level, and help in understanding the diversity of buildings across multiple cities, from detecting cookie cutter neighbourhoods or cities to identifying those with a highly variable architecture.

To give an impression of the scale of the data presented in this section, we note that 110 tables are produced in the database and the exported files of the exemplified administrative regions take 223GB in size. To generate the examples, we have used R and QGIS, making the entire pipeline of GBMI and this paper entirely free and open-source

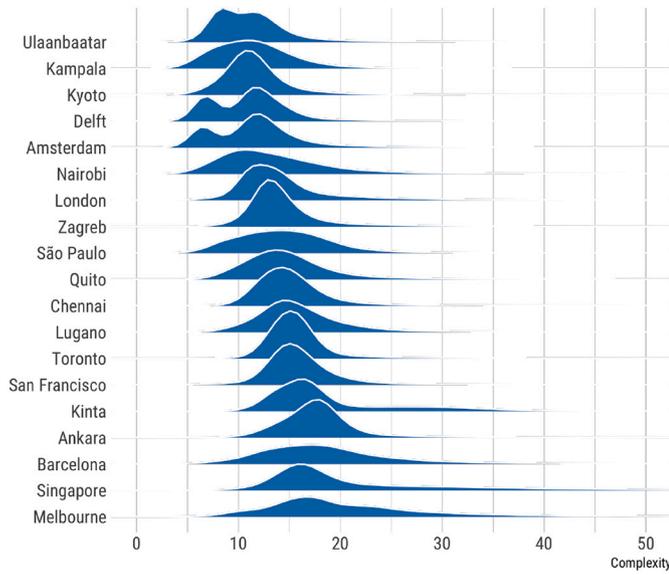


Fig. 11. Distribution of complexities of building footprints by city.

(PostgreSQL/PostGIS, OpenStreetMap and zonal data, and the mentioned statistical and geospatial software).

6. Discussion

6.1. Limitations

Our work is mainly limited by shortcomings of input data, which are out of our control. In our implementation, we focus on OSM, as it is a centralised repository of building data from around the world, with some locations mapped better than others. While in many places around the world, especially urbanised areas, building footprints have reached full completeness, in many areas heterogeneous completeness has been noticed by researchers (Varentsov, Samsonov, & Demuzere, 2020). For some of the indicators, imperfect building completeness may not be a significant issue. Provided that the sample is sufficiently representative, indicators that measure central tendency and dispersion, such as mean area and their standard deviation, may be accurate even if the full completeness is not reached. Further, the rapid growth of OSM, including in less developed countries (So & Duarte, 2020; Yuan et al., 2018), assures the addition of more cities in the repository. Also, recent datasets openly released by companies, such as Microsoft (United States) and Google (Africa) (Heris, Foks, et al., 2020; Sirko et al., 2021), and the growing volume of data released openly by governments (Biljecki, Chew, Milojevic-Dupont, & Creutzig, 2021), provide confidence that much of the globe could be processed in the future and included in our repository.

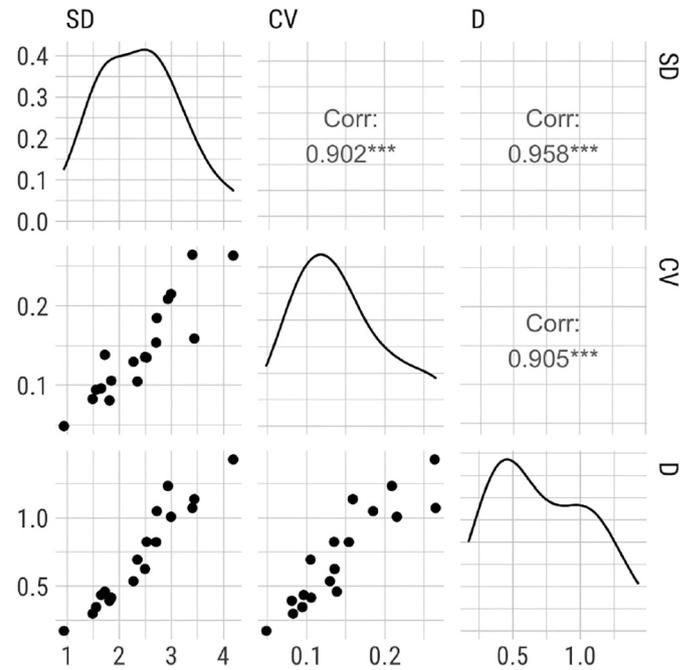


Fig. 13. The relationship between three summary indicators suggesting the dispersion of a particular indicator (complexity).

Disparity in architecture

Variation of building shapes and sizes within each city reveals their entropy

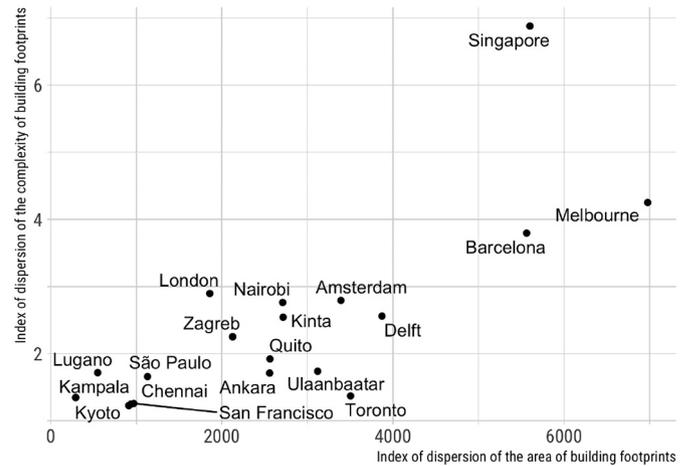


Fig. 14. Various urban form metrics, when aggregated at city level, enable us understanding flavours of cities and facilitate comparisons.

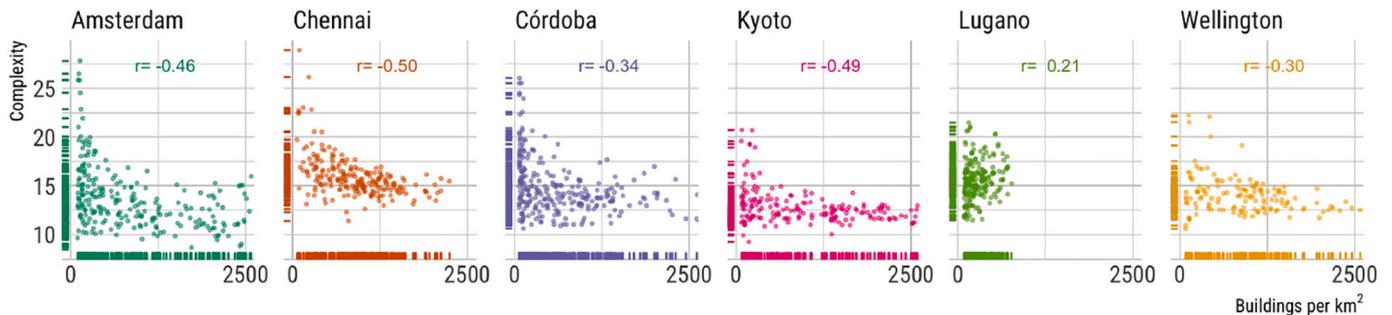


Fig. 12. The relationship between the complexity of buildings and their normalised number in a zone. These two aggregated indicators are usually negatively correlated, with some exceptions.

The completeness of attributes is another issue that will have a direct effect on the quality of the generated data, but there is some optimism as an increasing number of locations around the world, including in developing countries, is being covered with high completeness of semantic information (Biljecki, 2020). Building height datasets are a bottleneck in many areas around the world, diminishing the scalability of indicators that require the building height. Data on building heights that are fully complete are in some cases available from authoritative (government) datasets in form of building footprints enriched with attribute information on heights or as point clouds obtained from airborne lidar, but these are limited to few geographic areas. Despite commendable advancements in large-scale mapping of buildings using satellite remote sensing techniques, there are still no global open datasets on heights of individual buildings, and many instances are generated at a coarse spatial resolution (e.g. average building height at the scale of a block), limited in coverage, and/or their positional accuracy may not be fully adequate for studying the urban form at high resolution (Chen, Zhang, Wong, & Ignatius, 2020; Esch et al., 2022; Frantz et al., 2021; Geis et al., 2019; Li et al., 2020; Li, Herfort, Huang, Zia, & Zipf, 2020; Tian, Tsendbazar, van Leeuwen, Fensholt, & Herold, 2022; Zhu et al., 2022). This limitation solely pertains to our input dataset (OSM) and geographies with completeness issues. Recent open data developments such as the work of Dukai et al. (2020); Dukai, Peters, Vitalis, van Liempt, and Stoter (2021), Peters, Dukai, Vitalis, van Liempt, and Stoter (2022), and Yang and Zhao (2022), which offer nation-wide building open datasets including heights of buildings, give some assurance. As our work is designed to be data-agnostic, supporting other input geospatial datasets such as high quality instances released by governments and researchers, it remains an issue only in the generated dataset rather than the developed methodology, and using such datasets will alleviate this issue.

6.2. Directions for future work

While datasets such as point clouds and higher LoD 3D models give the means to additional indicators (Bonczak & Kontokosta, 2019; Chen, Qiu, et al., 2020), in the context of our work, it is unlikely that highly detailed 3D data will become available widely any time soon. We focus on the form of data that is available at a large scale, and we believe that 2D building footprints with attributes will continue to reign for a long time in such studies.

In our case, we have used the particular grid as it based on a dataset containing population data, potentially being useful for further analyses analysing the relationships with socio-economic variables (an example of such analysis is illustrated in Fig. 8). Notwithstanding, our tool allows using any other grid as input, so the database could be further amalgamated with other types of data from other sources that are used in urban form studies, e.g. energy consumption (Chen, Wu, & Biljecki, 2021), air temperature (Xu, Chen, Zhou, Wu, & Liu, 2020), land cover and land use (Byahut, Patel, & Mehta, 2020; Hertwig et al., 2020), wind (Allen-Dumas et al., 2020), and liveability and sustainability measures (Benita, Kalashnikov, & Tunçer, 2021; Patias, Rowe, Cavazzi, & Arribas-Bel, 2021).

In future work, it might be beneficial to recompute the indicators on another dataset of spatial boundaries and release such data openly as well, making use of other freely available large scale gridded population datasets (Palacios-Lopez et al., 2019) and aligning it to complementary work in urban morphology such as a recently released global open dataset on street network indicators (Boeing, 2021). Connecting our datasets with others may support an even wider range of studies that rely on indicators from multiple topographic features (Gamero-Salinas, Kishnani, Monge-Barrio, López-Fidalgo, & Sánchez-Ostiz, 2020; Lindberg & Grimmond, 2011; Song, Zhang, & Han, 2021). Further, recent developments in urban morphology employ deep learning and street view imagery (Biljecki & Ito, 2021; Chen, Zhang, & Zheng, 2021; Wurm et al., 2021). A viable direction for future work would be to bring them

together and complement our approach.

Formalised data integration is another potential future direction of this project. As building datasets are disseminated in formats that allow storing attribute content in a standardised manner to facilitate data exchange and interoperability, we will investigate integrating the metrics in the building dataset using a more formal approach. For example, CityGML, a prominent standard for managing and exchanging 3D building data (Kutzner, Chaturvedi, & Kolbe, 2020), allows extending it with schemas to store additional attributes (Biljecki et al., 2021), which may be applicable to the metrics we computed.

Finally, data on the urban form have been involved in temporal studies to study its evolution, with those derived with remote sensing techniques dominating such papers (Colucci, Ruvo, Lingua, Matrone, & Rizzo, 2020; Zhao, Weng, & Hersperger, 2020). Our approach enables temporal studies, as data from different periods may be processed separately and compared in a consistent manner. However, in future work, it might be beneficial to enable storing such data concurrently and provide better support for it.

7. Conclusions

Spatial data has for a long time been instrumental in characterising urban areas with regard to land use, street networks, and buildings, but not without research gaps and limitations. This work, focusing on building data and accompanying their unprecedented growth, presents a triplex contribution in understanding the structure of cities.

First, we compiled a thorough overview of metrics to quantify the urban form, based on an exploration of recent work and supplemented it with an introduction of a new array of indicators, forming a structured approach. We believe that this is the most comprehensive list on building-related morphological indicators hitherto and that — given the stringent nature of the review — our paper doubles also as a ‘mini review paper’, and we have also shared some observations encountered during the review. We have derived a comprehensive list of hundreds of indicators at multiple scales. While we do not establish a claim that our work serves as an authoritative and conclusive list of metrics of building morphology and while we do not attempt to impose a consensus on them, we hope that the framework will contribute towards their standardisation and definitive catalogue. Considering the breadth of indicators we include in GBMI, and that our literature review reveals that nearly all studies use less than 10 indicators, we believe that it may inspire researchers to experiment with metrics they would not have taken into account otherwise, potentially furthering applications and revealing new insights. Next, we believe that a combination of multiple metrics may introduce new applications in studying the urban form. Furthermore, fellow researchers are welcome to add new indicators in our modular and extensible open-source pipeline. We also believe that our inventory will serve researchers working on other types of urban data such as point clouds, from which many of the same set of indicators can be extracted.

Second, a key contribution is the implementation of a free and open-source solution, which is database-based, a novelty in the field. It supports the computation of the indicators at a very large scale and allowing using different input datasets and varying parameters, serving a multitude of studies across a range of disciplines. Statistic computations could be done directly in the database or exported to a set of files and used in statistical and geospatial software (Section 5).

Third, generation of detailed ready-to-use datasets, and their open release, providing a convenient instance that may be found useful by others and minimising their efforts in computing such data on their own, especially if they do not have access to high performance computing facilities. Considering the growing coverage of the dataset, we believe that our work may contribute towards having more studies that include more than one city, and cover cities that have previously not been subject of studies, as our literature review (Section 3.3) reveals that cross-city studies are still not common and that large swaths of land are

not subject of research. The examples in Section 5 present just a fraction of the possibilities. Further, we believe the same examples present a scientific contribution as some analyses have not been conducted before. Thanks to the administrative tags, bridges to other data can be established, with the work leading to new explorations and revealing previously unexplored relationships with a variety of indicators, some of which are novel.

In conclusion, one of the principal contributions of GBMI is that it presents a method for formalised, structured, modular, and extensible computation and management of urban indicators at a massive scale and high resolution, while the precomputed dataset allows easy and fast comparative studies in a variety of software, which is an advancement with respect to related work.

Our work is designed to carry on as a continuous development, primarily continuing the computations for further locations.

CRedit authorship contribution statement

Filip Biljecki: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Yoong Shin Chow:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – review & editing, Project administration.

Acknowledgements

We thank the editor and the reviewers for their comments. We appreciate the valuable contributions of the OpenStreetMap community. We thank our colleagues at the NUS Urban Analytics Lab for the discussions, and April Zhu for the design of illustrations. Hugo Ledoux, Anna Labetski, and Stelios Vitalis (TU Delft) are gratefully acknowledged for the discussions on data on buildings, software, and urban form. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133 and by the AWS Cloud Credits for Research.

References

- Adelia, A., Yuan, C., Liu, L., & Shan, R. (2019). Effects of urban morphology on anthropogenic heat dispersion in tropical high-density residential areas. *Energy and Buildings*, 186, 368–383. <https://doi.org/10.1016/j.enbuild.2019.01.026>
- Aktas, Y. D., Wang, K., Zhou, Y., Othman, M., Stocker, J., Jackson, M., ... Hunt, J. (2020). Outdoor thermal comfort and building energy use potential in different land-use areas in tropical cities: Case of Kuala Lumpur. *Atmosphere*, 11, 652. <https://doi.org/10.3390/atmos11060652>
- Allen-Dumas, M. R., Rose, A. N., New, J. R., Omitaomu, O. A., Yuan, J., Branstetter, M. L., Sylvester, L. M., Seals, M. B., Carvalhaes, T. M., Adams, M. B., Bhandari, M. S., Shrestha, S. S., Sanyal, J., Berres, A. S., Kolosna, C. P., Fu, K. S., & Kahl, A. C. (2020). Impacts of the morphology of new neighborhoods on microclimate and building energy. *Renewable and Sustainable Energy Reviews*, 133, Article 110030. <https://doi.org/10.1016/j.rser.2020.110030>
- Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 8, 232. <https://doi.org/10.3390/ijgi8050232>
- Angel, S., Parent, J., & Civco, D. L. (2010). Ten compactness properties of circles: Measuring shape in geography. *The Canadian Geographer*, 54, 441–461. <https://doi.org/10.1111/j.1541-0064.2009.00304.x>
- Asadi, A., Arefi, H., & Fathipour, H. (2020). Simulation of green roofs and their potential mitigating effects on the urban heat island using an artificial neural network: A case study in Austin, Texas. *Advances in Space Research*, 66, 1846–1862. <https://doi.org/10.1016/j.asr.2020.06.039>
- Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PLoS One*, 12, Article e0180698. <https://doi.org/10.1371/journal.pone.0180698>
- Barrington-Leigh, C., & Millard-Ball, A. (2019). A global assessment of street-network sprawl. *PLoS One*, 14, Article e0223078. <https://doi.org/10.1371/journal.pone.0223078>
- Basaraner, M., & Cetinkaya, S. (2017). Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. *International Journal of Geographical Information Science*, 31, 1–26. <https://doi.org/10.1080/13658816.2017.1346257>

- Benita, F., Kalashnikov, V., & Tunçer, B. (2021). A spatial livability index for dense urban centers. *Environment and Planning B: Urban Analytics and City Science*, 48, 2006–2022. <https://doi.org/10.1177/2399808320960151>
- Biljecki, F. (2020). Exploration of open data in Southeast Asia to generate 3D building models. *ISPRS annals of photogrammetry. Remote Sensing and Spatial Information Sciences*, VI-4(W1–2020), 37–44. <https://doi.org/10.5194/isprs-annals-vi-4-w1-2020-37-2020>
- Biljecki, F., Chew, L. Z. X., Milojevic-Dupont, N., & Creutzig, F. (2021). *Open government geospatial data on buildings for planning sustainable and resilient cities*. arXiv: 2107.04023.
- Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215, Article 104217. <https://doi.org/10.1016/j.landurbplan.2021.104217>
- Biljecki, F., Lim, J., Crawford, J., Moraru, D., Tauscher, H., Konde, A., Adouane, K., Lawrence, S., Janssen, P., & Stouffs, R. (2021). Extending CityGML for IFC-sourced 3d city models. *Automation in Construction*, 121, 103440. <https://doi.org/10.1016/j.autcon.2020.103440>
- Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, 126–139. <https://doi.org/10.1016/j.compenvurbysys.2017.05.004>
- Boeing, G. (2020). A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science*, 47, 590–608. <https://doi.org/10.1177/2399808318784595>
- Boeing, G. (2021). Street network models and indicators for every urban area in the world. *Geographical Analysis*. <https://doi.org/10.1111/gean.12281>
- Bonzak, B., & Kontokosta, C. E. (2019). Large-scale parameterization of 3D building morphology in complex urban landscapes using aerial LiDAR and city administrative data. *Computers, Environment and Urban Systems*, 73, 126–142. <https://doi.org/10.1016/j.compenvurbysys.2018.09.004>
- Botta, F., & Gutiérrez-Roig, M. (2021). Modelling urban vibrancy with mobile phone and OpenStreetMap data. *PLoS One*, 16, Article e0252015. <https://doi.org/10.1371/journal.pone.0252015>
- Braun, R., Padsala, R., Malmir, T., Mohammadi, S., & Eicker, U. (2021). Using 3D CityGML for the modeling of the food waste and wastewater generation—A case study for the City of Montréal. *Frontiers in Big Data*, 4, Article 662011. <https://doi.org/10.3389/fdata.2021.662011>
- Byahut, S., Patel, B., & Mehta, J. (2020). Emergence of sub-optimal land utilization patterns in Indian cities. *Journal of Urban Design*, 25, 758–777. <https://doi.org/10.1080/13574809.2020.1752646>
- Byrne, R., Hewitt, N., Griffiths, P., & MacArtain, P. (2021). Measured wind and morphological characteristics of a peri-urban environment and their impact on the performance of an operational large-scale wind turbine. *Journal of Wind Engineering and Industrial Aerodynamics*, 212, Article 104592. <https://doi.org/10.1016/j.jweia.2021.104592>
- Cao, Q., Luan, Q., Liu, Y., & Wang, R. (2021). The effects of 2D and 3D building morphology on urban environments: A multi-scale analysis in the Beijing metropolitan region. *Building and Environment*, 192, Article 107635. <https://doi.org/10.1016/j.buildenv.2021.107635>
- Cao, S., Du, M., Zhao, W., Hu, Y., Mo, Y., Chen, S., Cai, Y., Peng, Z., & Zhang, C. (2020). Multi-level monitoring of three-dimensional building changes for megacities: Trajectory, morphology, and landscape. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 54–70. <https://doi.org/10.1016/j.isprsjprs.2020.06.020>
- Carlucci, M., Zambon, I., & Salvati, L. (2019). Diversification in urban functions as a measure of metropolitan complexity. *Environment and Planning B: Urban Analytics and City Science*, 47, 1289–1305. <https://doi.org/10.1177/2399808319828374>
- Chen, H. C., Han, Q., & Vries, B. D. (2019). Urban morphology indicator analyzes for urban energy modeling. *Sustainable Cities and Society*, 52, Article 101863. <https://doi.org/10.1016/j.scs.2019.101863>
- Chen, S., Wong, N. H., Ignatius, M., Zhang, W., He, Y., Yu, Z., & Hii, D. J. C. (2022). ATLAS: Software for analysing the relationship between urban microclimate and urban morphology in a tropical city. *Building and Environment*, 208, 108591. <https://doi.org/10.1016/j.buildenv.2021.108591>
- Chen, S., Zhang, W., Wong, N. H., & Ignatius, M. (2020). Combining CityGML files and data-driven models for microclimate simulations in a tropical city. *Building and Environment*, 185, Article 107314. <https://doi.org/10.1016/j.buildenv.2020.107314>
- Chen, S., Zhang, Y., & Zheng, J. (2021). Assessment on global urban photovoltaic carrying capacity and adjustment of photovoltaic spatial planning. *Sustainability*, 13, 3149. <https://doi.org/10.3390/su13063149>
- Chen, T. H. K., Qiu, C., Schmitt, M., Zhu, X. X., Sabel, C. E., & Prishchepov, A. V. (2020). Mapping horizontal and vertical urban densification in Denmark with Landsat time-series from 1985 to 2018: A semantic segmentation solution. *Remote Sensing of Environment*, 251, Article 112096. <https://doi.org/10.1016/j.rse.2020.112096>
- Chen, W., Wu, A. N., & Biljecki, F. (2021). Classification of urban morphology with deep learning: Application on urban vitality. *Computers, Environment and Urban Systems*, 90, Article 101706. <https://doi.org/10.1016/j.compenvurbysys.2021.101706>
- Colucci, E., Ruvo, V. D., Lingua, A., Matrone, F., & Rizzo, G. (2020). HBIM-GIS integration: From IFC to CityGML standard for damaged cultural heritage in a multiscale 3D GIS. *Applied Sciences*, 10, 20–1356. <https://doi.org/10.3390/app10041356>
- Crooks, A., Pfoser, D., Jenkins, A., Croitoru, A., Stefanidis, A., Smith, D., Karagiorgou, S., Efentakis, A., & Lamprianidis, G. (2015). Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 29, 720–741. <https://doi.org/10.1080/13658816.2014.977905>

- Demuzere, M., Bechtel, B., Middel, A., & Mills, G. (2019). Mapping Europe into local climate zones. *PLoS One*, 14, Article e0214474. <https://doi.org/10.1371/journal.pone.0214474>
- Dibble, J., Prelorendjios, A., Romice, O., Zanella, M., Strano, E., Pagel, M., & Porta, S. (2017). On the origin of spaces: Morphometric foundations of urban form evolution. *Environment and Planning B: Urban Analytics and City Science*, 46, 707–730. <https://doi.org/10.1177/2399808317725075>
- Dukai, B., Peters, R., Vitalis, S., van Lierp, J., & Stoter, J. (2021). Quality assessment of a nationwide data set containing automatically reconstructed 3D building models. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVI-4/W4-2021* (pp. 17–24). <https://doi.org/10.5194/isprs-archives-xlvi-4-w4-2021-17-2021>
- Dukai, B., Peters, R., Wu, T., Commandeur, T., Ledoux, H., Baving, T., ... Stoter, J. (2020). Generating, storing, updating, and disseminating a country-wide 3D model. In *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIV-4/W1-2020* (pp. 27–32). <https://doi.org/10.5194/isprs-archives-xliv-4-w1-2020-27-2020>
- Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., & Zeidler, J. (2022). World settlement footprint 3d - a first three-dimensional survey of the global building stock. *Remote Sensing of Environment*, 270, 112877. <https://doi.org/10.1016/j.rse.2021.112877>
- Esch, T., Zeidler, J., Palacios-Lopez, D., Marconcini, M., Roth, A., Mönks, M., ... Dech, S. (2020). Towards a large-scale 3D modeling of the built environment – Joint analysis of TanDEM-X, Sentinel-2 and open street map data. *Remote Sensing*, 12, 2391. <https://doi.org/10.3390/rs12152391>
- Feldmeyer, D., Nowak, W., Jamshe, A., & Birkmann, J. (2021). An open resilience index: Crowdsourced indicators empirically developed from natural hazard and climatic event data. *Science of the Total Environment*, 145734. <https://doi.org/10.1016/j.scitotenv.2021.145734>
- Fleischmann, M. (2019). Momepy: Urban morphology measuring toolkit. *Journal of Open Source Software*, 4, 4–1807. <https://doi.org/10.21105/joss.01807>
- Fleischmann, M., Feliciotti, A., & Kerr, W. (2021). Evolution of urban patterns: Urban morphology as an open reproducible data science. *Geographical Analysis*. <https://doi.org/10.1111/gean.12302>
- Fleischmann, M., Feliciotti, A., Romice, O., & 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, Article 101441. <https://doi.org/10.1016/j.compenvurbysys.2019.101441>
- Fleischmann, M., Romice, O., & Porta, S. (2020). Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environment and Planning B: Urban Analytics and City Science*, 7. <https://doi.org/10.1177/2399808320910444>, 2399808320910444.
- Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., & Hostert, P. (2021). National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. *Remote Sensing of Environment*, 252, Article 112128. <https://doi.org/10.1016/j.rse.2020.112128>
- Gamero-Salinas, J., Kishnani, N., Monge-Barrio, A., López-Fidalgo, J., & Sánchez-Ostiz, A. (2020). The influence of building form variables on the environmental performance of semi-outdoor spaces. A study in mid-rise and high-rise buildings of Singapore. *Energy and Buildings*, 110544. <https://doi.org/10.1016/j.enbuild.2020.110544>
- Gao, S., Zhan, Q., Yang, C., & Liu, H. (2020). The diversified impacts of urban morphology on land surface temperature among urban functional zones. *International Journal of Environmental Research and Public Health*, 17, 9578. <https://doi.org/10.3390/ijerph17249578>
- Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, 114, 13108–13113. <https://doi.org/10.1073/pnas.1700035114>
- Geis, C., Leichte, T., Wurm, M., Pelizari, P. A., Standfuss, I., Zhu, X. X., ... Taubenböck, H. (2019). Large-area characterization of urban morphology—Mapping of built-up height and density using TanDEM-X and Sentinel-2 data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12, 2912–2927. <https://doi.org/10.1109/jstars.2019.2917755>
- Gisbert, F. J. G., Mart, I. C., & Gielen, E. (2017). Clustering cities through urban metrics analysis. *Journal of Urban Design*, 22, 1–20. <https://doi.org/10.1080/13574809.2017.1305882>
- Hacar, M. (2022). Analyzing the behaviors of OpenStreetMap volunteers in mapping building polygons using a machine learning approach. *ISPRS International Journal of Geo-Information*, 11, 70. <https://doi.org/10.3390/ijgi11010070>
- Heris, M. P., Foks, N. L., Bagstad, K. J., Troy, A., & Ancona, Z. H. (2020). A rasterized building footprint dataset for the United States. *Scientific Data*, 7, 207. <https://doi.org/10.1038/s41597-020-0542-3>
- Heris, M. P., Middel, A., & Muller, B. (2020). Impacts of form and design policies on urban microclimate: Assessment of zoning and design guideline choices in urban redevelopment projects. *Landscape and Planning*, 202, 103870. <https://doi.org/10.1016/j.landurbplan.2020.103870>
- Hermosilla, T., Palomar-Vázquez, J., Balaguer-Beser, T., Balsa-Barreiro, J., & Ruiz, L. A. (2014). Using street based metrics to characterize urban typologies. *Computers, Environment and Urban Systems*, 44, 68–79. <https://doi.org/10.1016/j.compenvurbysys.2013.12.002>. cited by 7
- Hertwig, D., Grimmond, S., Hendry, M. A., Saunders, B., Wang, Z., Jeoffrion, M., ... Kotthaus, S. (2020). Urban signals in high-resolution weather and climate simulations: Role of urban land-surface characterisation. *Theoretical and Applied Climatology*, 142, 701–728. <https://doi.org/10.1007/s00704-020-03294-1>
- Hu, Y., Dai, Z., & Guldmann, J. M. (2020). Modeling the impact of 2d/3d urban indicators on the urban heat island over different seasons: A boosted regression tree approach. *Journal of Environmental Management*, 266, Article 110424. <https://doi.org/10.1016/j.jenvman.2020.110424>
- Ito, K., & Biljecki, F. (2021). Assessing bikeability with street view imagery and computer vision. *Transportation Research Part C: Emerging Technologies*, 132, Article 103371. <https://doi.org/10.1016/j.trc.2021.103371>
- Jehling, M., & Hecht, R. (2021). Do land policies make a difference? A data-driven approach to trace effects on urban form in France and Germany. *Environment and Planning B: Urban Analytics and City Science*. <https://doi.org/10.1177/2399808321995818>, 2399808321995818.
- Lin, H., Cui, P., Wong, N. H., & Ignatius, M. (2018). Assessing the effects of urban morphology parameters on microclimate in Singapore to control the urban Heat Island effect. *Sustainability*, 10, 206. <https://doi.org/10.3390/su10010206>
- Jochem, W. C., & Tatem, A. J. (2021). Tools for mapping multi-scale settlement patterns of building footprints: An introduction to the R package foot. *PLoS One*, 16, Article e247535. <https://doi.org/10.1371/journal.pone.0247535>
- Keil, J., Edler, D., Schmitt, T., & Dickmann, F. (2021). Creating immersive virtual environments based on open geospatial data and game engines. *KN - Journal of Cartography and Geographic Information*, 1–13. <https://doi.org/10.1007/s42489-020-00069-6>
- Kim, Y. L. (2020). Data-driven approach to characterize urban vitality: How spatiotemporal context dynamically defines Seoul's nighttime. *International Journal of Geographical Information Science*, 34, 1235–1256.
- Kraff, N. J., Wurm, M., & Taubenböck, H. (2020). The dynamics of poor urban areas - analyzing morphologic transformations across the globe using earth observation data. *Cities*, 107, Article 102905. <https://doi.org/10.1016/j.cities.2020.102905>
- Ku, C. A., & Tsai, H. K. (2020). Evaluating the influence of urban morphology on urban wind environment based on computational fluid dynamics simulation. *ISPRS International Journal of Geo-Information*, 9, 399. <https://doi.org/10.3390/ijgi9060399>
- Kutzner, T., Chaturvedi, K., & Kolbe, T. H. (2020). CityGML 3.0: New functions open up new applications. *PFG. Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88, 43–61. <https://doi.org/10.1007/s41064-020-00095-z>
- Lemoine-Rodríguez, R., Inostroza, L., & Zepp, H. (2020). The global homogenization of urban form. An assessment of 194 cities across time. *Landscape and Urban Planning*, 204, Article 103949. <https://doi.org/10.1016/j.landurbplan.2020.103949>
- Lemoine-Rodríguez, R., Inostroza, L., & Zepp, H. (2020). Urban form datasets of 194 cities delineated based on the contiguous urban fabric for 1990 and 2015. *Data in Brief*, 33, Article 106369. <https://doi.org/10.1016/j.dib.2020.106369>
- de Lemos Martins, T. A., Adolphe, L., Bastos, L. E. G., & de Lemos Martins, M. A. (2016). Sensitivity analysis of urban morphology factors regarding solar energy potential of buildings in a Brazilian tropical context. *Solar Energy*, 137, 11–24. <https://doi.org/10.1016/j.solener.2016.07.053>
- León, J., Vicuña, M., Ogeda, A., Guzmán, S., Gubler, A., & Mokrani, C. (2021). From urban form analysis to metrics for enhancing tsunami evacuation: Lessons from twelve Chilean cities. *International Journal of Disaster Risk Reduction*, 58, Article 102215. <https://doi.org/10.1016/j.ijdr.2021.102215>
- Li, H., Herfort, B., Huang, W., Zia, M., & Zipf, A. (2020). Exploration of OpenStreetMap missing built-up areas using twitter hierarchical clustering and deep learning in Mozambique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166, 41–51. <https://doi.org/10.1016/j.isprsjprs.2020.05.007>
- Li, H., Liu, Y., Zhang, H., Xue, B., & Li, W. (2021). Urban morphology in China: Dataset development and spatial pattern characterization. *Sustainable Cities and Society*, 71, 102981. <https://doi.org/10.1016/j.scs.2021.102981>
- Li, M., Koks, E., Taubenböck, H., & van Vliet, J. (2020). Continental-scale mapping and analysis of 3D building structure. *Remote Sensing of Environment*, 245, Article 111859. <https://doi.org/10.1016/j.rse.2020.111859>
- Li, S., Wu, C., Lin, Y., Li, Z., & Du, Q. (2020). Urban morphology promotes urban vibrancy from the spatiotemporal and synergetic perspectives: A case study using multisource data in Shenzhen, China. *Sustainability*, 12, 4829. <https://doi.org/10.3390/su12124829>
- Li, X., Cheng, S., Lv, Z., Song, H., Jia, T., & Lu, N. (2020). Data analytics of urban fabric metrics for smart cities. *Future Generation Computer Systems*, 107, 871–882. <https://doi.org/10.1016/j.future.2018.02.017>
- Li, X., Ying, Y., Xu, X., Wang, Y., Hussain, S. A., Hong, T., & Wang, W. (2020). Identifying key determinants for building energy analysis from urban building datasets. *Building and Environment*, 181, Article 107114. <https://doi.org/10.1016/j.buildenv.2020.107114>
- Li, X., Zhou, Y., Gong, P., Seto, K. C., & Clinton, N. (2020). Developing a method to estimate building height from Sentinel-1 data. *Remote Sensing of Environment*, 240, Article 111705. <https://doi.org/10.1016/j.rse.2020.111705>
- Li, Y., Schubert, S., Kropp, J. P., & Rybski, D. (2020). On the influence of density and morphology on the urban heat island intensity. *Nature Communications*, 11. <https://doi.org/10.1038/s41467-020-16461-9>
- Liao, W., Hong, T., & Heo, Y. (2021). The effect of spatial heterogeneity in urban morphology on surface urban heat islands. *Energy and Buildings*, 244, Article 111027. <https://doi.org/10.1016/j.enbuild.2021.111027>
- Lindberg, F., & Grimmond, C. S. B. (2011). The influence of vegetation and building morphology on shadow patterns and mean radiant temperatures in urban areas: Model development and evaluation. *Theoretical and Applied Climatology*, 105, 311–323. <https://doi.org/10.1007/s00704-010-0382-8>
- Lindenthal, T. (2017). Beauty in the eye of the home-owner: Aesthetic zoning and residential property values. *Real Estate Economics*, 48, 530–555. <https://doi.org/10.1111/1540-6229.12204>

- Litardo, J., Palme, M., Borbor-Cordova, M., Caiza, R., Macias, J., Hidalgo-Leon, R., & Soriano, G. (2020). Urban heat island intensity and buildings' energy needs in duran, Ecuador: Simulation studies and proposal of mitigation strategies. *Sustainable Cities and Society*, 62, Article 102387. <https://doi.org/10.1016/j.scs.2020.102387>
- Liu, X., Wang, M., Qiang, W., Wu, K., & Wang, X. (2020). Urban form, shrinking cities, and residential carbon emissions: Evidence from Chinese city-regions. *Applied Energy*, 261, Article 114409. <https://doi.org/10.1016/j.apenergy.2019.114409>
- Liu, Y., Chen, C., Li, J., & Chen, W. Q. (2020). Characterizing three dimensional (3-d) morphology of residential buildings by landscape metrics. *Landscape Ecology*, 35, 2587–2599. <https://doi.org/10.1007/s10980-020-01084-8>
- Liu, Y., Li, Q., Yang, L., Mu, K., Zhang, M., & Liu, J. (2020). c. Urban heat island effects of various urban morphologies under regional climate conditions. *Science of the Total Environment*, 743, Article 140589. <https://doi.org/10.1016/j.scitotenv.2020.140589>
- Liu, Y., Xu, Y., Zhang, F., & Shu, W. (2020). d. a preliminary study on the influence of Beijing urban spatial morphology on near-surface wind speed. *Urban Climate*, 34, Article 100703. <https://doi.org/10.1016/j.uclim.2020.100703>
- Lloyd, C. T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F. R., Gaughan, A. E., Nieves, J. J., Hornby, G., MacManus, K., Sinha, P., Bondarenko, M., Sorichetta, A., & Tatem, A. J. (2019). Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data*, 3, 108–139. <https://doi.org/10.1080/20964471.2019.1625151>
- Lloyd, C. T., Sorichetta, A., & Tatem, A. J. (2017). High resolution global gridded data for use in population studies. *Scientific Data*, 4, Article e1298. <https://doi.org/10.1038/sdata.2017.1>
- Ma, T., & Chen, T. (2020). Classification and pedestrian-level wind environment assessment among Tianjin's residential area based on numerical simulation. *Urban Climate*, 34, Article 100702. <https://doi.org/10.1016/j.uclim.2020.100702>
- Milojevic-Dupont, N., & Creutzig, F. (2020). Machine learning for geographically differentiated climate change mitigation in urban areas. *Sustainable Cities and Society*, 102526. <https://doi.org/10.1016/j.scs.2020.102526>
- Milojevic-Dupont, N., Hans, N., Kaack, L. H., Zumwald, M., Andrieux, F., de Barros Soares, D., ... Creutzig, F. (2020). Learning from urban form to predict building heights. *PLoS One*, 15, Article e0242010. <https://doi.org/10.1371/journal.pone.0242010>
- Mocnik, F. B. (2021). Benford's law and geographical information – The example of OpenStreetMap. *International Journal of Geographical Information Science*, 35, 1–27. <https://doi.org/10.1080/13658816.2020.1829627>
- Mohabat Doost, D., Buffa, A., Brunetta, G., Salata, S., & Mutani, G. (2020). Mainstreaming energetic resilience by morphological assessment in ordinary land use planning. The case study of Moncalieri, Turin (Italy). *Sustainability*, 12, 4443. <https://doi.org/10.3390/su12114443>
- Morganti, M., Salvati, A., Coch, H., & Cecere, C. (2017). Urban morphology indicators for solar energy analysis. *Energy Procedia*, 134, 807–814. <https://doi.org/10.1016/j.egypro.2017.09.533>
- Ng, E., Yuan, C., Chen, L., Ren, C., & Fung, J. C. (2011). Improving the wind environment in high-density cities by understanding urban morphology and surface roughness: A study in Hong Kong. *Landscape and Urban Planning*, 101, 59–74. <https://doi.org/10.1016/j.landurbplan.2011.01.004>
- Nievas, C. I., Pilz, M., Prehn, K., Schorlemmer, D., Weatherill, G., & Cotton, F. (2022). Calculating earthquake damage building by building: The case of the city of Cologne, Germany. *Bulletin of Earthquake Engineering*, 20, 1519–1565. <https://doi.org/10.1007/s10518-021-01303-w>
- Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). Reproducible research and GIScience: An evaluation using AGILE conference papers. *PeerJ*, 6, Article e5072. <https://doi.org/10.7717/peerj.5072>
- Othman, H. A. S., & Alshboul, A. A. (2020). The role of urban morphology on outdoor thermal comfort: The case of Al-Sharq City – Az Zarqa. *Urban Climate*, 34, Article 100706. <https://doi.org/10.1016/j.uclim.2020.100706>
- Paiva, S., Santos, D., & Rossetti, R. J. F. (2018). A methodological approach for inferring urban indicators through computer vision. In *2018 IEEE International Smart Cities Conference (ISC2)* (pp. 1–7). <https://doi.org/10.1109/isc2.2018.8656934>
- Palacios-Lopez, D., Bachofer, F., Esch, T., Heldens, W., Hirner, A., Marconcini, M., ... Reinartz, P. (2019). New perspectives for mapping global population distribution using world settlement footprint products. *Sustainability*, 11, 6056. <https://doi.org/10.3390/su11216056>
- Palliwal, A., Song, S., Tan, H. T. W., & Biljecki, F. (2021). 3D city models for urban farming site identification in buildings. *Computers, Environment and Urban Systems*, 86, Article 101584. <https://doi.org/10.1016/j.compenurbysys.2020.101584>
- Palme, M., Privitera, R., & Rosa, D. L. (2020). The shading effects of green infrastructure in private residential areas: Building performance simulation to support urban planning. *Energy and Buildings*, 229, Article 110531. <https://doi.org/10.1016/j.enbuild.2020.110531>
- Patias, N., Rowe, F., Cavazzi, S., & Arribas-Bel, D. (2021). Sustainable urban development indicators in Great Britain from 2001 to 2016. *Landscape and Urban Planning*, 214, Article 104148. <https://doi.org/10.1016/j.landurbplan.2021.104148>
- Peng, L. L., Jiang, Z., Yang, X., Wang, Q., He, Y., & Chen, S. S. (2020). Energy savings of block-scale facade greening for different urban forms. *Applied Energy*, 279, Article 115844. <https://doi.org/10.1016/j.apenergy.2020.115844>
- Peters, R., Dukai, B., Vitalis, S., van Liempt, J., & Stoter, J. (2022). Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands. *Photogrammetric Engineering & Remote Sensing*, 88, 165–170. <https://doi.org/10.14358/pers.21-00032r2>
- Ribeiro, I., Martilli, A., Falls, M., Zonato, A., & Villalba, G. (2021). Highly resolved WRF-BEP/BEM simulations over Barcelona urban area with LCZ. *Atmospheric Research*, 248, Article 105220. <https://doi.org/10.1016/j.atmosres.2020.105220>
- Ryu, H., Park, I. K., Chun, B. S., & Chang, S. I. (2017). Spatial statistical analysis of the effects of urban form indicators on road-traffic noise exposure of a city in South Korea. *Applied Acoustics*, 115, 93–100. <https://doi.org/10.1016/j.apacoust.2016.08.025>
- Sadeghi, M., Wood, G., Samali, B., & de Dear, R. (2020). Effects of urban context on the indoor thermal comfort performance of windcatchers in a residential setting. *Energy and Buildings*, 219, Article 110010. <https://doi.org/10.1016/j.enbuild.2020.110010>
- Schwarz, N. (2010). Urban form revisited—Selecting indicators for characterising European cities. *Landscape and Urban Planning*, 96, 29–47. <https://doi.org/10.1016/j.landurbplan.2010.01.007>
- Sharifi, A. (2019). Resilient urban forms: A review of literature on streets and street networks. *Building and Environment*, 147, 171–187. <https://doi.org/10.1016/j.buildenv.2018.09.040>
- Silva, L. T., Oliveira, M., & Silva, J. F. (2014). Urban form indicators as proxy on the noise exposure of buildings. *Applied Acoustics*, 76, 366–376. <https://doi.org/10.1016/j.apacoust.2013.07.027>
- Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keyzers, D., Neumann, M., Cisse, M., & Quinn, J. (2021). *Continental-scale building detection from high resolution satellite imagery*. arXiv:2107.12283.
- So, W., & Duarte, F. (2020). Cartographers of North Korea: Who are they and what are the technical, political, and social issues involved in mapping North Korea. *Geoforum*, 110, 147–156. <https://doi.org/10.1016/j.geoforum.2020.02.008>
- Song, D., Lu, M., & Xing, J. (2020). Analysis of the spatial morphology facing wind environment in Harbin central area. In *Sustainability in Energy and Buildings 2020* (pp. 1–15). Singapore: Springer. https://doi.org/10.1007/978-981-15-8783-2_1
- Song, S., Leng, H., Xu, H., Guo, R., & Zhao, Y. (2020). Impact of urban morphology and climate on heating energy consumption of buildings in severe cold regions. *International Journal of Environmental Research and Public Health*, 17, 8354. <https://doi.org/10.3390/ijerph17228354>
- Song, Y., Zhang, Y., & Han, D. (2021). Access structure. *Environment and Planning B: Urban Analytics and City Science*. <https://doi.org/10.1177/2399808320988560>, 2399808320988560.
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One*, 10, Article e0107042. <https://doi.org/10.1371/journal.pone.0107042>
- Stojanovski, T. (2019). Urban form and mobility choices: Informing about sustainable travel alternatives, carbon emissions and energy use from transportation in Swedish neighbourhoods. *Sustainability*, 11, 548. <https://doi.org/10.3390/su11020548>
- Szarka, N., & Biljecki, F. (2022). Population estimation beyond counts—Inferring demographic characteristics. *PLoS One*, 17, Article e0266484. <https://doi.org/10.1371/journal.pone.0266484>
- Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, 4, Article 170004. <https://doi.org/10.1038/sdata.2017.4>
- Tian, Y., Tsendbazar, N. E., van Leeuwen, E., Fensholt, R., & Herold, M. (2022). A global analysis of multifaceted urbanization patterns using earth observation data from 1975 to 2015. *Landscape and Urban Planning*, 219, 104316. <https://doi.org/10.1016/j.landurbplan.2021.104316>
- Tian, Y., & Yao, X. (2021). Urban form, traffic volume, and air quality: A spatiotemporal stratified approach. *Environment and Planning B: Urban Analytics and City Science*. <https://doi.org/10.1177/2399808321995822>, 2399808321995822.
- Tikhonova, O., & Beirão, J. (2020). A tale of two cities - a comparative study of historical urban cores. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 13, 448–465. <https://doi.org/10.1080/17549175.2020.1753227>
- Tong, H., & Kang, J. (2021). Characteristics of noise complaints and the associations with urban morphology: A comparison across densities. *Environmental Research*, 111045. <https://doi.org/10.1016/j.envres.2021.111045>
- Tong, S., Wong, N. H., Jusuf, S. K., Tan, C. L., Wong, H. F., Ignatius, M., & Tan, E. (2018). Study on correlation between air temperature and urban morphology parameters in built environment in northern China. *Building and Environment*, 127, 239–249. <https://doi.org/10.1016/j.buildenv.2017.11.013>
- Usui, H. (2020). Building height distribution under zoning regulations: Theoretical derivation based on allometric scaling analysis and application to harmonise building heights. *Environment and Planning B: Urban Analytics and City Science*. <https://doi.org/10.1177/2399808320977867>, 2399808320977867.
- Valdes, J., Wöllmann, S., Weber, A., Klaus, G., Sigl, C., Prem, M., Bauer, R., & Zink, R. (2020). A framework for regional smart energy planning using volunteered geographic information. *Advances in Geosciences*, 54, 179–193. <https://doi.org/10.5194/adgeo-54-179-2020>
- Varentsov, M., Samsonov, T., & Demuzere, M. (2020). Impact of urban canopy parameters on a Megacity's Modelled thermal environment. *Atmosphere*, 11, 1349. <https://doi.org/10.3390/atmos11121349>
- Wang, B., Cot, L., Adolphe, L., Geoffroy, S., & Sun, S. (2017). Cross indicator analysis between wind energy potential and urban morphology. *Renewable Energy*, 113, 989–1006. <https://doi.org/10.1016/j.renene.2017.06.057>
- Wang, B., Geoffroy, S., & Bonhomme, M. (2021). Urban form study for wind potential development. *Environment and Planning B: Urban Analytics and City Science*. <https://doi.org/10.1177/2399808321994449>, 2399808321994449.
- Wang, M., & Debbage, N. (2021). Urban morphology and traffic congestion: Longitudinal evidence from US cities. *Computers, Environment and Urban Systems*, 89, Article 101676. <https://doi.org/10.1016/j.compenurbysys.2021.101676>
- Wang, M., Madden, M., & Liu, X. (2017). Exploring the relationship between urban forms and CO2 emissions in 104 Chinese cities. *Journal of Urban Planning and Development*, 143, 04017014. [https://doi.org/10.1061/\(asce\)jup.1943-5444.0000400](https://doi.org/10.1061/(asce)jup.1943-5444.0000400)
- Wilson, J. P., Butler, K., Gao, S., Hu, Y., Li, W., & Wright, D. J. (2021). A five-star guide for achieving replicability and reproducibility when working with GIS software and

- algorithms. *Annals of the American Association of Geographers*, 111, 1–7. <https://doi.org/10.1080/24694452.2020.1806026>
- Wong, N. H., Tan, C. L., Kolokotsa, D. D., & Takebayashi, H. (2021). Greenery as a mitigation and adaptation strategy to urban heat. *Nature Reviews Earth & Environment*, 1–16. <https://doi.org/10.1038/s43017-020-00129-5>
- WorldPop. (2018). *Global 1km population*. <https://doi.org/10.5258/SOTON/WP00647>
- Wu, J., Lu, Y., Gao, H., & Wang, M. (2022). Cultivating historical heritage area vitality using urban morphology approach based on big data and machine learning. *Computers, Environment and Urban Systems*, 91, Article 101716. <https://doi.org/10.1016/j.compenvurbsys.2021.101716>
- Wurm, M., Droin, A., Stark, T., Geiß, C., Sulzer, W., & Taubenböck, H. (2021). Deep learning-based generation of building stock data from remote sensing for urban heat demand modeling. *ISPRS International Journal of Geo-Information*, 10, 23. <https://doi.org/10.3390/ijgi10010023>
- Xia, B., & Li, Z. (2021). Optimized methods for morphological design of mesoscale cities based on performance analysis: Taking the residential urban blocks as examples. *Sustainable Cities and Society*, 64, Article 102489. <https://doi.org/10.1016/j.scs.2020.102489>
- Xia, C., Yeh, A. G. O., & Zhang, A. (2020). Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: A case study of five Chinese megacities. *Landscape and Urban Planning*, 193, Article 103669.
- Xia, Y., Yabuki, N., & Fukuda, T. (2021). Sky view factor estimation from street view images based on semantic segmentation. *Urban Climate*, 40, Article 100999. <https://doi.org/10.1016/j.uclim.2021.100999>
- Xu, H., Chen, H., Zhou, X., Wu, Y., & Liu, Y. (2020). Research on the relationship between urban morphology and air temperature based on mobile measurement: A case study in Wuhan, China. *Urban Climate*, 34, Article 100671. <https://doi.org/10.1016/j.uclim.2020.100671>
- Yan, Y., Feng, C. C., Huang, W., Fan, H., Wang, Y. C., & Zipf, A. (2020). Volunteered geographic information research in the first decade: A narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science*, 34, 1–27. <https://doi.org/10.1080/13658816.2020.1730848>
- Yang, C., & Zhao, S. (2022). A building height dataset across China in 2017 estimated by the spatially-informed approach. *Scientific Data*, 9. <https://doi.org/10.1038/s41597-022-01192-x>
- Yu, Z., Chen, S., & Wong, N. H. (2020). Temporal variation in the impact of urban morphology on outdoor air temperature in the tropics: A campus case study. *Building and Environment*, 181, Article 107132. <https://doi.org/10.1016/j.buildenv.2020.107132>
- Yuan, C., Adelia, A. S., Mei, S., He, W., Li, X. X., & Norford, L. (2020). Mitigating intensity of urban heat island by better understanding on urban morphology and anthropogenic heat dispersion. *Building and Environment*, 176, Article 106876. <https://doi.org/10.1016/j.buildenv.2020.106876>
- Yuan, C., & Chen, L. (2011). Mitigating urban heat island effects in high-density cities based on sky view factor and urban morphological understanding: A study of Hong Kong. *Architectural Science Review*, 54, 305–315. <https://doi.org/10.1080/00038628.2011.613644>
- Yuan, C., Shan, R., Zhang, Y., Li, X. X., Yin, T., Hang, J., & Norford, L. (2019). Multilayer urban canopy modelling and mapping for traffic pollutant dispersion at high density urban areas. *Science of the Total Environment*, 647, 255–267. <https://doi.org/10.1016/j.scitotenv.2018.07.409>
- Yuan, J., Chowdhury, P. K. R., McKee, J., Yang, H. L., Weaver, J., & Bhaduri, B. (2018). Exploiting deep learning and volunteered geographic information for mapping buildings in Kano, Nigeria. *Scientific Data*, 5. <https://doi.org/10.1038/sdata.2018.217>
- Zhang, F., Wu, L., Zhu, D., & Liu, Y. (2019). Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. *ISPRS Journal of Photogrammetry and Remote Sensing*, 153, 48–58. <https://doi.org/10.1016/j.isprsjprs.2019.04.017>
- Zhang, J., Cui, P., & Song, H. (2020). Impact of urban morphology on outdoor air temperature and microclimate optimization strategy base on pareto optimality in Northeast China. *Building and Environment*, 180, Article 107035. <https://doi.org/10.1016/j.buildenv.2020.107035>
- Zhang, J., Xu, L., Shabunko, V., Tay, S. E. R., Sun, H., Lau, S. S. Y., & Reindl, T. (2019). Impact of urban block typology on building solar potential and energy use efficiency in tropical high-density city. *Applied Energy*, 240, 513–533. <https://doi.org/10.1016/j.apenergy.2019.02.033>
- Zhang, Z., Qian, Z., Zhong, T., Chen, M., Zhang, K., Yang, Y., ... Yan, J. (2022). Vectorized rooftop area data for 90 cities in China. *Scientific Data*, 9. <https://doi.org/10.1038/s41597-022-01168-x>
- Zhao, C., Weng, Q., & Hersperger, A. M. (2020). Characterizing the 3-D urban morphology transformation to understand urban-form dynamics: A case study of Austin, Texas, USA. *Landscape and Urban Planning*, 203, Article 103881. <https://doi.org/10.1016/j.landurbplan.2020.103881>
- Zhou, Q., Lin, H., & Bao, J. (2021). Spatial autoregressive analysis of nationwide street network patterns with global open data. *Environment and Planning B: Urban Analytics and City Science*. <https://doi.org/10.1177/2399808320987846>, 239980832098784.
- Zhu, R., Wong, M. S., You, L., Santi, P., Nichol, J., Ho, H. C., Lu, L., & Ratti, C. (2020). The effect of urban morphology on the solar capacity of three-dimensional cities. *Renewable Energy*, 153, 1111–1126. <https://doi.org/10.1016/j.renene.2020.02.050>
- Zhu, X. X., Qiu, C., Hu, J., Shi, Y., Wang, Y., Schmitt, M., & Taubenböck, H. (2022). The urban morphology on our planet – Global perspectives from space. *Remote Sensing of Environment*, 269, 112794. <https://doi.org/10.1016/j.rse.2021.112794>