3D building metrics for urban morphology

Anna Labetski, Stelios Vitalis, Filip Biljecki, Ken Arroyo Ohori & Jantien Stoter

Published online: 01 Aug 2022.

Submit your article to this journal ⬚

Article views: 2105

View related articles ⬚

View Crossmark data ⬚

Citing articles: 1 View citing articles ⬚

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

3D building metrics for urban morphology

Anna Labetski[a] , Stelios Vitalis[a] , Filip Biljecki[b,c] , Ken Arroyo Ohori[a] and Jantien Stoter[a]

[a]3D Geoinformation Group, Delft University of Technology, Delft, The Netherlands; [b]Department of Architecture, National University of Singapore, Singapore, Singapore; [c]Department of Real Estate, National University of Singapore, Singapore, Singapore

ABSTRACT

Urban morphology is important in a broad range of investigations across the fields of city planning, transportation, climate, energy, and urban data science. Characterising buildings with a set of numerical metrics is fundamental to studying the urban form. Despite the rapid developments in 3D geoinformation science, and the growing 3D data availability, most studies simplify buildings to their 2D footprint, and when taking their height into account, they at most assume one height value per building, i.e. simple 3D. We take the first step in elevating building metrics into full/true 3D, uncovering the use of higher levels of detail, and taking into account the detailed shape of a building. We set the foundation of the new research line on 3D urban morphology by providing a comprehensive set of 3D metrics, implementing them in openly released software, generating an open dataset containing 2D and 3D metrics for 823,000 buildings in the Netherlands, and demonstrating a use case where clusters and architectural patterns are analysed through time. Our experiments suggest the added value of 3D metrics to complement existing counterparts, reducing ambiguity, and providing advanced insights. Furthermore, we provide a comparative analysis using different levels of detail of 3D building models.

1. Introduction

In the past years, research in 3D city modelling, as a means to model the built environment in three dimensions, has gained momentum. Its ecosystem, from acquisition to data management and applications, is now relatively developed (Dukai *et al.* 2020, Gil 2020, Vitalis *et al.* 2020, Lucks *et al.* 2021, Noardo *et al.* 2021, Nys and Billen 2021, Santhanavanich and Coors 2021, Virtanen *et al.* 2021, Wysocki *et al.* 2021, Biljecki *et al.* 2021b), prompting us to investigate their application in characterising the urban form and architecture. 3D city models can be acquired with airborne and satellite techniques, and they are useful across many use cases, from understanding noise pollution

to estimating the urban heat island effect to supporting urban farming (Chen *et al.* 2020, Beran *et al.* 2021, Palliwal *et al.* 2021, Park *et al.* 2021), and with this paper, we aspire to add 3D urban morphology to the list.

Data on buildings is a key ingredient in studying the urban form. The most common instance of geospatial data describing the location, shape, and size of buildings are 2D footprints (Biljecki *et al.* 2021a). Thanks to recent advancements in their acquisition (Xie *et al.* 2019, Sun *et al.* 2020), footprints have been increasingly available worldwide (Brovelli and Zamboni 2018, Heris *et al.* 2020, Zhang *et al.* 2022), and they may contain associated attributes such as building type and number of storeys (Malhotra *et al.* 2022). Such footprint data has been instrumental in the majority of studies that study the form of the built environment to establish patterns with respect to its effects on energy, climate, transportation, and other urban phenomena.

Urban morphology is a popular approach to the study of the characteristics of a city. One component of urban morphology studies involves computing indicators that encapsulate the urban form numerically, often by dissecting and aggregating the size, shape, and spatial distribution of city features, fueling studies in city planning, transportation, climate, energy, urban data science, and more (Boeing 2021, Fleischmann *et al.* 2021, Luan and Fuller 2021, Mei and Yuan 2021, Tian and Yao 2021, Tong and Kang 2021, Wang and Debbage 2021, Chen *et al.* 2021a, 2021b, Biljecki and Chow 2022, Coutrot *et al.* 2022). Much of the recent developments focuses on buildings, being spurred by the large-scale availability of data, primarily 2D building footprints (Huang and Wang 2020, Sarretta and Minghini 2021, Sirko *et al.* 2021, Wu and Biljecki 2022).

Several cities around the world have openly released their 3D building data, such as Berlin,[1] Helsinki,[2] and New York City.[3] However, it appears that the multidisciplinary community driving urban morphology has not yet fully embraced 3D city models, as much of their potential is still not taken advantage of. That is, the detailed geometry of buildings, offered by contemporary 3D building models, remains underused, as related work continues to settle with basic descriptors of buildings such as a uniform height, which from the perspective of GIScience, is considered 2.5D (2D feature with an associated single height) rather than true 3D. The same goes for related developments in the geospatial community, such as point clouds and street view imagery, which offer a more detailed insight into the building form, but have not been fully exploited in the context of urban morphology and developments continue to utilise basic concepts that have evolved little from their chiefly 2D/2.5D origins (Leng *et al.* 2020, Ito and Biljecki 2021).

In this paper, as we posit that 3D city models have not been made the most use of in the realm of urban morphology, we take the first step towards true 3D. This topic is important because 2D and 2.5D metrics, which we will outline in the next section (Section Context and related work), may not fully capture the complexity of the real-world urban form depending on a specific use case. Furthermore, some 2D and 2.5D metrics entail ambiguity that may impair analyses (Figure 1). In this illustration, two common 2D metrics (footprint area and convexity, which quantifies the deviation of a polygon from its convex hull) have approximately the same values for three
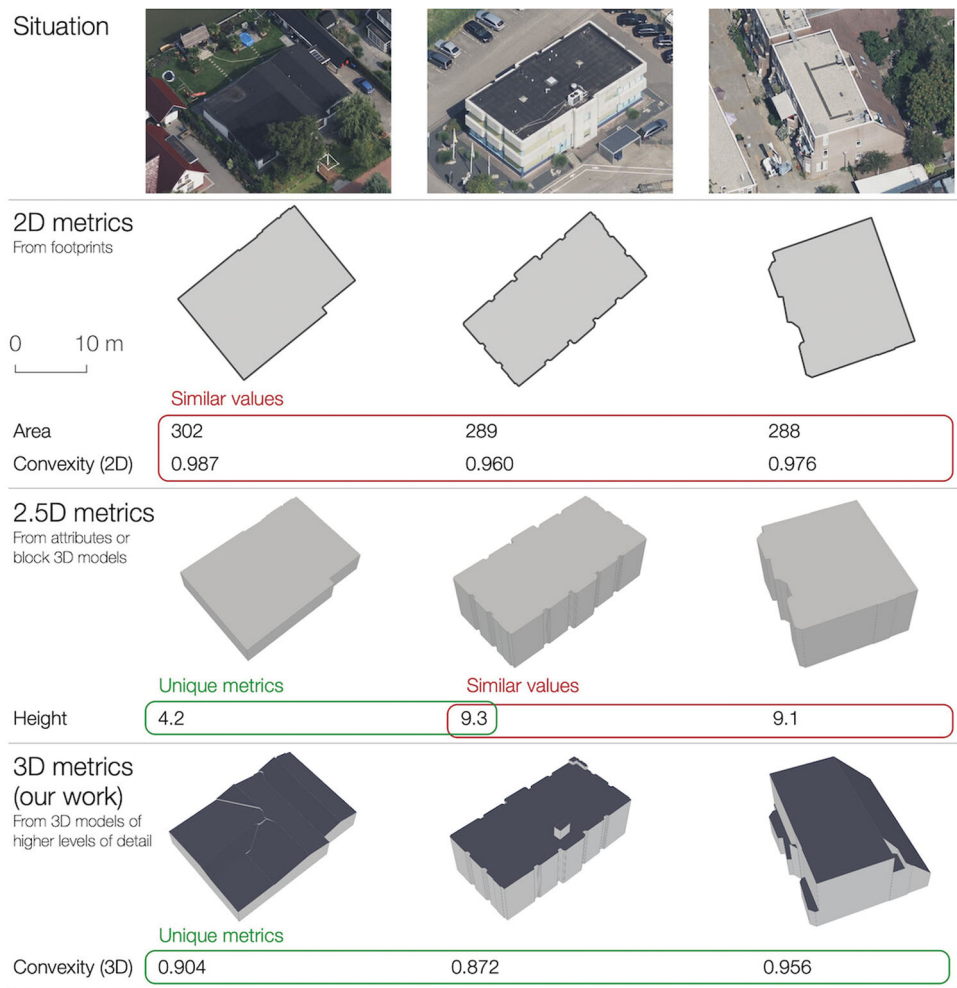
Figure 1. Motivation for advancing the dimensionality of urban morphology: distinct buildings may have the same values of traditionally used 2 D and 2.5 D metrics, thus, going fully 3 D is required to reduce ambiguity and enhance the characterisation. Data courtesy of 3 D BAG[11] and Bing Maps.[12]

significantly different buildings. 2.5D metrics, e.g. including the height as a descriptor, set apart the first building from the last two. However, the last two buildings have the same height, and thus, the 3 metrics shown so far are approximately equal, even though the architecture of the buildings is substantially different. Taking advantage of more detailed 3D building models, by computing more advanced and descriptive 3D metrics, alleviates the ambiguity and is more indicative in capturing the shape of buildings. Higher level of detail (LoD) 3D models often enable semantically rich information (also illustrated in Figure 1), such as the area of the roof, another potentially useful descriptor that we seek to leverage in our work.

The work presented in this paper is timely given the increasing availability of open 3D datasets, as there are opportunities to extend traditional 2D metrics to a 3D perspective, not only in theory but also in practice and at a large scale. Therefore, we

have developed a set of 3D metrics to characterise buildings in a three-dimensional manner (Section The metrics). These metrics give a considerably better insight into the building form, which we demonstrate with experiments (Section Data analysis). It is important to note that advanced metrics may also be applied to simple (block) 3D models (third row in Figure 1), and thus, our work also provides value to the simpler data that is already widely in use in the community, without requiring the less common datasets with a higher level of detail.

In this paper, we present our software package that ingests 3D city models and computes their 2D and 3D metrics at the building level and stores them in a structured manner to enable data analyses, including those at the urban scale with aggregated metrics (Section Implementation and Section Use cases). We release the tool as open-source software. As detailed 3D models of buildings are more widely used and there is an increase in availability worldwide, our tool is ready to make use of them where available, as demonstrated by another result of the work—we generate a dataset characterising 823,000 buildings in four cities in the Netherlands, which we release openly (Section Implementation). We provide a detailed data analysis to understand the benefits of the newly introduced metrics as well as difficulties and how to overcome them, and we compare between 2D and 3D as well as between different levels of detail (Section Data analysis). Two use cases demonstrating applications of the dataset are also presented in the paper with A) principal components analysis was driven clustering and B) examining the change of architectural patterns over time (Section Use cases).

2. Context and related work

2.1. Data on buildings and 3D city models

3D city models support the representation of both geometry and semantics. For every 3D building, individual surfaces can be augmented with semantic information, such as *ground*, *roof*, or *wall*. 3D city models can be modelled in many different ways based primarily on the method to generate them. They can be described by their level of detail (LoD). A commonly accepted general taxonomy of 3D city models consists of three LoDs: block models (LoD1), simple 3D models with a generalised roof shape (LoD2), and architecturally detailed datasets (LoD3). The last one is not common, thus, in this work, we focus on the first two (i.e. LoD1 and LoD2), which are also illustrated in Figure 1. In our work, as we capitalise on the availability of 3D data at multiple LoDs, we investigate their relationship and include an analysis of how the LoD influences the metrics—an important consideration in this topic, given that different geographies have 3D datasets available in different LoDs.

2.2. 2D and 2.5D urban morphology, and a hint of 3D

Many urban morphology studies are confined to 2D. For example, Byahut et al. (2020) compute metrics such as footprint area to establish the efficiency of land utilisation. There is also a body of knowledge dedicated to studying 2D metrics. One of the key papers in this domain, and the main inspiration for developing our method, is the work by Basaraner and Cetinkaya (2017). They study 20 approaches to characterise 2D

shapes in GIS, and based on experiments, they identify a substantial overlap among several metrics and select a few distinct indicators. We take their work as a basis, and design 3D counterparts by extending the dimensionality of 2D indicators. Not all 2D indicators are straightforward in 3D and not all are meaningful; furthermore, many new ones are necessary which we study in this paper.

An increasing number of studies in the past years have been advancing the methodologies by taking into account the vertical extent of a building. In general, studies use rough measures of volume, envelope area, and ground floor area of buildings (Geis et al. 2019, Zhang et al. 2019, Milojevic-Dupont et al. 2020, Shirowzhan et al. 2020, Li et al. 2021, Santos et al. 2021, Wu et al. 2022). In certain cases, such indicators are derived from LoD1 models, while in a large number of instances, they are computed from attributes of semantically rich building footprints without any 3D/volumetric data (e.g. footprint area multiplied by the number of storeys results in height of the building, assuming a set height per storey). A substantial body of comparative research, especially in microclimate studies, has demonstrated the advantages of 2.5D metrics, such as the general volume and height of buildings over flat (2D) indicators (Huang and Wang 2019, Tian et al. 2019, Cao et al. 2021, Liu et al. 2021, Lu et al. 2021), as we also hint at in Figure 1. For example, Liu et al. (2020) use the height of buildings together with a series of 2D metrics, to establish different types of residential communities in the study area; and Hu et al. (2020) compute a series of 2D and 2.5D metrics, such as the ratio of building height to the footprint area and volume, to understand the impact of the urban form on the urban heat island across different seasons. These building-level metrics are often aggregated at a higher level, such as cells of regular grids or administrative districts, using summary statistics such as mean and standard deviation, e.g. the aforementioned work of Liu et al. (2020) which computes the variation of heights in a district and Yang et al. (2022) which generate 3D morphological zones based on building metrics.

Most of these analyses purport to be in 3D, which technically is not incorrect, but we argue that the current '3D' analysis is primitive (or pseudo-3D) and should rather be considered as what is known in GIS as 2.5D because they rely on a single height per footprint. Furthermore, the metrics that are used are simple, merely scratching the surface of understanding the 3D urban form. Therefore, apart from the lack of more detailed 3D models, we believe that even LoD1 models offer much more potential beyond just being utilised for these simple indicators, and in our work, we demonstrate that the more advanced 3D metrics we introduce can be applied to LoD1 datasets.

Another issue with the work so far is that much of the large-scale work relies on data that is often not accurate or it does not have a high resolution and/or quality, e.g. no individual buildings or height at a coarse resolution or low accuracy, see Touchaei and Wang (2015), Xu et al. (2017), Ren et al. (2020), Li et al. (2020), Fibaek et al. (2021), Zhu et al. (2022). In our paper, we conduct an analysis that is advanced both in scale (geographic coverage) and detail (individual buildings with precise geometry).

To the best of our knowledge, the only work that truly takes advantage of higher 3D data in the context of characterising the urban form is the one of Lindenthal

(2020). The study uses 3D building models with generalised roof shapes (LoD2) to compute the architectural similarity of buildings and relate it to data on real estate transactions. It was found that homogeneous districts tend to command a price premium. Therefore, the gaps in the analysis of the building form in true 3D are clear. In our work, we develop a comprehensive catalogue of 3D metrics for studying the shapes of buildings, and software to implement them.

3. The metrics

We compute a number of metrics that are related to the 3D characteristics of the buildings and their models (Table 1). We divide these metrics into four categories, based on the methodological similarities of their computation:

Geometric properties These are mostly statistics that refer to the properties of the model of a building.

Derived properties Metrics that require some form of analysis and calculations, but whose meaning is mostly direct and deterministic. They represent generic notions related to the size and proportions of a building.

Spatial distribution Metrics that focus on the relationship between a building and its neighbours.

Space indices Complex metrics that highlight more details about the shape of the building. These are mostly based on the work of Basaraner and Cetinkaya (2017). Their computation relies on complex calculations and they are independent of the building's size (see Appendix Table 1).

One interesting metric to highlight here is the concept of volume. There are multiple approaches to calculating the volume of a building. Figure 2 summarises the four approaches that we test for calculating volumes in this paper. There is the actual volume, the volume of the convex hull, the volume of the object-oriented bounding box and the volume of the axis-aligned bounding box. They are all useful depending on

**Table 1.** Metrics are computed per building based on category.

Geometric properties	Number of vertices, Number of surfaces, Number of vertices by semantic type (i.e. ground, roof, wall), Number of surfaces by semantic type (i.e. ground, roof, wall), Min/Max/Range/Mean/Median/Std/Mode height
Derived properties	Footprint perimeter, Volume, Volume of convex hull, Volume of Object-Oriented Bounding Box, Volume of Axis-Oriented Bounding Box, Volume of voxelised building, Length and width of the Object-Oriented Bounding Box, Surface area, Surface area by semantic surface, Horizontal elongation, Min/Max vertical elongation, Form factor
Spatial distribution	Shared walls, Nearest neighbour
Space indices (see Table 3)	Circularity/Hemisphericality*, Convexity 2D/3D*, Fractality 2D/3D*, Rectangularity/Cuboidness*, Squareness/Cubeness*, Cohesion 2D/3D*, Proximity 2D/3D[+], Exchange 2D/3D[+], Spin 2D/3D[+], Perimeter/Circumference*, Depth 2D/3D[+], Girth 2D/3D[+], Dispersion 2D/3D[x], Range 2D/3D*, Equivalent Rectangular/Cuboid*, Roughness[x]

*Formula-based index, size-independent by definition.
[+]Index based on interior grid points (discretised), normalised.
[x]Index based on surface grid points (discretised), normalised.
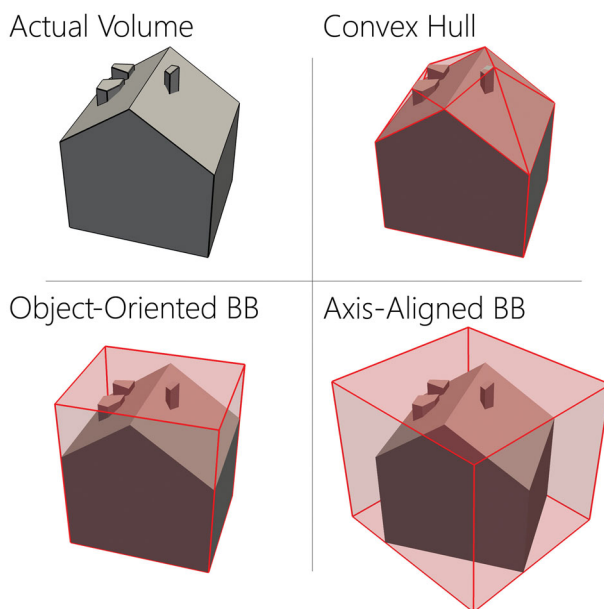
Figure 2. The different possible volume calculations for a building.

the use case but can have drastically different values depending on the building shape which can impact an application significantly.

Finally, we compute differences to compare the different LoDs of a building. Figure 3 demonstrates the various Boolean operations that we use to compare LoD1 (a simple block structure) and LoD2 (a more detailed model with variations including the roof, dormers, and chimney) of the same building. These differences can impact use cases such as calculating the energy consumption of a building.

Some metrics are more complicated than others to calculate, specific metrics in the *spatial distribution* and *shape indices* categories. Therefore, the following two sections (Sections Spatial distribution and Shape indices—measures of shape complexity in 2D and 3D) will describe those two in greater detail.

3.1. Spatial distribution

Calculating the nearest neighbour of a building is based on the *minimum* distance between a building and its nearest neighbour. A spatial index is utilised to identify neighbours of a building based on its bounding box. If there are no neighbours found based on this, then a search for the 5 closest buildings is used. There is a maximum search radius of 10 km for computational optimisation purposes.

We, also, compute the total surface area of shared walls between a building and all its neighbours. We use a method that computes the intersection between surfaces in 3D: first we compute the plane parameters for every triangle and we use clustering to group them per plane;[4] then, we project the triangles on their common plane and compute their respective 2D counterparts; finally, we dissolve the triangles to polygons and compute their intersection in 2D, from which we compute the resulting
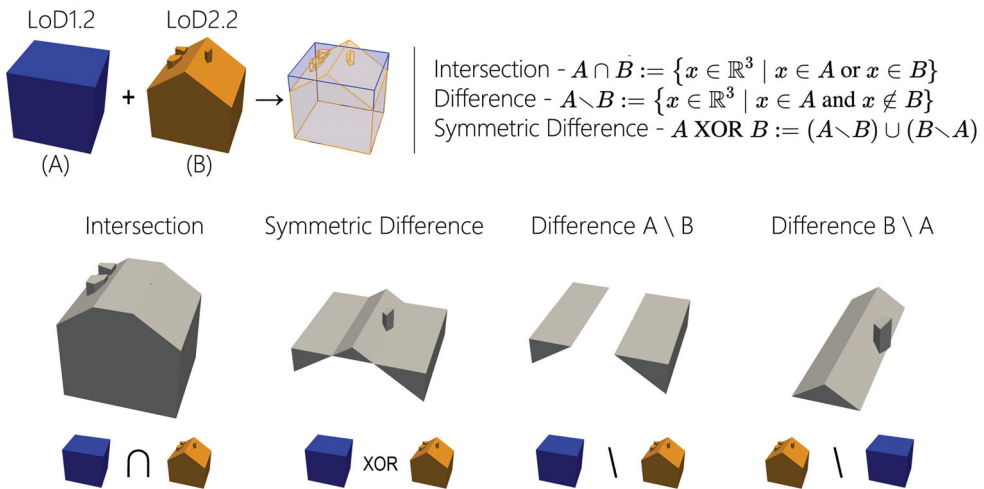
Figure 3. Boolean operations for the same building with different LoD representations.

area. This calculates the total surface area that is shared between the current building and all of its neighbours.

3.2. Shape indices—measures of shape complexity in 2D and 3D

Shape indices are measures of shape complexity that derive from interpreting the 2D measures in a 3D context from the respective 2D measures as outlined in Basaraner and Cetinkaya (2017), with additions based on the Python library *momepy* based on Fleischmann (2019). We focus on these indices given that the Basaraner and Cetinkaya (2017) paper has a very comprehensive overview of indices for 2D building footprints with a replicable methodology for us to test. We compute the indices by adopting the respective formulas to 3D space (see Appendix Table 1). Therefore we replace a polygon's area and perimeter with a polyhedron's volume and surface area, respectively. To respect the analogy of the dimensions, we adjust any constants and factors so that the resulting formulas in 3D represent similar measures to their 2D counterparts.

All shape indices are defined in a way that their values are expected to range between 0 and 1, although not in a strict sense as some indices might sometimes have a value slightly higher than 1 (but not too far from it). This is due to the methodology employing discretisation to approximate the actual value of the measures; in order to practically compute measures such as, for example, the average distance of points from the center we utilise a grid of interior points created from a voxelisation of the building and a grid of surface points on its boundaries (Figure 4). In addition, all indices are independent of a building's size, therefore only expressing its shape. To ensure this, the indices that are not by definition size-independent are normalised by using the value of the equivalent area circle of the footprint's polygon, in 2D. Similarly, in 3D, we decided to use the notion of the equivalent volume sphere of the building's polyhedron.[5] Therefore, an important aspect of 3dfying these metrics was to
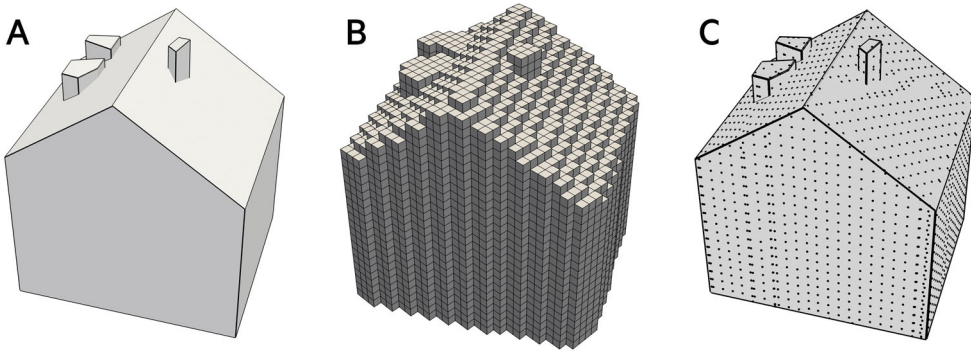
Figure 4. Example of the grid construction for a building. The two grids (interior and surface points) are used for the calculation of certain indices (see Table 1). (A) is the original building, (B) is the voxels whose centers are used as interior grid points, and (C) is the points created for the surface grid.

discover the formula of the respective measure for a sphere and use this as the denominator.

Some indices are more rudimentary, representing the resemblance of a shape with the respective primitive; these are *circularity*, *hemisphericality*, *rectangularity*, *cuboidness*, *squareness*, *cubeness*, *equivalent rectangularity*, and *equivalent cuboidness*. The rest of the indices are mostly designed to represent characteristics of the shape, for instance with respect to its compactness and thickness. Table 2 lists the definition of every index, which provides its context and purpose.

4. Implementation

4.1. Software and engineering decisions

We implemented our methodology in a Python library '3DBM – 3D Building Metrics', and a set of command-line tools that computes the metrics of buildings in a CityJSON file (Ledoux *et al.* 2019). For geometric processing, manipulation, and visualisation we used PyVista (Sullivan and Kaszynski 2019) and PyMesh.[6] We release our tool openly.[7]

In the implementation of the software, we needed to make certain engineering decisions. Due to the nature of the definitions of the indices, and the non-deterministic aspect of some elements used in them, the calculation of the indices can vary based on the implementation. Many of them are based on calculations over interior or surface grid points (Figure 4). Therefore, the density of the grid can impact the accuracy of the calculation.

We decided to use a voxelisation of the building, with the voxels being aligned to the axes of the space. The size of voxels we used for our experiments was 0.5 m, which through a process of experimentation provided the best balance between precision and computational complexity. The same reasoning applies to the surface grid points. This distance, though, can be configured accordingly to use a more dense or sparse grid.

The definition of the centroid is another non-deterministic aspect of the calculations. To begin with, there is no clear consensus on what a centroid is and what its

Table 2. Definitions of 2D and 3D measures of shape complexity.

	2D	3D
Circularity/ Hemisphericality	It measures the area deviation between a polygon and its equal-perimeter circle. Circle is generally assumed as the most compact shape.	It measures the volume deviation between a polyhedron and its equal-area hemisphere. A hemisphere was selected to represent the space above ground.
Convexity	It measures the area deviation between a polygon and its convex hull. Thus, it reveals a polygon's degree of being curved inward or outward.	It measures the volume deviation between a polyhedron and its convex hull. Thus, it reveals a building's degree of being curved inward or outward.
Fractality	It measures the edge roughness or smoothness. Based on Wentz (2010).	It measures the surface roughness or smoothness.
Rectangularity/ Cuboidness	It measures the area deviation between a polygon and its minimum area bounding rectangle. Thus, it reveals a polygon's degree of being curved inward.	It measures the volume deviation between a polyhedron and its minimum volume bounding box. Thus, it reveals a polyhedral's degree of being curved inwards.
Squareness/ Cubeness	It measures the perimeter deviation between a polygon and its equal-area square.	It measures the surface area deviation between a polyhedron and its equal-volume cube.
Cohesion	It is a measure of overall accessibility from all points to others within a polygon.	It is a measure of overall accessibility from all points to others within a polyhedron.
Proximity	It is a measure of overall accessibility from all inner points to the centre of a polygon.	It is a measure of overall accessibility from all inner points to the centre of a polyhedron.
Exchange	It measures how much of the area inside a circle is exchanged with the area outside it to create the polygon.	It measures how much of the volume inside a sphere is exchanged with the volume outside it to create the polyhedron.
Spin	It is appropriate for measuring compactness when focus is on shape extremities.	It is appropriate for measuring compactness when focus is on shape extremities.
Perimeter/ Circumference	It focuses on the compactness of a polygon's boundary.	It focuses on the compactness of a polyhedron's boundary.
Depth	It focuses on the irregular changes along the boundary of a polygon.	It focuses on the irregular changes along the boundary of a polyhedron.
Girth	It is a measure of the thickness of the layer insulating its innermost core from its periphery.	It is a measure of the thickness of the layer insulating its innermost core from its periphery.
Dispersion	As a small variant of Boyce and Clark's approach, it indicates whether a phenomenon is propagating from an epicentre equally in all directions.	As a small variant of Boyce and Clark's approach adapted to 3D. It indicates whether a phenomenon is propagating from an epicentre equally in all directions.
Range	It focuses attention on the distance between the furthest edges of a given polygon, and is therefore subject to the undue influence of small patches that are part of the polygon but far away from the rest of the polygon.	It focuses attention on the distance between the furthest faces of a given polyhedron, and is therefore subject to the undue influence of small patches that are part of the polyhedron but far away from the rest of the polygon.
Equivalent rectangular/ cuboid index	This index aims to measure deviation of a polygon from an equivalent rectangle, improving a drawback of rectangularity. Rectangularity is too sensitive to the protrusions along the boundary of a building footprint. This can cause a significant increase in the size of the	This index aims to measure deviation of a building from an equivalent cuboid. Rectangularity is too sensitive to the protrusions along the boundary of a building. This can cause a significant increase in the size of the object's minimum volume bounding box, and thus can

(*continued*)

Table 2. Continued.

|	2D	3D
	polygon's minimum area bounding rectangle, and thus can produce a misleading value for the shape. This index largely overcomes this problem by scaling the minimum area bounding rectangle until its area equals to the polygon's area.	produce a misleading value for the shape. This index largely overcomes this problem by scaling the minimum volume bounding box until its volume equals to the polyhedron's volume.
Roughness index	This index was created as a measure of compactness. It has two advantages against the indices based on area-perimeter ratio such as circularity: (1) less sensitive to the elongation (i.e. the aspect ratio of a polygon) and (2) more responsive to the roughness (i.e. intrusions and protrusions along the boundary of a polygon).	This index was created as a measure of compactness. It has two advantages against the indices based on volume to surface area ratio such as hemisphericality: (1) less sensitive to the elongation (i.e. the aspect ratio of a polyhedron) and (2) more responsive to the roughness (i.e. intrusions and protrusions along the boundary of a polyhedron).
Elongation*	Calculates elongation of a shape, seen as the elongation of its minimum bounding rectangle. Based on Gil et al. (2009).	Calculates elongation of a polyhedron, seen as the elongation of its minimum bounding box towards two different sets of axis; one is the shortest side of the bounding box against its height and the other one is the longest side of the bounding box against its height.
Form Factor*	–	A measure of the compactness of a 3D object which attempts to remove the bias introduced by the size of an object. Adapted from Bourdic et al. (2012).

2D definitions were acquired from Basaraner and Cetinkaya (2017).
*Indicates metrics acquired from Fleischmann (2019).

requirements are (e.g. if it always has to lie in the interior of an object or not); nor is there a clear definition provided in the existing bibliography about the subject. Therefore, for the calculation of the centroid of a building, we decided to use the voxel-derived interior grid. As a result, the centroid is the 'center of mass' of the voxels of a building (i.e. the mean of ordinates of the interior grid points). We chose this in order to ensure a good approximation of a center point as, to our knowledge, there is no robust algorithm to compute a 3D centroid without ensuring a valid and water-tight volume. In theory, this does not ensure that the centroid of a volume lies at its interior. Nevertheless, taking into consideration that buildings in the real world are of relatively convex and compact shapes we believe that this approach provides a reasonable approximation.

4.2. Generation of the open dataset

The second part of our implementation is the generation of a dataset containing the metrics following the methodology and using the developed software. We used the software on data describing the four largest cities in the Netherlands: Utrecht (161,545 buildings), The Hague (192,535 buildings), Rotterdam (260,414 buildings), and Amsterdam (208,895 buildings). They are the largest cities in the country and they are also different in their built environment which offers the opportunity for analysing
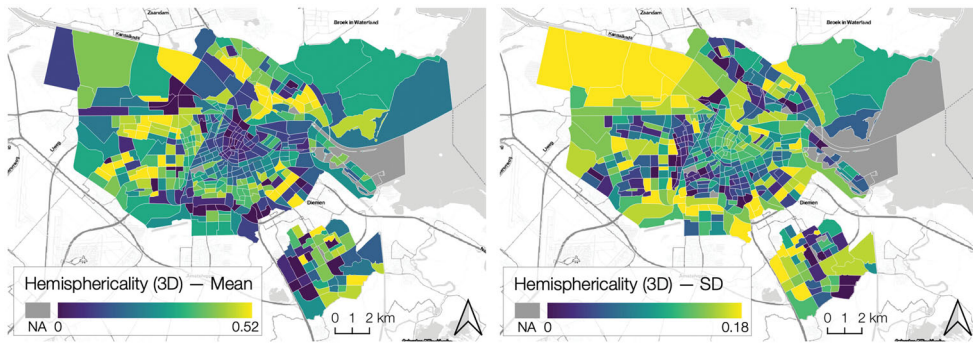
Figure 5. The 3D urban form of a city. Aggregations of one of the 3D metrics (*hemisphericality*) at the level of the administrative neighbourhoods (*buurt*) of the municipality of Amsterdam—mean as a measure of central tendency (left); and another one (standard deviation) indicating its dispersion within each area (right). The maps highlight the difference between the historical city centre with the well-known canal houses (ranking low on the *hemisphericality* index) and the newer neighbourhoods. The administrative boundaries and the basemap are courtesy of Statistics Netherlands, Stamen, and OpenStreetMap contributors.

diverse building forms. As input data (3D city models), we used 3D BAG,[8] an up-to-date open dataset containing 3D building models of the whole of the Netherlands (Peters *et al.* 2022). It contains 3D buildings at multiple LoDs, which are generated by combining a point cloud with building footprints. The 3D BAG was selected to test our implementation due to the open availability of the data, the vast geographic scope that it covers, the various LoDs, and the relatively low level of geometric errors in the data (as confirmed by Dukai *et al.* (2021) with the tool developed by Ledoux (2018)).

The result is a series of 2D and 3D metrics for each of the 823,000 buildings in the study area containing the four cities. We ran the full analysis with LoD1 and LoD2 data as well as the Boolean comparison between LoD1 and LoD2. We release our work as open data under the licence CC BY 4.0. Our resulting dataset is available at https://doi.org/10.7910/DVN/6QCRRF. The data is released as a set of CSV files, one for each LoD.

The content of the dataset will be showcased in the subsequent sections in tabular and visual ways, and also as part of two use cases in which we demonstrate the usability of the metrics. Figure 5 gives insight into the generated dataset. The building-level metrics (e.g. see the ones in Figure 1, which have been computed using our tool) have been aggregated at the government administrative level (i.e. the generated dataset was associated with a government instance (neighbourhood area))—it shows two statistical measures of an aggregation of one of the metrics, aiding in understanding the spatial distribution of the building form of a city, i.e. central tendency and and dispersion in each area.

5. Data analysis

5.1. Errors

3D calculations are not trivial and errors can occur for two main reasons. First, the input data can have geometric errors that were introduced during data creation.
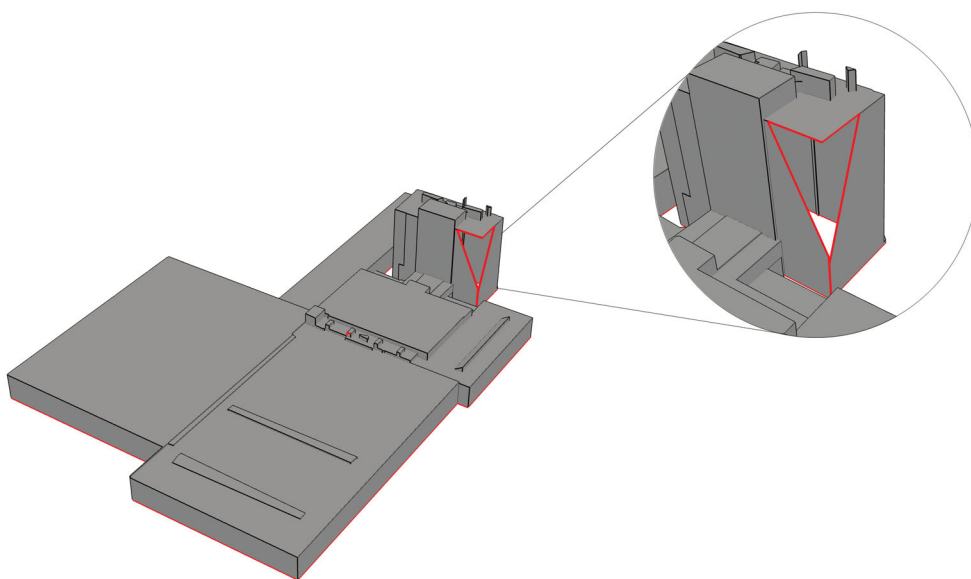
Figure 6. Example of a building with holes (i.e. not watertight) that impact calculations.

These are errors such as non-manifold edges or buildings that are non-watertight (i.e. containing holes). Second, the shapes can be too small, as the data creation method can introduce small artefacts that do not have a big enough volume to provide a reliable result for many of the indices (especially with respect to the grid density used).

In order to ensure that erroneous values do not skew our analysis, we filtered the data to exclude such cases, which was 36% of our input dataset. The filtering is based on the following rules:

1. Is a building a closed 2-manifold object based on val3dity?[9] val3dity is an open-source tool for the validation of 3D primitives according to the international standard ISO19107 (Ledoux 2018). Our main concern is that the validity of a building affects the calculation of its volume and its voxelisation. Not all geometric errors affect these calculations equally; for example, errors such as duplicate points are not an issue, while non-manifold objects are a major issue. Figure 6 shows an example of a non-watertight building with holes, for which no reliable volume or voxelisation can be computed.
2. Does the building have holes after reconstruction of the data by our implementation? This is because there can be issues with open edges introduced during the parsing of data (for example, issue related to the triangulation which is necessary for PyVista to process the data or because of float precision limitations of the software implementation).
3. What is the number of interior and surface grid points in 2D and 3D? See Figure 4. If an object is too small in relation to the provided grid density, then there will not be enough grid points generated to calculate some indices.

4. Is the actual volume bigger than the convex hull volume or is the difference between the two absurdly large in general? The convex hull calculation is more reliable than the actual volume calculation which is more sensitive to errors. The actual volume should never be significantly larger than the convex hull and therefore extreme difference are a good indicator of some implementation error in the underlying libraries used for this analysis.
5. What is the volume? Sometimes there are small fragments that are a result of the reconstruction method or smaller structures that do not constitute a building that are present in the original data. Therefore, we filter out volumes that are smaller than an average one-car garage, i.e. 40m^3.

5.2. Summary of the metric results

Table 3 summarises all of the metrics that we computed across the four cities. An overview of the indices across the four cities is summarised in Figure 7. The specific indices values for buildings of different shapes and complexities in 2D and 3D (see Figure 8) are summarised in Table 4.

Examining Table 3 we can see some interesting patterns emerge. We can see that the cities examined are quite dense as is demonstrated by a median value of 0 for the closest distance to the nearest neighbour. This is likely driven by the high level of terraced houses that are present in large Dutch cities. There are also variations in the volume calculations as is expected but it is especially stark in the axis-aligned bounding box. This is often the default setting for calculating the bounding box of a feature given that it is the easiest to compute, but it can be heavily influenced by the orientation of the building and therefore the object-oriented bounding box is a better approximation. We can also observe that while the majority of buildings in the study area are relatively small, there is a median range of 3.51 m in the range of roof values which indicates relatively large roof structures. Finally, it is interesting to observe that while the ground of a building is always one surface, the median for the number of wall surfaces is 25 while the roof modelling is typically simpler with a median value of 3.

5.2.1. Differences between 2D and 3D indices

The analysis in this section is based on the shapes of a sample of buildings as shown in Figure 8 and the values as summarised in Table 4. These buildings represent a mix of simple shapes of common building types (B1, B2, B3, and B4) as well as examples of unique and more complex buildings such as churches (B13), stadiums (B4 and B5), rounded buildings (B5, B6, B7, and B8), skyscrapers (B9, B11, and B12), and buildings with courtyards (B4 and B14).

When comparing 2D and 3D indices, B1 and B2 have very similar values (i.e. a difference of maximum 0.1) in all respective pairs, while no other building shows the same pattern. Coincidentally, they also portray very similar horizontal and maximum elongations; meaning that the horizontal and vertical profiles have very similar aspect ratios.

The relationship between the indices can also be studied. Figure 9 summarises the correlations between the 2D and 3D metrics. There are strong positive correlations between many of the metrics especially because many of the indices may be

Table 3. Summary statistics for the 3D urban morphology metrics for LoD2 for the four Dutch cities.

	Minimum	1st Quartile	Median	Mean	Third quartile	Maximum
				Stats		
Vertex count	18.00	92.00	156.00	206.07	250.00	13545.00
Unique vertex count	6.00	31.00	52.00	67.99	82.00	4485.00
Surface count	5.00	17.00	29.00	37.59	45.00	2404.00
Ground vertex count	3.00	14.00	20.00	23.85	29.00	1860.00
Wall vertex count	12.00	60.00	102.00	135.18	164.00	9140.00
Roof vertex count	3.00	20.00	33.00	46.97	57.00	4140.00
Ground surface count	1.00	1.00	1.00	1.00	1.00	1.00
Wall surface count	3.00	15.00	25.00	33.33	40.00	2251.00
Roof surface count	1.00	2.00	3.00	3.26	4.00	187.00
Max roof z (m)	−4.08	5.57	8.74	8.99	11.72	175.82
Min roof z (m)	−6.51	2.24	4.27	4.90	6.94	98.98
Height range (m)	0.00	1.07	3.51	4.10	6.08	140.63
Mean roof z (m)	−4.57	4.14	6.59	7.06	9.44	137.37
Median roof z (m)	−4.60	4.05	6.53	7.11	9.80	175.81
STD roof z (m)	0.00	0.43	1.42	1.49	2.16	63.21
Mode roof z (m)	−5.39	4.37	7.59	7.84	10.74	175.81
Ground min z (m)	−11.55	−1.38	0.11	−0.25	1.11	22.72
				Properties		
Actual volume (m^3)	40.00	314.41	438.76	1363.35	826.02	1000985.64
Convex hull volume (m^3)	40.01	330.32	480.19	1681.57	919.68	1187713.12
OBB volume (m^3)	40.07	392.26	603.39	2151.23	1140.39	1905213.43
AABB volume (m^3)	40.49	616.44	1038.07	3515.02	1878.86	2504540.69
Footprint perimeter (m)	4.13	29.35	34.46	43.08	43.66	2233.79
Surface area (m^2)	65.96	295.19	386.54	732.66	614.20	373242.85
Ground area (m^2)	0.85	48.29	60.38	154.42	91.94	125476.63
Wall area (m^2)	11.95	180.57	239.89	368.37	391.21	46158.81
Roof area (m^2)	0.85	53.82	70.48	165.24	105.90	125476.63
Horizontal elongation	0.00	0.29	0.44	0.42	0.54	0.98
Min vertical elongation	0.00	0.15	0.30	0.35	0.52	1.20
Max vertical elongation	0.00	0.21	0.38	0.38	0.53	1.20
				Relationships		
Shared walls area (m^2)	0.00	53.79	117.49	130.99	172.13	113763.06
Closest distance (m)	0.00	0.00	0.00	1.46	0.00	10000.00
				Indices		
Circularity (2D)	0.03	0.60	0.68	0.65	0.73	1.00
Hemisphericality (3D)	0.03	0.42	0.46	0.45	0.49	0.60
Convexity (2D)	0.11	0.94	0.99	0.96	1.00	1.00
Convexity (3D)	0.07	0.88	0.95	0.92	1.00	1.00
Fractality (2D)	0.20	0.40	0.42	0.42	0.43	1.05
Fractality (3D)	0.19	0.30	0.31	0.32	0.33	0.56
Rectangularity (2D)	0.07	0.87	0.97	0.92	1.00	1.00
Cuboidness (3D)	0.04	0.68	0.80	0.78	0.90	1.00
Squareness (2D)	0.18	0.88	0.93	0.91	0.97	1.13
Cubeness (3D)	0.15	0.88	0.93	0.91	0.97	1.11
Equivalent rectangularity (2D)	0.18	0.94	0.99	0.96	1.00	1.13
Equivalent cuboidness (3D)	0.23	0.95	0.98	0.97	1.00	1.18
Proximity (2d)	0.17	0.87	0.92	0.90	0.95	1.19
Proximity (3D)	0.11	0.82	0.87	0.84	0.89	1.08
Spin (2D)	0.02	0.71	0.80	0.78	0.88	1.20
Spin (3D)	0.01	0.64	0.72	0.69	0.78	1.14
Perimeter (2D)	0.16	0.78	0.83	0.81	0.86	1.00
Circumference (3D)	0.12	0.71	0.75	0.73	0.78	0.90
Depth (2D)	0.15	0.73	0.79	0.78	0.84	1.19
Depth (3D)	0.08	0.60	0.65	0.64	0.69	1.01
Girth (2D)	0.11	0.59	0.65	0.65	0.72	1.00
Girth (3D)	0.02	0.51	0.57	0.56	0.63	0.87
Dispersion (2D)	0.23	0.74	0.79	0.78	0.84	1.00
Dispersion (3D)	0.13	0.76	0.81	0.79	0.84	0.93
Range (2D)	0.13	0.67	0.72	0.70	0.76	1.00
Range (3D)	0.00	0.59	0.64	0.62	0.67	0.84
Roughness (2D)	0.04	0.73	0.80	0.77	0.83	1.19
Roughness (3D)	0.26	0.83	0.87	0.88	0.92	1.20
Form factor (3D)	0.00	0.04	0.05	0.06	0.07	1.00
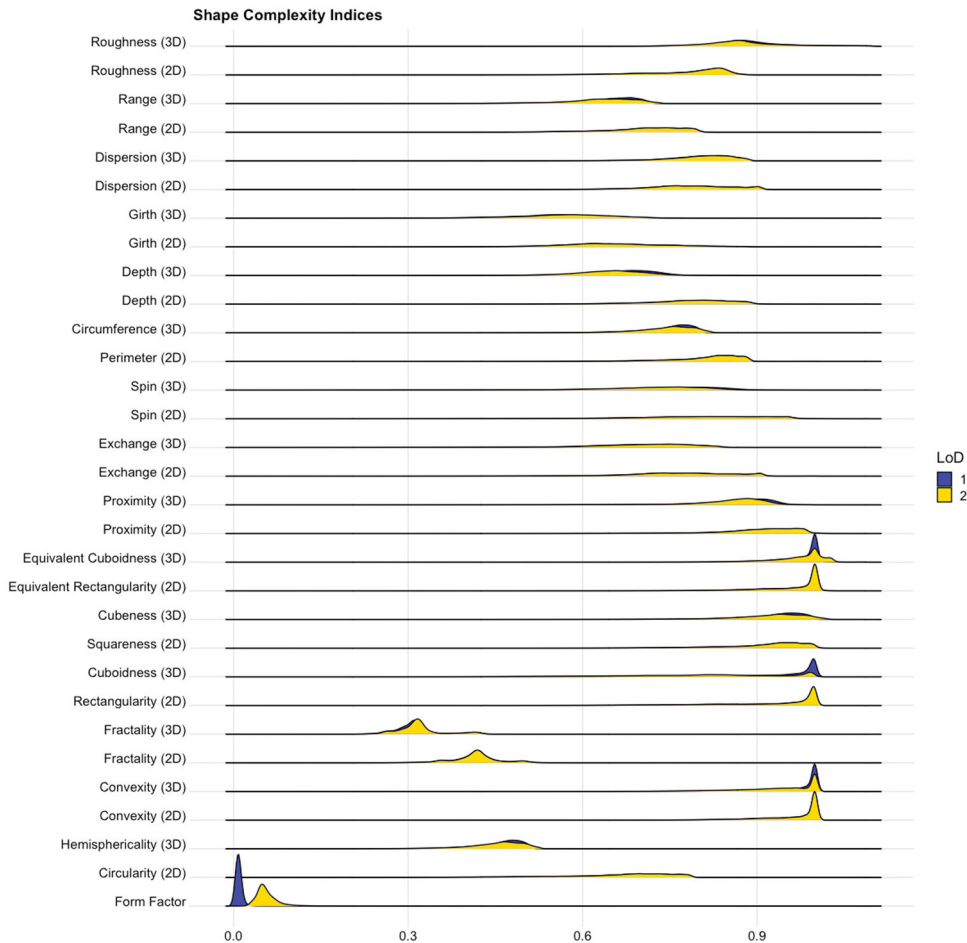
Figure 7. Index values for the measures of shape complexity across the study area by LoD.

influenced by similar variables and specifically elongation. Nevertheless, there are important nuances between the indices that can contribute independently to an analysis, such as the principal components analysis conducted in Section Clustering demonstrates. An interesting observation here is that *range* has a strong negative correlation with many of the indices, which is perhaps due to the fact that *range* focuses on the distance between the furthest faces of a given object and does not care about how regular a shape is.

5.2.2. Differences between buildings

Intuitively, we can consider B1, B2, B3, and B10 as archetypal residential buildings for the Netherlands, based on the fact that: a) they have a slanted roof (with one or two peaks); b) they are of relatively average size and proportions; and c) they have no big spikes or protrusions. For these four buildings, some of their 2D indices vary significantly, while their 3D indices are very similar (Figure 10). This is despite them having

**Table 4.** Urban morphology metrics for several building examples visualised in Figure 8.

	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14
							2D							
Circularity	0.52	0.51	0.76	0.83	0.48	0.28	0.91	0.29	0.46	0.72	0.64	0.51	0.39	0.20
Convexity	0.92	0.90	0.99	0.98	0.94	0.86	0.98	0.73	0.84	0.98	0.94	0.80	0.86	0.68
Fractality	0.43	0.45	0.36	0.21	0.24	0.32	0.32	0.32	0.28	0.40	0.31	0.30	0.30	0.32
Rectangularity	0.85	0.81	0.99	0.92	0.74	0.69	0.79	0.58	0.83	0.96	0.89	0.61	0.72	0.58
Squareness	0.81	0.81	0.98	1.03	0.78	0.60	1.08	0.61	0.76	0.96	0.90	0.80	0.70	0.50
Equivalent rectangularity	0.92	0.90	0.99	1.03	0.78	0.60	1.08	0.61	0.77	0.98	0.90	0.82	0.73	0.54
Proximity	0.82	0.83	0.98	0.99	0.99	0.98	1.00	0.84	0.91	0.95	0.98	0.85	0.92	0.67
Exchange	0.68	0.68	0.90	0.92	0.93	0.90	0.98	0.75	0.81	0.83	0.92	0.74	0.78	0.53
Spin	0.61	0.61	0.94	0.97	0.97	0.94	1.00	0.71	0.81	0.88	0.95	0.69	0.80	0.43
Perimeter	0.72	0.71	0.87	0.91	0.69	0.53	0.96	0.54	0.68	0.85	0.80	0.71	0.62	0.45
Depth	0.67	0.67	0.88	0.91	0.88	0.84	0.97	0.52	0.67	0.83	0.85	0.64	0.74	0.41
Girth	0.56	0.60	0.82	0.84	0.83	0.92	0.90	0.51	0.58	0.70	0.79	0.52	0.67	0.44
Dispersion	0.70	0.70	0.90	0.93	0.93	0.91	0.98	0.75	0.79	0.84	0.91	0.74	0.79	0.56
Range	0.59	0.59	0.79	0.87	0.87	0.85	0.97	0.73	0.73	0.75	0.78	0.63	0.71	0.54
Roughness	0.70	0.69	0.82	0.85	0.49	0.29	0.93	0.29	0.50	0.80	0.67	0.57	0.43	0.21
							3D							
Hemisphericality	0.42	0.43	0.41	0.15	0.17	0.29	0.52	0.28	0.20	0.44	0.21	0.30	0.25	0.14
Convexity	0.88	0.89	0.91	0.23	0.67	0.85	0.94	0.58	0.59	0.87	0.85	0.74	0.55	0.49
Fractality	0.32	0.33	0.31	0.25	0.21	0.23	0.24	0.23	0.22	0.30	0.24	0.21	0.23	0.27
Cuboidness	0.62	0.60	0.83	0.17	0.43	0.64	0.70	0.42	0.52	0.70	0.73	0.48	0.27	0.30
Cubeness	0.88	0.90	0.87	0.44	0.49	0.68	1.00	0.66	0.54	0.90	0.56	0.71	0.62	0.42
Equivalent cuboidness	0.96	0.97	0.96	0.61	0.57	0.90	1.01	0.69	0.57	0.93	0.64	0.76	0.64	0.66
Proximity	0.83	0.83	0.78	0.32	0.68	0.76	0.96	0.76	0.73	0.89	0.69	0.75	0.78	0.41
Exchange	0.71	–	0.57	0	–	0.48	0.89	–	–	0.78	–	0.52	–	0.14
Spin	0.65	0.66	0.58	0.10	0.46	0.55	0.91	0.56	0.51	0.78	0.43	0.51	0.56	0.15
Circumference	0.71	0.72	0.70	0.36	0.39	0.55	0.81	0.53	0.44	0.73	0.45	0.57	0.5	0.34
Depth	0.59	0.59	0.55	0.32	0.40	0.55	0.79	0.5	0.43	0.63	0.32	0.51	0.49	0.30
Girth	0.55	0.55	0.43	0.40	0.37	0.34	0.80	0.47	0.44	0.64	0.30	0.36	0.44	0.30
Dispersion	0.79	0.79	0.74	0.40	0.73	0.71	0.93	0.75	0.69	0.85	0.63	0.69	0.71	0.47
Range	0.57	0.60	0.60	0.32	0.57	0.58	0.80	0.59	0.54	0.65	0.47	0.50	0.51	0.28
Roughness	0.87	0.87	0.91	3.54	0.43	0.84	0.79	0.63	0.47	0.77	0.58	0.79	0.65	1.03
Form factor	1.11	1.04	2.04	8.87	2.75	2.55	1.04	1.55	0.83	0.98	0.49	0.58	1.84	4.65
							Elongations							
Horizontal elongation	0.64	0.62	0.16	0.12	0.00	0.12	0.01	0.02	0.14	0.35	0.05	0.31	0.42	0.54
Minimum vertical elongation	0.19	0.11	0.46	0.83	0.70	0.78	0.24	0.44	0.44	0.14	0.71	0.46	0.31	0.90
Maximum vertical elongation	0.56	0.57	0.36	0.81	0.69	0.75	0.24	0.42	0.52	0.24	0.72	0.63	0.16	0.78

very different values on vertical elongation (which, from Section Differences between 2D and 3D indices, seems to contribute to some extent).

We believe this highlights an important finding: 3D indices are more reliable than 2D for identifying the real shape of a building. If we were to observe just their footprints, we would not be able to know that these four buildings have such similar characteristics. Nevertheless, we can intuitively consider them quite similar, to a certain degree, by taking into account their 3D shape.

5.3. Comparison between LoDs

A comparison between different LoDs can be useful to determine the appropriate LoD for an application, for example, energy consumption. A higher level of detail does not mean a better level of detail and is instead use case dependent. Table 5 examines the Boolean difference between all LoD1 and LoD2 buildings as based on Figure 3. By examining Figure 7, we can also observe that there are differences in the distribution of the index values between LoD1 and LoD2. LoD1 tends to have more normal distributions for indices such as *equivalent cuboidness* and *convexity*, which makes sense
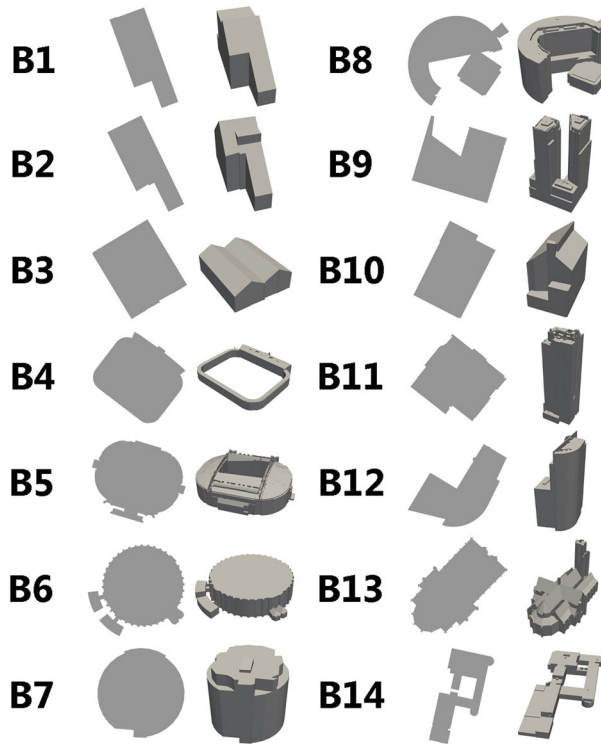
Figure 8. Metric values for buildings of different complexities for the 2 D footprint and the 3 D building shape. B1, B2, …, B14 are used to name the individual buildings from the sample we specifically selected as representative of the diversity of 3 D BAG.

given that LoD1 buildings are mainly simple block structures. Based on Table 5, it can be observed that the intersection values tend to be higher for LoD2 buildings, which might be due to the fact that the regular shapes of LoD1 models tend to be larger. Median values for differences between LoDs range between 8% and 14%, which will have a significant influence based on the application in question. While differences can be observed, and are expected, it is also important to note that these differences vary building by building, so the metrics can be utilised to check for cases where a higher level of detail does not add any additional benefits, and also cases where 2D or 2.5D is sufficient.

6. Use cases

6.1. Clustering

To verify the usefulness of 3D indices in identifying different building forms, we conducted a hierarchical clustering and evaluated the uniformity and distinctiveness of the resulting clusters on a random sample of 200,000 buildings (due to computational constraints we could not run the analysis on the full dataset). We ran an Agglomerative clustering algorithm[10] using 11 features created based on a principal
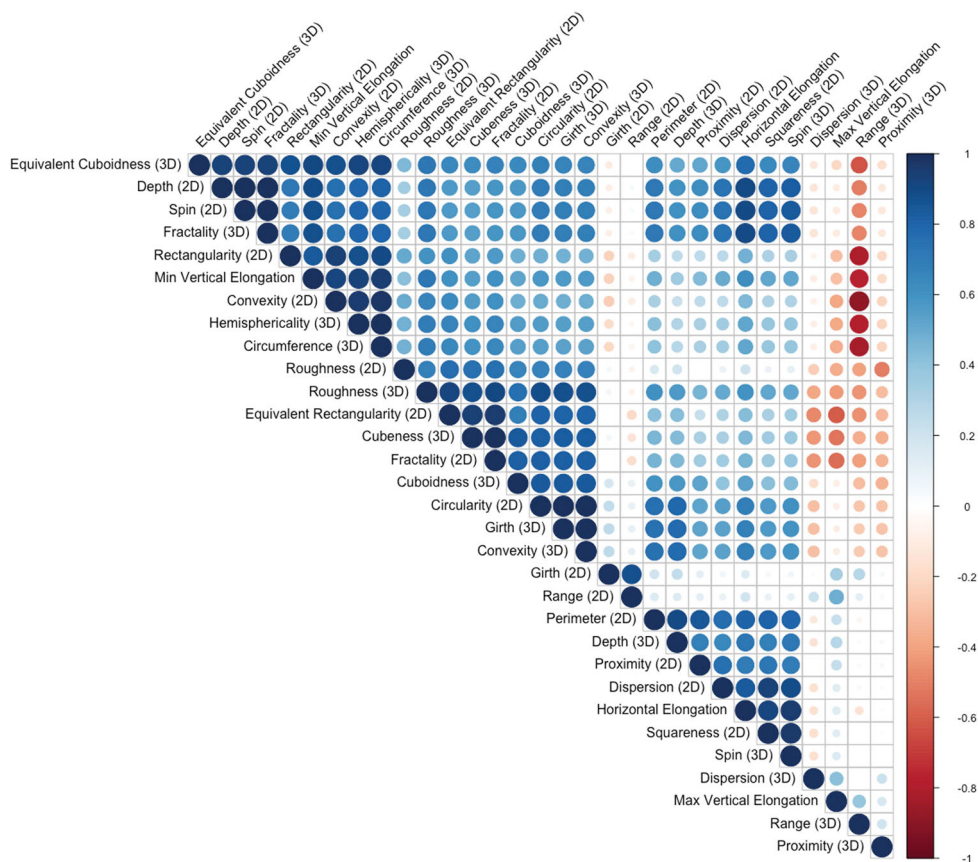
**Figure 9.** Correlation matrix between the indices of shape complexity.

component analysis (PCA) of the 3D indices. Furthermore, the PCA also demonstrated higher variability in the 3D indices than the 2D indices, see Figure 11. The clustering was conducted based on the average linkage method with a fixed number of 30 clusters.

The resulting clusters were of varying size, with three clusters dominating the vast majority of buildings (Figure 12). The rest of the clusters were relatively small (mostly ranging in size from 1 to 1000 buildings), with some clusters demonstrating very unique and distinct features. Figure 13 shows four such clusters which are particularly interesting. Figure 13(a) demonstrates a cluster of 166 buildings, all with "thin" and "wide" buildings (i.e. short buildings with large footprints). Figure 13(b) shows a cluster of 10 buildings with a distinct "L" shape or a perpendicular protrusion. Figure 13(c) shows a cluster of 58 buildings, which highlight "spikes" or long protrusions. Finally, Figure 13(d) shows a cluster of 322 buildings with mostly circular and cylindrical shapes.

What we established by observing the results of the clustering is that 3D indices seem to successfully identify similar shapes and group them together. Nevertheless, given that all three dimensions contribute equally to the calculation of an index, there is no guarantee that the clusters will identify "similar buildings" similar to how a
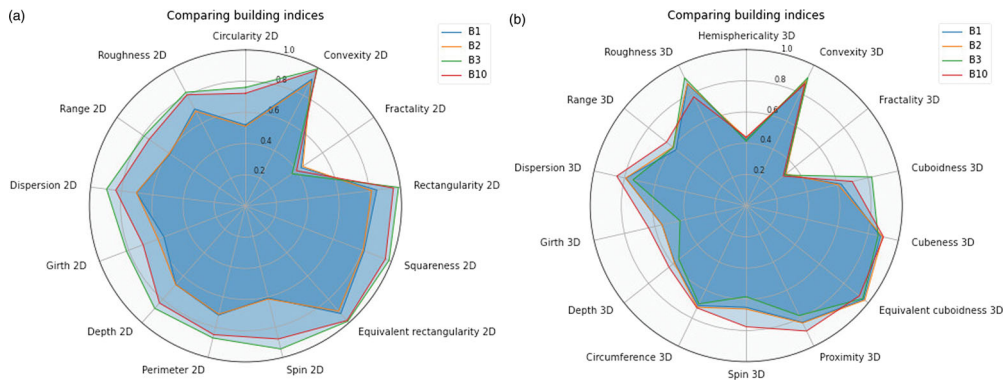
Figure 10. Comparison of indices between buildings B1, B2, B3, and B10 from Figure 8. (a) 2D indices. (b) 3D indices.

Table 5. Summary statistics for the Boolean intersections and differences between LoD 1 (Source) and LoD 2 (Destination) in terms of volume.

	Minimum	1st Quartile	Median	Mean	Third quartile	Maximum
			Amsterdam			
Source intersection	16%	88%	93%	92%	99%	100%
Destination intersection	20%	98%	100%	99%	100%	100%
Source difference	0%	2%	8%	9%	14%	405%
Destination difference	0%	2%	11%	13%	20%	235%
			The Hague			
Source intersection	30%	85%	91%	90%	99%	100%
Destination intersection	5%	98%	100%	99%	100%	100%
Source difference	0%	2%	10%	11%	17%	1817%
Destination difference	0%	2%	8%	11%	16%	531%
			Rotterdam			
Source intersection	24%	86%	92%	91%	98%	100%
Destination intersection	16%	98%	99%	99%	100%	100%
Source difference	0%	2%	10%	11%	16%	542%
Destination difference	0%	2%	10%	12%	19%	323%
			Utrecht			
Source intersection	31%	82%	89%	88%	95%	100%
Destination intersection	45%	97%	99%	98%	100%	100%
Source difference	0%	6%	13%	13%	20%	123%
Destination difference	0%	6%	14%	16%	24%	224%

person might group them; in other words, the indices do not differentiate between the horizontal and vertical profile. For example, in Figure 13(c), while most buildings show a vertical spike, there is one building whose protrusion is expanding horizontally. Clustering would be more meaningful when based on application-specific needs, in which case the clustering parameters would need to be adjusted accordingly.

6.2. Change of patterns over time

The second demonstration of a use case is understanding architectural patterns through time (Figure 14). The 3D BAG dataset contains the year of construction of each building, following which we aggregated buildings by decade and century. The left part of the figure illustrates the distribution of one of the 3D indicators (*cubeness*) in one of the cities (Den Haag) by century. There is a notable pattern change in both
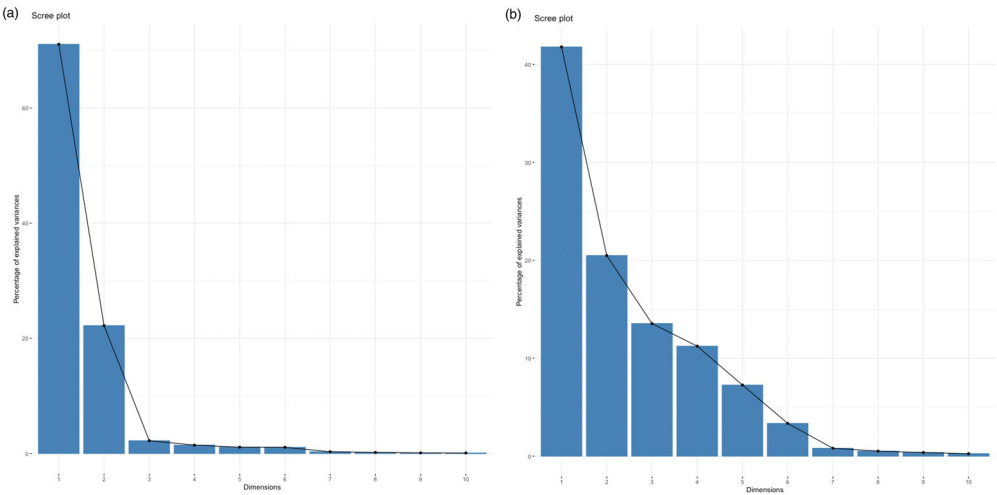
Figure 11. PCA comparison between 2 D and 3 D indices. (a) 2D indices. (b) 3D indices.
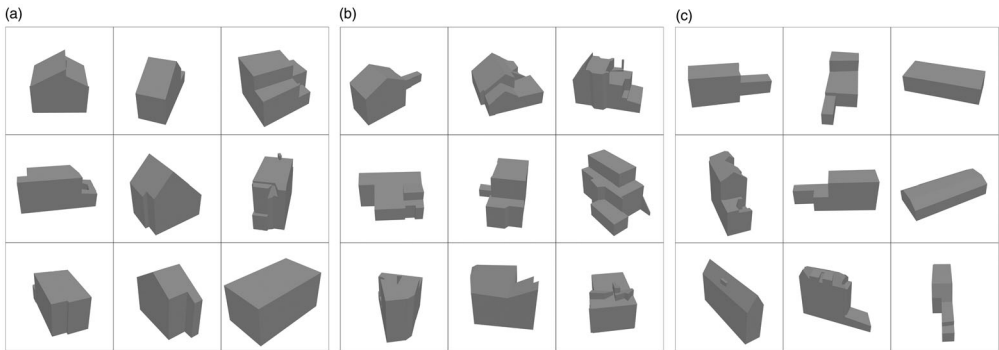
Figure 12. The three largest clusters of the hierarchical clustering analysis conducted on a sample of 200,000 buildings based on the 3 D indices. The clustering input was eleven features based on principal components that represented the original 3 D indices. We used the average linkage method to classify buildings in 30 clusters. These three clusters represent 86%, 7% and 4% of the sample, respectively. Clusters (a) and (b) are very similar, although on average objects in the latter seem to be slightly more complex. Cluster (c) portrays more elongated objects, mostly with protrusions.

the median value and dispersion of this morphological aspect over time. Next, the right part of the example shows the evolution of the relationship of two metrics (*dispersion* index together with the *convexity* of the footprint) in the last hundred years. There is a pattern in the increase of both metrics until the 1950s, after which the relationship takes a departure (distinguished in different colours for convenience). This example also doubles as a demonstration of an analysis that combines the 2D and 3D, and affirms our stance that the two families of metrics may be complementary.

7. Discussion

The availability of open 3D tools such as PyVista and PyMesh offers the ability to create software for studying 3D urban morphology. The 3D metrics we developed are, in
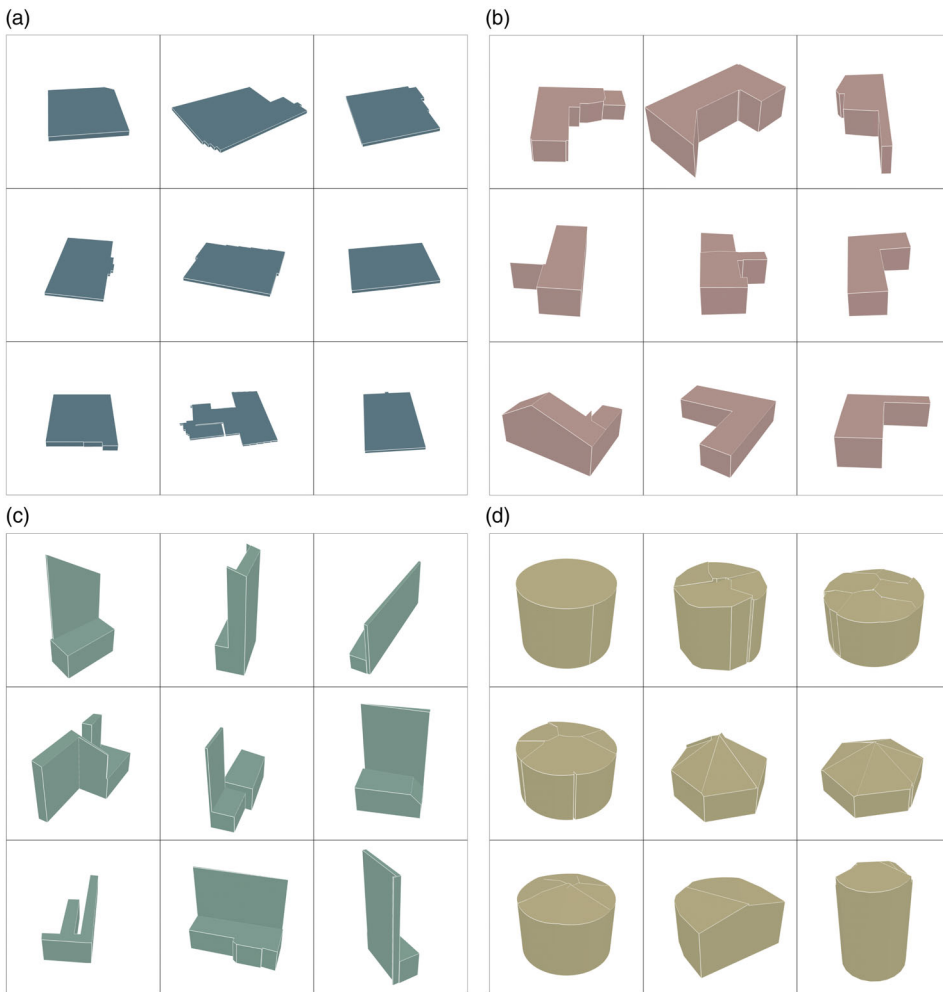
Figure 13. Examples of classes with distinct characteristics from the hierarchical clustering analysis based on the 3 D indices. The clustering input was eleven features based on principal components that represented the original 3 D indices. We used the average linkage method to classify buildings in 30 clusters (the number of clusters was selected after empirical experimentation with different parameters, with the goal to form meaningful clusters).

general, applicable to any type of 3D geometries and therefore can be used to conduct a large variety of urban analysis from any 3D dataset, e.g. bridges, bus stops, etc. At the same time, we identified an additional benefit from using 3D city models as our source data, given that they support the storage of semantic data, such as roof and wall surfaces, which allowed us to compute more properties that can be useful in a range of different analyses (i.e. vertex and surface counts and areas based on semantics).

When "3dfying" the originally 2D indices, we needed to slightly deviate from the original methodology in Basaraner and Cetinkaya (2017) in order to look for a more meaningful output in the real world. For example, instead of extending from *circularity* to *sphericality*, we chose to compute the *hemisphericality* as any resemblance of a building shape to a sphere would not be realistic. It remains an open question,
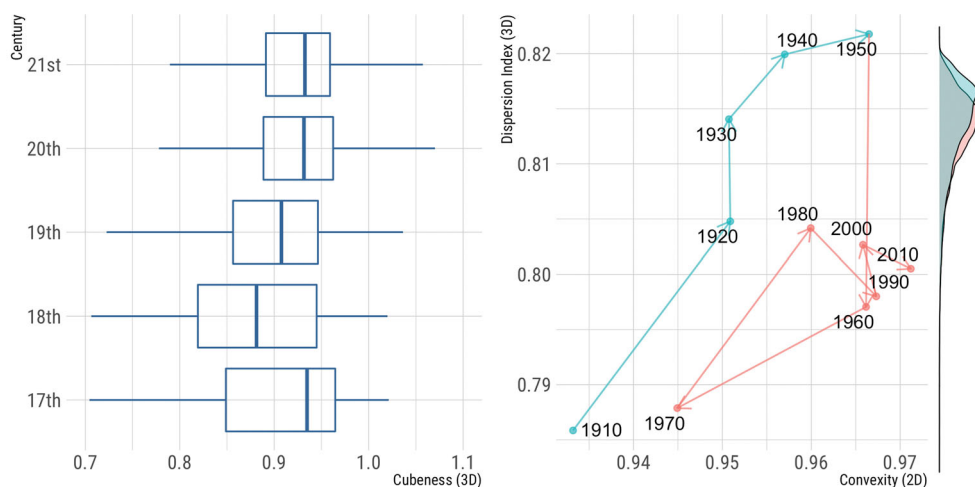
Figure 14. Associating the dataset with cadastral and other records enables a series of spatio-temporal analyses of urban morphology. In this case, the 3 D morphological patterns through time have been dissected by era.

though, if this needs to be further extended to other shapes as well. For example, one might want to experiment with the idea of using a cylinder or cone as a reference shape that can turn one index into many.

Some metrics/indices are more closely related to each other than others, e.g. *cubeness* and *cuboidness*. We do not know to which extent this is due to the peculiarities of the dataset used in this paper. Therefore, it would be interesting to compare the relationship between the indices in various geographic contexts. In addition, we cannot infer to which extent some metrics might be more important than others. Nevertheless, when running a comparison analysis it is important to be cognisant of the relationship between indices. Depending on the use case, PCA can be a useful tool for working with multiple indices at the same time. Furthermore, the 3D indices are not sensitive to the orientation of a polyhedron, therefore combining them with 2D measures can solve issues around the orientation of a building if it is important to a specific use case.

Intuitively, the purpose of metrics is to quantify the characteristics of geometries that a person can cognitively perceive when they observe them. Nevertheless, many times a person can decompose a geometry into more primitive components, which is something that our metrics are incapable of. For example, in Figure 8 building B8 would intuitively be considered to have around half the 2D *roughness* value of B10, the latter being a very smooth object as it can be perceived as a slightly rough half circle; interestingly, though, it has a very low *roughness* value of 0.29. Such a case can highlight how even though there is a primitive "normal" shape that contributes to the overall shape of geometry, the metrics cannot expose this. We expect this to be even more dominant in the 3D metrics.

One of the problems when computing 3D metrics is the increased processing time needed in order to compute them. This is because many of the indices are based on computing values for a number of grid points, which in 3D are exponentially more

than in 2D. In particular, we decided to exclude the calculation of the *cohesion* index, as its complexity was of $O(n^2)$ with respect to the number of grid points (n), which would render the computational needs excessive for our resources.

Some shape indices require a discretisation to be computed, using either interior points formed from a grid or surface grid points. As mentioned in Section Software and engineering decisions, for our implementation we chose to use voxels of size 0.5 m aligned to the axes. Based on empirical experiments we ran on sample buildings with different grid densities we concluded that the grid of 0.5 m cells is a good balance between accuracy and performance. This is because we observed that increasing the density to produce more cells did not change the values of the indices significantly. In addition, we believe that assuming a grid dense enough to provide an accurate value for the indices, would not be influenced enough by the axes over which it is built as its resolution is high enough to be representative of the shape despite the outcome. To ensure this, we also applied the post-processing steps described in Section Errors, so that only volumes with enough grid points are included in our analysis. Nevertheless, we do believe that further investigation on the subject can be conducted on the details of how the grid can be created and what its exact influence on the final indices is. This is especially true with respect to how the grid creation can be affected by the specific application for which the metrics are computed.

Another problematic aspect of working with 3D is the fundamental complexity of 3D geometric representation and its immediate complications. In this paper, we are working with boundary representation of 3D geometries, where a building is represented by its bounding polygons. This kind of representation, though, is very susceptible to validity issues, especially when it is used for representing solids (i.e. closed volumes). While this is similar in 2D, there are more comprehensive and robust methods to fix validity issues for 2D geometries than in 3D. Therefore, in many cases, we had to settle for checking the validity of 3D geometries and exclude those that are not valid, as described in Section Errors. In addition, some metrics are more sensitive to validity issues and water-tightness. This is especially true for the exchange index in 3D, which requires a Boolean operation to be computed; such operations can only be computed for valid and watertight objects, which is a requirement that is hard to meet in 3D. In our dataset, roughly 10% of volumes are not watertight, and therefore the exchange index could not be computed for those buildings.

A topic for further study is if the metrics are influenced by the modelling choices of the datasets and the reconstruction method to create the geometries. For example, it can be that a reconstruction method does not allow for consecutive surfaces of small angles between them and, therefore, this would have a big influence on metrics such as the roughness index. Therefore, we believe that one of the use cases for these metrics is to evaluate the influence of methodologies in an automatic reconstruction process. This would be beneficial when comparing different reconstruction methods in the same city, but at the same time, it also makes comparing cities with different reconstruction methods more difficult. On a similar topic, these metrics could be used to identify problematic unrealistic geometries, even if they are geometrically valid. Some of the examples from our clustering experiment (Section Clustering) demonstrate how the indices can be used to extract geometries with "spikes" or irregular shapes, in general.

Finally, the appropriate unit of analysis for the metrics is very important and heavily depends on the application for which the analysis will be conducted. This is mostly with respect to the subdivision of the urban space. For example, from an architectural perspective, it can be more useful to compute the metrics at the building level. On the other hand, for applications such as urban design, where a neighbourhood-level analysis is more interesting, the input should be dissolved so that adjacent buildings are merged together and treated as one geometry.

8. Conclusion

In this paper, we endeavoured to augment the multifaceted research line of urban morphology by bringing it into full 3D with a focus on buildings specifically. We put forward a robust set of metrics that is consistent with widely used existing lower-dimensional (2D) metrics, implement them in a software package that we release openly, generate a comprehensive dataset characterising more than 800,000 buildings that we share publicly, and explore use cases that may benefit from understanding the detailed three-dimensional urban fabric. In addition, by conducting a comprehensive data analysis and discussing a variety of aspects such as data, challenges, and engineering decisions, we believe that with this paper we set a solid foundation for this new research direction. Our work, besides building-level characterisation, supports urban scale and multi-city comparative analyses, as suggested by many examples in the paper (e.g. Figures 5, 7, 13 and 14).

The newly introduced metrics are not necessarily intended to replace the existing widely-used metrics. However, as demonstrated by this paper, they are beneficial in supplementing the metrics used so far as they reduce ambiguities and do more justice in parametrising the growing complexity of cities. We hope that our work will also be useful for the computer graphics community, which deals with 3D models of a diverse variety of features beyond buildings, to characterise 3D shape complexity.

Impediments to going fully 3D are the paucity of data and the complexity of implementing the 3D metrics. First, in comparison to building footprints, 3D city models are still not widely available around the world, and less so at higher levels of detail. However, recent developments in 3D reconstruction indicate a trend of the increase in both coverage and detail of 3D building models (Gui and Qin 2021, Zhang *et al*. 2021, Esch *et al*. 2022, Klimkowska *et al*. 2022, Pang and Biljecki 2022, Peters *et al*. 2022), potentially alleviating this issue. Our work establishes the value of data at various LoDs, and it demonstrates that even LoD1 (block) models, which are simple and increasingly available, provide benefits when they are used to compute advanced 3D metrics, rather than being used for only simple indices as in research so far. Second, urban morphology in 2D and 2.5D is nowadays fairly straightforward thanks to its established history, available datasets, defined metrics, and growing software support (Jhaldiyal *et al*. 2018, Fleischmann 2019, Jochem and Tatem 2021, Biljecki and Chow 2022, Yap *et al*. 2022), going 3D, for now, entails some complexity and requires capacity building, e.g. more complex formulas, intricate implementations, an increased likelihood of errors, and handling less established data formats.

For future work, we aspire to investigate the application of the work to other urban features, such as roads, vegetation, underground spaces, etc. in order to have a true 3D urban morphology analysis. Furthermore, the aggregation of different urban scales will be an important aspect of a true urban morphology analysis in our future work. Scalability is an important consideration as well. In this work, we focused on the Netherlands due to the rich availability of open data and the ability to compare different geographical areas that are reconstructed with the same methodology. We hope that our freely available software development will contribute towards computing the metrics in further geographies where data such as point clouds are available.

Finally, the set of 3D metrics we propose in this work may spur a variety of use cases, of which we suggest a few examples in the paper. Other possible use cases may be understanding the patterns of morphological indicators with various socio-economic aspects, e.g. average income by administrative neighbourhood (c.f. Figure 5), for which data may be available in some geographies.

Notes

1. https://www.businesslocationcenter.de/en/economic-atlas/download-portal
2. http://kartta.hel.fi/3d
3. http://www1.nyc.gov/site/doitt/initiatives/3d-building.page
4. Clustering is used to compensate for precision issues between similar planes. We use Agglomerative clustering, where we compute the distances between the vectors of plane parameters and we cluster them with an average linkage method using a threshold of 0.1. The distance matrix is computed so that triangles with opposite normals will still have a close distance and, therefore, end up in the same cluster.
5. Table 1 highlights which indices require such normalisation.
6. https://pymesh.readthedocs.io/
7. https://github.com/tudelft3d/3d-building-metrics
8. https://3dbag.nl/
9. http://geovalidation.bk.tudelft.nl/val3dity/
10. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html
11. https://3dbag.nl
12. https://bing.com/maps

Authors contributions

AL came up with the concept. AL and SV developed the metrics and implementation. AL, SV, and FB analysed the results and tested use cases. KAO and JS provided consultation and review. All authors contributed to writing and reviewing the paper.

Disclosure statement

Funding

Notes on contributors

Anna Labetski is a PhD candidate in the 3D geoinformation group at the Delft University of Technology. She holds an MSc in Geospatial Analysis from University College London.

Stelios Vitalis is a GIS software engineer, working as a PhD candidate at the 3D geoinformation group of TU Delft. He holds a degree in Rural and Surveying Engineering from NTU of Athens and an MSc in Informatics from the University of Piraeus.

Filip Biljecki is an assistant professor at the National University of Singapore and the principal investigator of the NUS Urban Analytics Lab. He holds an MSc in Geomatics and a PhD in 3D GIS from the Delft University of Technology.

Ken Arroyo Ohori is a lecturer and researcher at the 3D geoinformation group of the Delft University of Technology in the Netherlands. His research covers various aspects of geographic information, with a focus on the use of novel techniques to process data geometrically and topologically.

Prof dr Jantien Stoter is full professor and chairs the 3D Geoinformation research group at the Delft University of Technology. She did her PhD on 3D Cadastre (2004), and received personal grants both of the Dutch Science Foundation (2011–2016) and the European Research Council (2016-ongoing). She (co) leads several projects on GeoBIM, 3D city modelling and urban applications. She has fulfilled several roles in the Open Geospatial Consortium and the European Spatial Data Research association.

ORCID

Anna Labetski   http://orcid.org/0000-0003-2672-1220
Stelios Vitalis   http://orcid.org/0000-0003-1886-0722
Filip Biljecki   http://orcid.org/0000-0002-6229-7749
Ken Arroyo Ohori   http://orcid.org/0000-0002-9863-0152
Jantien Stoter   http://orcid.org/0000-0002-1393-7279

Data and codes availability statement

The source code of our implementation is available as open source on GitHub: https://github.com/tudelft3d/3d-building-metrics. The full dataset is available openly: https://doi.org/10.7910/DVN/6QCRRF

References

Basaraner, M., and 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 (10), 1952–1977.

Beran, D., *et al.*, 2021. The third dimension in noise visualization – a design of new methods for continuous phenomenon visualization. *The Cartographic Journal*, 1–17.

Biljecki, F., and Chow, Y.S., 2022. Global building morphology indicators. *Computers, Environment and Urban Systems*, 95, 101809.

Biljecki, F., *et al.*, 2021a. Open government geospatial data on buildings for planning sustainable and resilient cities. arXiv preprint arXiv:2107.04023.

Biljecki, F., *et al.*, 2021b. Extending CityGML for IFC-sourced 3D city models. *Automation in Construction*, 121, 103440.

Boeing, G., 2021. Street network models and indicators for every urban area in the world. *Geographical Analysis*.

Bourdic, L., Salat, S., and Nowacki, C., 2012. Assessing cities: a new system of cross-scale spatial indicators. *Building Research & Information*, 40 (5), 592–605.

Brovelli, M.A., and Zamboni, G., 2018. A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. *ISPRS International Journal of Geo-Information*, 7 (8), 289.

Byahut, S., Patel, B., and Mehta, J., 2020. Emergence of sub-optimal land utilization patterns in Indian cities. *Journal of Urban Design*, 25 (6), 758–777.

Cao, Q., *et al.*, 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, 107635.

Chen, S., *et al.*, 2021a. ATLAS: software for analysing the relationship between urban microclimate and urban morphology in a tropical city. *Building and Environment*, 208, 108591.

Chen, S., *et al.*, 2020. Combining CityGML files and data-driven models for microclimate simulations in a tropical city. *Building and Environment*, 185, 107314.

Chen, W., Wu, A.N., and Biljecki, F., 2021b. Classification of urban morphology with deep learning: application on urban vitality. *Computers, Environment and Urban Systems*, 90, 101706.

Coutrot, A., *et al.*, 2022. Entropy of city street networks linked to future spatial navigation ability. *Nature*, 604 (7904), 104–107.

Dukai, B., *et al.*, 2021. Quality assessment of a nationwide data set containing automatically reconstructed 3D building models. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-4/W4-2021, 17–24.

Dukai, B., *et al.*, 2020. Generating, storing, updating, and disseminating a country-wide 3D model. *International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, XLIV-4/W1-2020:27–32.

Esch, T., *et al.*, 2022. World settlement footprint 3D – a first three-dimensional survey of the global building stock. *Remote Sensing of Environment*, 270, 112877.

Fibaek, C.S., Keßler, C., and Arsanjani, J.J., 2021. A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning. *International Journal of Applied Earth Observation and Geoinformation*, 105, 102628.

Fleischmann, M., 2019. Momepy: urban morphology measuring toolkit. *Journal of Open Source Software*, 4 (43), 1807.

Fleischmann, M., *et al.*, 2021. Methodological foundation of a numerical taxonomy of urban form. *Environment and Planning B: Urban Analytics and City Science*, 49 (4), 1283–1299.

Geis, C., *et al.*, 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 (8), 2912–2927.

Gil, J., 2020. City information modelling: a conceptual framework for research and practice in digital urban planning. *Built Environment*, 46 (4), 501–527.

Gil, J., *et al.*, 2009. On the discovery of urban typologies: data mining the multi-dimensional character of neighbourhoods. In: *27th eCAADe conference*. Istanbul, Turkey: eCAADe (Education and research in Computer Aided Architectural Design in Europe), 269–278.

Gui, S., and Qin, R., 2021. Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 1–19.

Heris, M.P., *et al.*, 2020. A rasterized building footprint dataset for the United States. *Scientific Data*, 7 (1), 207.

Hu, Y., Dai, Z., and 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, 110424.

Huang, X., and Wang, C., 2020. Estimates of exposure to the 100-year floods in the conterminous United States using national building footprints. *International Journal of Disaster Risk Reduction*, 50 (101731), 101731.

Huang, X., and Wang, Y., 2019. Investigating the effects of 3D urban morphology on the surface urban heat island effect in urban functional zones by using high-resolution remote sensing data: a case study of Wuhan, Central China. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 119–131.

Ito, K., and Biljecki, F., 2021. Assessing bikeability with street view imagery and computer vision. *Transportation Research Part C: Emerging Technologies*, 132, 103371.

Jhaldiyal, A., *et al.*, 2018. Urban morphology extractor: a spatial tool for characterizing urban morphology. *Urban Climate*, 24, 237–246.

Jochem, W.C., and Tatem, A.J., 2021. Tools for mapping multi-scale settlement patterns of building footprints: an introduction to the R package foot. *PLOS One*, 16 (2), e0247535.

Klimkowska, A., *et al.*, 2022. Detailed three-dimensional building façade reconstruction: a review on applications, data and technologies. *Remote Sensing*, 14 (11), 2579.

Ledoux, H., 2018. val3dity: validation of 3D GIS primitives according to the international standards. *Open Geospatial Data, Software and Standards*, 3 (1), 1.

Ledoux, H., *et al.*, 2019. CityJSON: a compact and easy-to-use encoding of the CityGML data model. *Open Geospatial Data, Software and Standards*, 4 (1), 4.

Leng, H., *et al.*, 2020. Urban morphology and building heating energy consumption: evidence from Harbin, a severe cold region city. *Energy and Buildings*, 224, 110143.

Li, H., *et al.*, 2021. Quantifying 3D building form effects on urban land surface temperature and modeling seasonal correlation patterns. *Building and Environment*, 204, 108132.

Li, M., *et al.*, 2020. Continental-scale mapping and analysis of 3D building structure. *Remote Sensing of Environment*, 245, 111859.

Lindenthal, T., 2020. Beauty in the eye of the home-owner: aesthetic zoning and residential property values. *Real Estate Economics*, 48 (2), 530–555.

Liu, Y., *et al.*, 2020. Characterizing three dimensional (3-d) morphology of residential buildings by landscape metrics. *Landscape Ecology*, 35 (11), 2587–2599.

Liu, Y., *et al.*, 2021. Complexity of relationship between 2d/3d urban morphology and land surface temperature: a multi-scale perspective. *Environmental Science and Pollution Research International*, 28 (47), 66804–66818.

Lu, H., *et al.*, 2021. Multi-scale impacts of 2D/3D urban building pattern in intra-annual thermal environment of Hangzhou. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102558.

Luan, H., and Fuller, D., 2021. Urban form in Canada at a small-area level: quantifying "compactness" and "sprawl" with Bayesian multi variate spatial factor analysis. *Environment and Planning B: Urban Analytics and City Science*, 49 (4), 1300–1313.

Lucks, L., *et al.*, 2021. Improving trajectory estimation using 3D city models and kinematic point clouds. *Transactions in GIS*, 25 (1), 238–260.

Malhotra, A., *et al.*, 2022. Information modelling for urban building energy simulation—a taxonomic review. *Building and Environment*, 208, 108552.

Mei, S.-J., and Yuan, C., 2021. Analytical and numerical study on transient urban street air warming induced by anthropogenic heat emission. *Energy and Buildings*, 231, 110613.

Milojevic-Dupont, N., *et al.*, 2020. Learning from urban form to predict building heights. *PLOS One*, 15 (12), e0242010.

Noardo, F., *et al.*, 2021. Reference study of CityGML software support: the GeoBIM benchmark 2019—Part II. *Transactions in GIS*, 25 (2), 842–868.

Nys, G., and Billen, R., 2021. From consistency to flexibility: a simplified database schema for the management of CityJSON 3D city models. *Transactions in GIS*, 25 (6), 3048–3066.

Palliwal, A., *et al.*, 2021. 3D city models for urban farming site identification in buildings. *Computers, Environment and Urban Systems*, 86, 101584.

Pang, H.E., and Biljecki, F., 2022. 3D building reconstruction from single street view images using deep learning. *International Journal of Applied Earth Observation and Geoinformation*, 112, 102859.

Park, Y., Guldmann, J.-M., and Liu, D., 2021. Impacts of tree and building shades on the urban heat island: combining remote sensing, 3D digital city and spatial regression approaches. *Computers, Environment and Urban Systems*, 88, 101655.

Peters, R., *et al.*, 2022. Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands. *Photogrammetric Engineering & Remote Sensing*, 88 (3), 165–170.

Ren, C., *et al.*, 2020. Developing a rapid method for 3-dimensional urban morphology extraction using open-source data. *Sustainable Cities and Society*, 53, 101962.

Santhanavanich, T., and Coors, V., 2021. CityThings: an integration of the dynamic sensor data to the 3D city model. *Environment and Planning B: Urban Analytics and City Science*, 48 (3), 417–432.

Santos, T., *et al.*, 2021. Assessing sustainable urban development trends in a dynamic tourist coastal area using 3D spatial indicators. *Energies*, 14 (16), 5044.

Sarretta, A., and Minghini, M., 2021. Towards the integration of authoritative and OpenStreetMap geospatial datasets in support of the European strategy for data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-4/W2-2021, 159–166.

Shirowzhan, S., *et al.*, 2020. Data mining for recognition of spatial distribution patterns of building heights using airborne lidar data. *Advanced Engineering Informatics*, 43, 101033.

Sirko, W., *et al.*, 2021. Continental-scale building detection from high resolution satellite imagery. arXiv preprint arXiv:2107.12283.

Sullivan, C.B., and Kaszynski, A., 2019. PyVista: 3D plotting and mesh analysis through a streamlined interface for the visualization toolkit (VTK). *Journal of Open Source Software*, 4 (37), 1450.

Sun, Y., *et al.*, 2020. Automatic registration of a single SAR image and GIS building footprints in a large-scale urban area. *ISPRS Journal of Photogrammetry and Remote Sensing*, 170, 1–14.

Tian, Y., and Yao, X., 2021. Urban form, traffic volume, and air quality: a spatiotemporal stratified approach. *Environment and Planning B: Urban Analytics and City Science*, 49 (1), 92–113.

Tian, Y., *et al.*, 2019. The effect of urban 2D and 3D morphology on air temperature in residential neighborhoods. *Landscape Ecology*, 34 (5), 1161–1178.

Tong, H., and Kang, J., 2021. Characteristics of noise complaints and the associations with urban morphology: a comparison across densities. *Environmental Research*, 197, 111045.

Touchaei, A., and Wang, Y., 2015. Characterizing urban heat island in Montreal (Canada)—effect of urban morphology. *Sustainable Cities and Society*, 19, 395–402.

Virtanen, J.-P., *et al.*, 2021. Near real-time semantic view analysis of 3D city models in web browser. *ISPRS International Journal of Geo-Information*, 10 (3), 138.

Vitalis, S., Arroyo Ohori, K., and Stoter, J., 2020. CityJSON in QGIS: development of an open-source plugin. *Transactions in GIS*, 24 (5), 1147–1164.):

Wang, M., and Debbage, N., 2021. Urban morphology and traffic congestion: longitudinal evidence from US cities. *Computers, Environment and Urban Systems*, 89, 101676.

Wentz, E.A., 2010. A shape definition for geographic applications based on edge, elongation, and perforation. *Geographical Analysis*, 32 (2), 95–112.

Wu, A.N., and Biljecki, F., 2022. GANmapper: geographical data translation. *International Journal of Geographical Information Science*, 36 (7), 1394–1422.

Wu, J., *et al.*, 2022. Cultivating historical heritage area vitality using urban morphology approach based on big data and machine learning. *Computers, Environment and Urban Systems*, 91, 101716.

Wysocki, O., *et al.*, 2021. Plastic surgery for 3D city models: a pipeline for automatic geometry refinement and semantic enrichment. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-4-2021, 17–24.

Xie, Y., *et al.*, 2019. A locally-constrained YOLO framework for detecting small and densely-distributed building footprints. *International Journal of Geographical Information Science*, 34 (4), 1–25.

Xu, Y., *et al.*, 2017. Urban morphology detection and computation for urban climate research. *Landscape and Urban Planning*, 167, 212–224.

Yang, L., *et al.*, 2022. Urban morphological regionalization based on 3d building blocks-a case in the central area of Chengdu, China. *Computers, Environment and Urban Systems*, 94, 101800.

Yap, W., Janssen, P., and Biljecki, F., 2022. Free and open source urbanism: software for urban planning practice. *Computers, Environment and Urban Planning*, 96(W1-2020):101825.

Zhang, C., Fan, H., and Kong, G., 2021. VGI3D: an interactive and low-cost solution for 3D building modelling from street-level VGI images. *Journal of Geovisualization and Spatial Analysis*, 5 (2), 18.

Zhang, J., *et al.*, 2019. Impact of urban block typology on building solar potential and energy use efficiency in tropical high-density city. *Applied Energy*, 240, 513–533.

Zhang, Z., *et al.*, 2022. Vectorized rooftop area data for 90 cities in China. *Scientific Data*, 9 (1), 66.

Zhu, X.X., *et al.*, 2022. The urban morphology on our planet – global perspectives from space. *Remote Sensing of Environment*, 269, 112794.

Appendix

The following notations are used in 2D:

- A_x = Area of *x*
- P_x = Perimeter of *x*
- PN = Polygon
- CH = Convex Hull of the polygon
- EAC = Equal Area Circle of the polygon
- MABR = Minimum Area Bounding Rectangle of the polygon
- PN ∩ EAC = Intersection of a polygon and its Equal Area Circle
- LIC = Largest Inscribed Circle of the polygon (where r_{LIC} is its radius)
- ADC = Average Distance Circle of the polygon (where r_{ADC} is its radius)
- SCC = Smallest Circumscribing Circle of the polygon (where r_{SCC} is its radius)
- $d_{igp_{ij}}$ = Distance between two internal grid points, *i* and *j*
- $\mu_{r_{igp}}$ = Average distance of internal grid points to the center
- $\mu_{r_{ibp}}$ = Average distance of surface points to the center
- $\mu_{r_{dev}}$ = Average distance between surface points and the Average Distance Circle along its radius

The following notations are used in 3D:

- V_x = Volume of *x*
- A_x = Total surface area of *x*
- MS = Mesh

- CH = Convex Hull of the mesh
- EVS = Equal Volume Sphere of the mesh
- OOBB = Object-Oriented Bounding Box of the mesh
- MS ∩ EVS = Intersection of a mesh and its Equal volume Sphere
- LIS = Largest Inscribed Sphere of the mesh (where r_{LIS} is its radius)
- ADS = Average Distance Sphere of the mesh (where r_{ADS} is its radius)
- SCS = Smallest Circumscribing Sphere of the mesh (where r_{SCS} is its radius)
- $d_{igp_{ij}}$ = Distance between two internal grid points, i and j
- $\mu_{r_{igp}}$ = Average distance of internal grid points to the center
- $\mu_{r_{ibp}}$ = Average distance of surface points to the center
- $\mu_{r_{dev}}$ = Average distance between surface points and the Average Distance Sphere along its radius

Appendix Table 1. Calculating shape indices in 2D and 3D.

Index	2D	3D
Circularity/Hemisphericality	$CI = \frac{A_{PN}}{A_{EPC}} = \frac{4\pi A_{PN}}{P_{PN}^2}$	$HEM = \frac{3\sqrt{2\pi V}}{A_{MS}^{\frac{3}{2}}}$
Convexity	$CNV_2 = \frac{A_{PN}}{A_{CH}}$	$CNV_3 = \frac{V_{MS}}{V_{CH}}$
Fractality	$FR_2 = 1 - \frac{\log(A_{PN})}{2 \times \log(P_{PN})}$	$FR_3 = 1 - \frac{\log(V_{MS})}{\frac{3}{2} \times \log(A_{MS})}$
Rectangularity/Cuboidness	$REC = \frac{A_{PN}}{A_{MABR}}$	$CBD = \frac{V_{MS}}{V_{OOBB}}$
Squareness/Cubeness	$SQN = \frac{P_{EAC}}{P_{PN}} = \frac{4\sqrt{A_{PN}}}{P_{PN}}$	$CBN = \frac{A_{EVS}}{A_{MS}} = \frac{6\sqrt[3]{V_{MS}^2}}{A_{MS}}$
Cohesion	$nCI_2 = \frac{0.9054 \times \sqrt{\frac{A_{PN}}{\pi}}}{\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n d_{igp_{ij}}}$	$nCI_3 = \frac{\frac{36}{35} \times \sqrt[3]{\frac{3 V_{MS}}{4\pi}}}{\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n d_{igp_{ij}}}$
Proximity	$nPxI_2 = \frac{\frac{2}{3} \times r_{EAC}}{\mu_{r_{igp}}} = \frac{\frac{2}{3} \times \sqrt{\frac{A_{PN}}{\pi}}}{\mu_{r_{igp}}}$	$nPxI_3 = \frac{\frac{3}{4} \times r_{EVS}}{\mu_{r_{igp}}} = \frac{\frac{3}{4} \times \sqrt[3]{\frac{3 V_{MS}}{4\pi}}}{\mu_{r_{igp}}}$
Exchange	$nEI_2 = \frac{A_{(PN \cap EAC)}}{A_{PN}}$	$nEI_3 = \frac{V_{(MS \cap EVS)}}{V_{MS}}$
Spin	$nSI_2 = \frac{0.5 \times \frac{A_{PN}}{\pi}}{\mu_{r_{igp}}^2}$	$nSI_3 = \frac{\frac{3}{5} \times \sqrt[3]{\frac{3 \times V_{MS}}{4 \times \pi}}^2}{\mu_{r_{igp}}^2}$
Perimeter/Circumference	$nPmI_2 = \frac{P_{EAC}}{P_{PN}} = \frac{2\sqrt{\pi A_{PN}}}{P_{PN}}$	$nPmI_3 = \frac{4\pi \times \sqrt[3]{\frac{3 \times V_{MS}}{4 \times \pi}}^2}{A_{MS}}$
Depth	$nDpI_2 = \frac{\mu_{PN_{d_{igp,b}}}}{\mu_{EAC_{d_{igp,b}}}} = \frac{3\mu_{PN_{d_{igp,b}}}}{\sqrt{\frac{A_{PN}}{\pi}}}$	$nDpI_3 = \frac{4 \times \mu_{MS_{d_{igp,b}}}}{\sqrt[3]{\frac{3 \times V_{MS}}{4 \times \pi}}}$
Girth	$nGI_2 = \frac{r_{LIC}}{r_{EAC}} = \frac{r_{LIC}}{\sqrt{\frac{A_{PN}}{\pi}}}$	$nGI_3 = \frac{r_{LIS}}{\sqrt[3]{\frac{3 \times V_{MS}}{4 \times \pi}}}$
Dispersion	$nDsI_2 = 1 - \frac{\mu_{r_{dev}}}{r_{ADC}} = 1 - \frac{\mu_{r_{dev}}}{\mu_{r_{ibp}}}$	$nDsI_3 = 1 - \frac{\mu_{r_{dev}}}{r_{ADS}} = 1 - \frac{\mu_{r_{dev}}}{\mu_{r_{ibp}}}$
Range	$nRI_2 = \frac{r_{EAC}}{r_{SCC}} = \frac{\sqrt{\frac{A_{PN}}{\pi}}}{r_{SCC}}$	$nRI_3 = \frac{\sqrt[3]{\frac{3 \times V_{MS}}{4 \times \pi}}}{r_{SCS}}$
Equivalent rectangular/ cuboid index	(1) $k = \sqrt{\frac{A_{PN}}{A_{MABR}}} \quad (k \leq 1)$ (2) $P_{EAR} = k \times P_{MABR}$ (3) $ERI = \frac{P_{EAR}}{P_{PN}} = \sqrt{\frac{A_{PN}}{A_{MABR}}} \times \frac{P_{MABR}}{P_{PN}}$	$ECI = \sqrt[3]{\frac{V_{MS}}{V_{OOBB}}}^2 \times \frac{A_{OOBB}}{A_{MS}}$
Roughness index*	$RI_2 = \frac{\mu_{r_{ibp}}^2}{A_{PN} + P_{PN}^2} \times 42.62$	$RI_3 = \frac{\mu_{r_{ibp}}^3}{V_{MS} + \sqrt{A_{MS}}^3} \times 48.735$
Elongation	$1 - \frac{S}{L}$ (Computed for the three axes respectively)	
Form factor	–	$\frac{A}{V^{\frac{2}{3}}}$

*For *roughness index* the constants 42.62 and 48.735 are used because without them the resulting values for a circle or a sphere would be $RI_2 = 1/42.62$ and $RI_3 = 1/48.735$, respectively.