Building and Environment 237 (2023) 110295

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes

Filip Biljecki [a,b,*], Yoong Shin Chow [a], Kay Lee [c]

[a] Department of Architecture, National University of Singapore, Singapore
[b] Department of Real Estate, National University of Singapore, Singapore
[c] Yale-NUS College, Singapore

ARTICLE INFO

Keywords:
User-generated content
3D city models
Digital twins
Standards
Volunteered Geographic Information

ABSTRACT

Geospatial data of the building stock is essential in many domains pertaining to the built environment. These datasets are often provided by governments, but crowdsourcing them has surged in the last decade. Nowadays, OpenStreetMap (OSM) – the most popular Volunteered Geographic Information (VGI) platform – contains geospatial and descriptive data on more than 500 million buildings worldwide collected by millions of contributors, and it is increasingly used in studies ranging from energy and microclimate to urban planning and life cycle assessment. However, large-scale understanding on their quality remains limited, which may hinder their use and management. In this paper, we seek to understand the state of building information in OSM and whether it is a reliable source of such data. We provide a comprehensive study to assess the quality of attribute (descriptive) data of the building stock mapped globally, e.g. building function, which are key ingredients in many analyses and simulations in the built environment. We examine three aspects: completeness, consistency, and accuracy. In this assessment, the first at such scale and the most comprehensive available hitherto, we find that quality continues to be highly heterogeneous — from poor quality in some, to very high completeness in other areas, potentially benefiting a range of application domains, e.g. we estimate that 3D building models of 443 administrative units (mostly cities and municipalities) around the world can be generated from OSM, underpinning the generation of digital twins. The number of floors and building type are the most frequent properties that contributors record, and in most cases are highly accurate, while mapping the interior of buildings did not gain momentum.

1. Introduction

Data on buildings play an important role in a wide range of domains, from energy and climate to cadastre and urban studies, and at different scales, from the architectural and precinct to national and continental scales [1–4]. Great strides have been made in methods to acquire information on buildings in the past decade [5–9], and data on the building stock is increasing both in volume and content. For example, data at the urban scale is gradually more geometrically detailed and it now routinely includes descriptive information such as type and number of storeys of buildings [10–12].

Another key development is the multiplication of stakeholders and their types — nowadays, many governments, companies, academia, and volunteers collect, maintain, and release building data openly [13–18]. Among these, crowdsourcing has gained particular attention in the past decade, especially OpenStreetMap[1] (OSM), the free editable map of the world and the leading instance of Volunteered Geographic Information (VGI) [19], which spans a variety of other types of data such as social media (e.g. Twitter, Flickr, Weibo) and street view imagery (e.g. Mapillary, KartaView).

OSM allows mapping and describing any real-world feature, from administrative areas and topographic features to amenities and street furniture, and buildings have emerged as a prominent one, reflecting their importance in the built environment [20]. Building data from this source, which can be mapped at different scales and detail and may contain a rich set of attributes describing the individual building stock, has been welcomed by the built environment research community thanks to the increasing coverage, quality, open licence, and uniqueness, as OSM remains the only such building data source worldwide. For example, building data available in OSM has been used for numerous studies in the built environment, e.g. on vulnerability and damage assessment [21–24], energy modelling and thermal simulations [25–30], microclimate studies [31–34], water and waste management [35,

* Corresponding author at: Department of Architecture, National University of Singapore, Singapore.
E-mail address: filip@nus.edu.sg (F. Biljecki).

[1] https://www.openstreetmap.org/

https://doi.org/10.1016/j.buildenv.2023.110295
Received 7 October 2022; Received in revised form 3 April 2023; Accepted 6 April 2023
Available online 18 April 2023

36], rooftop utilisation analyses [37], socio-economic studies [38], landscape perception [39], carbon emissions [40], mapping urban function [41], urban farming [42], urban morphology [43,44], and evacuation management [45]. In particular, building attributes such as the type and height, feed many such studies [46–48].

While OSM has proven its worth in numerous studies in the community, the heterogeneous provenance of the data has called attention to understanding their quality and it has increased scrutiny on the content of the data [49]. Therefore, many studies have been conducted on assessing the quality of a particular subset of OSM data, such as streets, amenities, and buildings, and they cover a variety of data quality aspects such as whether all amenities in an area have been mapped and how closely the mapped geometry of a building resembles the one in reality. The majority of state of the art relies on *authoritative* geospatial data (maintained and made available by government bodies), which is used as reference that is compared to the OSM data, from which the quality of OSM is gauged. These datasets are often released openly by authorities, and thus the studies are limited to jurisdictions where they are available and tend to focus on a finite geographical coverage — at the city level or at the national level.

Quality assessment studies have used a variety of approaches and focus on particular spatial data quality elements, from assessing the completeness of data (e.g. the percentage of amenities in a city that are mapped) to the correctness of descriptive data (e.g. whether the recorded type of an amenity is correct or not). Buildings have gained much attention in quality assessment studies, however, the descriptive content of building information has not been much in focus, and especially notably missing is large-scale research and an overarching study on this topic. That is, it remains undocumented what is the global content of the data and whether the buildings that are mapped (more than half billion of them at the time of the writing of this paper) are associated with accurate information that can be relied upon for such studies.

In this paper, following the growing coverage and use of OSM, we bridge the evident gap in state of the art and present a comprehensive global study on understanding the content and the quality of attributes of buildings in OSM. In our research, we focus on three aspects that are key to understanding the fit for purpose and quality, and are instrumental in a variety of studies — completeness, consistency, and accuracy of attributes. These represent measures of quantity and correctness of the data content. Conducting this research, we introduce insights that may be valuable also for studies that deal at smaller scales, such as at the country level. Further, besides outlining a series of statistics, we seek to explain their patterns and interpret the results, and provide meaning to them, e.g. implications for use cases. In analysing each of these quality aspects, we cut across multiple dimensions, such as geographical area, to understand spatial, socio-economic, and other patterns of building data quality worldwide. It is important to note that data quality, especially in the geospatial context, is often mixed up with resolution, which is not in the focus of our paper.

This study is designed to be of interest to a broad variety of disciplines pertaining to the built environment, with different applications of building-related geographic and descriptive information at the urban scale. Also, the discourse on building data quality has been overlooked and has not followed the proliferation of the use of such data in various domains in the built environment, thus, our paper brings attention to the topic and gives concrete insights and recommendations to the research community. Further, we believe that it is also relevant in the context of the growing interest in crowdsourced geospatial data in the sustainable development and smart city communities and for governance [50–54]. Our paper also expounds the basics of OSM building information to give an understanding of the platform and the data, potentially further raising awareness of this growing data source that is increasingly used in different application domains, and accompany its further adoption. Besides focusing on a general, broad, and agnostic analysis applicable to many use cases, we give particular attention to one — availability of the information on the building height, which is of key importance to several use cases and generating 3D building models that can be used for construction digital twins and conducting simulations [55], and for which OSM has been used in scores of papers [32,56–61]. Finally, we share observations and lessons learned that may serve as input to researchers and volunteers, and may lead to the improvement of both mapping in OSM and the use of data.

2. Background and related work

2.1. OpenStreetMap

As a prominent open and collaborative project [19], OpenStreetMap (OSM) is a volunteer-contributed worldwide geospatial database that offers the capability to model any urban feature spatially and descriptively. It was established in 2004 to counter the predominance of proprietary map data [62], and since then millions of contributors have mapped billions of features based on field surveys, aerial and satellite imagery, and using local knowledge of the area. Mappers, who are driven by a variety of motivations and interests, are from all over the world, from a variety of backgrounds and demographics, and can be local residents, visitors, or remote contributors, e.g. tourists visiting an area and even those who have never been in the location they map (using satellite imagery to help mapping a remote area) [63–65]. Besides manual mapping by contributors, a portion of the data has been adopted from authoritative sources where the data licence allows so, where it does not conflict with existing content in OSM, and where the quality is ensured [66]. As a result, its data quality is quite heterogeneous in multiple aspects [67].

As its name suggests, the project initially focused on roads, but it also emerged as an important data source of features such as points of interest (e.g. restaurants, pharmacies, schools, parking lots, religious objects), walkways, public transportation routes, administrative zones, public open spaces (e.g. parks, sport fields), natural topographic features such as waterbodies and mountain peaks, and — buildings.

In general, OpenStreetMap is both a geospatial database (the data can be downloaded and used in a variety of software packages) and a web-based map (the data is rendered and available to everyone for viewing, and editing is allowed), and its advantages are numerous, even when authoritative or other data is already available in the same area [12,68]. Most importantly, it is free and released under a liberal licence, globally present and harmonised, based on local knowledge, supported by corporations, enables historical versioning, and may be updated more often than (potentially outdated) government data, and may reveal additional attributes not available elsewhere [12,69,70]. Moreover, in some regions, OSM is often the only open data source available, especially across the Global South, and in areas where authorities and companies collect the data but are not keen on releasing them openly or provide them to researchers. Finally, another advantage of OSM is that it may contain informal settlements and slums, which may not be included in authoritative data [71,72].

A disadvantage is its varying level of quality, as – in contrast to authoritative sources – mapping efforts are scattered and voluntary (e.g. all amenities in a neighbourhood may be mapped by a diligent and enthusiastic contributor living there, but not in another one in the same city). Nevertheless, the data has in general been embraced by practitioners, governments, and the research community, and numerous studies across many fields have benefited from the semantic and physical representation of real-world features [19]. Buildings are among them, and they will be given a detailed introduction in Section 3 with an exploratory data analysis to give a better background for this study.

2.2. Geospatial data quality and assessment of OSM

Spatial data quality assessment has been of importance for decades and a prominent research pillar in the geospatial domain, with developments traversing many stakeholders, types of data, and sources [73–77], and having a direct influence on the usability. It has been formalised in the standard ISO 19157, which gives guidelines on a number of spatial data quality elements such as logical consistency, lineage, completeness, positional accuracy, and attribute accuracy [78,79].

As part of such developments, and spurred by the diverse means to acquire data, assessing the quality of OSM data has been a prominent topic in conferences and journals [80–86]. The quality of a large variety of categories of features around the world has been analysed, from shops and museums to toponyms and land use [87–90], and buildings have often been in focus of such studies [68,91–94]. Studies cover most spatial data quality elements [80], including the quality of attributes, such as whether the speed limits of roads are available and accurate [78,95].

Much of OSM quality studies are straightforward, and mostly rely on (often more trustworthy) authoritative (government) data that they employ for comparison [96,97] or manual work [98], so they are not scalable [99–101]. Alternatively, some studies use other sources such as satellite imagery. Thanks to the increasing availability and quality of such data and computational resources, there have been recent studies that have slightly expanded the typical study size and cover a few countries [102,103]. Besides such extrinsic studies that are usually used to gauge aspects such as completeness and accuracy, there are intrinsic counterparts [104–107]. These do not rely on an external source of data to gauge the quality, but focus on aspects for which reference data is not required, e.g. spatio-temporal analysis of contributions and understanding data consistency (e.g. verify the content of data against the mapping guidelines such as adherence to a set of determined values such as building types).

Most quality assessment studies pertaining to buildings focus on estimating their completeness (i.e. understanding the percentage of buildings that have been mapped in an area), and there has been little interest in the quality of building attributes [108,109]. In fact, no study has been conducted at the global scale, and one that is focusing on several attributes. While quality of the attributes is just one of several spatial data quality elements, and while buildings are just one type of features that are mapped, considering the importance of buildings and a large number of studies that rely on building information sourced from OSM, we deem that this is an omission that requires attention and comprehensive research.

The topic of building data quality is important not only because the data is seeing an increase in use (much of which is without an understanding of the quality), but also because at the global scale the practices may differ drastically. Further, while users have to map the location of a building, they are not obliged to provide also their attributes, e.g. a contributor may map the footprint (2D shape) of a building but skip entering any building characteristics such as address and year of construction.

Existing studies on OSM building attribute analyses tend to be similar as most OSM data quality assessment studies, e.g. they have a bounded and small geographical focus, often following administrative lines. For example, the paper of Goetz and Zipf [110] analyses the completeness of building attributes in Germany, e.g. percentage of buildings that have height information recorded. The work presented in [55] is similar, but focusing on eleven countries in Southeast Asia. In this paper, we follow similar approaches to assess spatial data quality and act in accordance with established definitions for the spatial data quality elements we focus on (elaborated further in Section 4), but we scale them, and we combine multiple aspects in the same study rather than focusing on one or few, and provide a comprehensive discussion focusing on the interpretation of the results, implications, and outlining recommendations.

3. Buildings in OpenStreetMap

This section overviews the basic concepts, explains the data structure, and it gives an exploratory analysis of the content of a recent snapshot of the database.

3.1. Overview

Mapping features in OpenStreetMap is based on a topological structure with four core elements: nodes, ways, closed ways and relations. Nodes are points that have their geographical position mapped with latitude and longitude. They are the basic elements that are used to map features best described as points (e.g. location of a bench in a park) and at the same time make up the other elements described in the continuation. Ways are connected lines of nodes which are often used to represent linear features such as roads, rivers, paths and so on. Closed ways are similar to ways — they are connected nodes but form a closed loop, yielding a polygon. They are the most common representation of buildings, and are used to map other areal features in the built environment such as a waterbodies and administrative areas. Relations are the most complex among the elements and consists of ordered lists of nodes, ways and/or relations as members. Relations are used to represent more complex shapes and structures, such as interstate highway that stretches over multiple sections, bus routes, multi-structure buildings, and areas with other nested allotments.

OSM elements can be attributed with one or more tags, which are key–value pairs that are used to store descriptive information (attributes) about the feature, e.g. postcode of a building, speed limit and width of a road, accessibility and working hours of an amenity, name and cuisine of a restaurant, circumference of the trunk and species of a tree, and type and operator of a vending machine. Some further self-explanatory examples pertaining to buildings are: `building=commercial`, `building:material=cement`, and `building:levels=4`. An example of an urban setting mapped in OSM is given in Fig. 1, with more details about a building modelled as a polygon together with several attributes associated to it.

The tag keys and values are flexible allowing users to input virtually any textual information, and to any extent (from no particular information to a long set). However, the OSM community maintains detailed documentation[2] with recommended tag keys and values for common map features. For many of these, there is comprehensive documentation that attempts to standardise mapping guidelines around the world, e.g. there are dozens of types of buildings listed as possible values, but using further values is allowed.

Some relevant tag keys are also encouraged to be used together to provide a richer description of the element. For example, a building tagged with `amenity=place_of_worship` and `religion=christian` indicates that the building not only has a religious function but it is a church, and a shop tagged with `shop=craft` and `level=3` describes a craft shop located on the third level of the building. It is also important to note that some tags are applicable for multiple feature type. For example, the tag `operator` is used to denote an entity that is in charge of an object such as building but also of a bus route and toll road. Such flexibility certainly has advantages, but it entails challenges when analysing it and requires a degree of data engineering. On the other hand, to avoid ambiguity, some of them specific to certain features are prefixed, e.g. `building:levels`, which is used to indicate the number of storeys of a building.

While most tag keys describe features that are universal globally, there may be some variations in interpreting them, and there are certain instances that only exist and/or are characteristic to certain geographic regions. For example, there are many buildings tagged with `static_caravan` in Arizona, USA, reflecting the local interest in

[2] https://wiki.openstreetmap.org/wiki/Main_Page

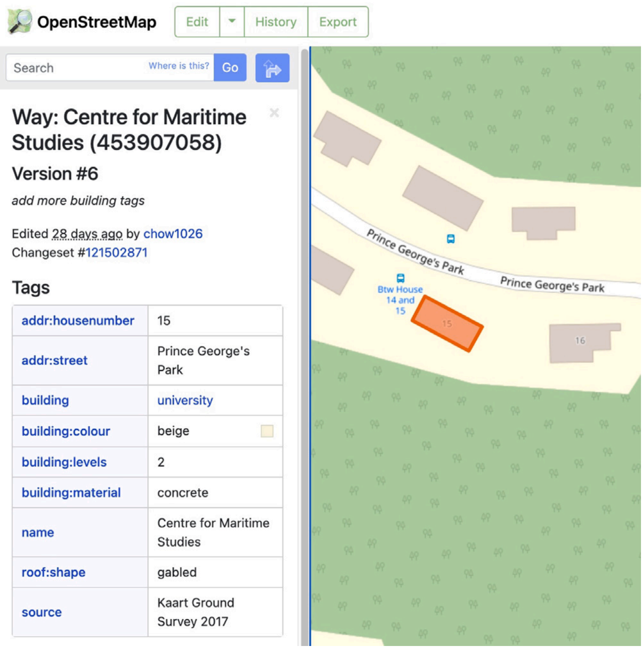


Fig. 1. An excerpt from OpenStreetMap illustrating a building with its tags such as the number of floors and material. The buildings are integrated with other urban features such as roads, bus stops, land cover, and restaurants. (c) OpenStreetMap contributors.

vagabond lifestyle and/or the particular interest of local contributors to map them.

The availability and selection of these building properties are quite varied and flexible, but in practice, they tend to be largely derived *visually* — e.g. a contributor may count the number of floors of a building from its exterior when passing next to it or from a street-level or aerial image. Therefore, information that is not always evident or easily obtainable, such as the year of a construction of a building, while supported by OSM, is scarce and often limited to buildings that have such records available publicly (e.g. from a wall plate or encyclopedia entry of a landmark or buildings in areas where the government releases cadastral information openly).

At the time of writing this paper, there are 507 million buildings mapped in OSM, which are described with more than 22 thousand unique tag keys. The large number of tag keys is caused by occasional spelling errors, auxiliary data (e.g. link to a local cadastral database), and some local variations and specifics (e.g. the intermittent use of a non-English language, and a very specific tag key that was used by a single contributor for just a few buildings in the entire global database) [46], which results in unique but isolated keys, and which are irrelevant in the context of this paper. Only a small number of the keys account for the vast majority of the attributes, which are in the focus of our research.

Because the underlying data in OSM is available freely according to an open licence and there are several means to obtain it, this wealth of information has been not only used by myriads of practitioners and researchers for spatial analyses, but also many accompanying web services and tools have been developed [19,111]. For example, Fig. 2 illustrates an instance of a web service that uses OSM data of buildings and generates 3D building models where their height information is available.

Finally, OSM allows mapping the indoor of features (e.g. floor plans), an aspect we also investigate in the paper owing also to the increased attention of indoor mapping in research [112], and importance for use cases in the built environment, e.g. assessing indoor air quality [113].

Table 1
Most frequent relevant building tag keys by categories.

Category	Tag keys
semantic	`name`
life cycle	`start_date`, `building:use`
location	`addr:street`, `addr:city`, `addr:postcode`, `addr:country`, `addr:state`, `addr:suburb`, `addr:place`, `addr:district`, `addr:housenumber`
structural	`building:levels`, `height`, `roof:shape`, `roof:levels`, `building:material`, `roof:material`, `roof:colour`, `building:colour`
interior	`building:flats`, `capacity`, `level`, `indoor`, `min_level`, `max_level`, `entrance`, `room`, `window`, `stairs`, `door`, `conveying`, `non_existent_levels`, `access`

3.2. Exploratory analysis

3.2.1. Contributors

The rapid growth of OpenStreetMap is much credited to the large community of dedicated volunteers. At the moment, there are more than 8 million registered users on OpenStreetMap. Among these users, 1.2M (14%) contributed at least once; and about 700k (approx. 7%) contributed to adding or editing buildings.

The contributions of these users also vastly varied. About 47% of the users contributed to less than 10 buildings, while 32% contributed between 10 and 99 buildings, 17% contributed between 100 and 999 buildings, and 3.5% contributed between 1000 to 9999 buildings. Less than 1% of the users contributed to mapping more than 10000 buildings. These high volume contributions are often a result of the bulk import of OSM data where contributors import government data or data from non-profit external sources after obtaining local buy-in and license to import [66,114,115]. Larger volume contributions could also be submitted by various organised initiatives such as the Humanitarian OpenStreetMap Team to help disaster response or development of less developed regions [116–118]. Further, there is an increasing trend of corporate contributors [70,119].

3.2.2. Tagging

Attributes are the main focus of this paper, thus, this section gives an overview of the prominent tag key pertaining to buildings. Among all the tag keys, our study focuses on 20 tags that are most frequently used to describe the properties of buildings in OSM. While that may seem as a very small subset, due to the flexibility and entangled tagging system, most of the tag keys are entirely scattered, and have no semantic meaning and no relevance in the building and geospatial sense. For example, they may refer to the source from which the information was provided, phone number, and smoking restrictions in the building. Further, as our results will show, virtually all but the selected tag keys are very seldom and localised, out of relevance of this study.

These selected building-related tag keys could be grouped into four categories: semantic, life cycle, location, and structural. Table 1 outlines them with their keys that are self-explanatory. In addition, we include several of them pertaining to indoor of buildings. Each of these pieces of building information has value for certain use cases. For example, the structural and interior tag keys are the attributes most pertinent to energy modelling, climate studies, vulnerability studies, and 3D building modelling as they describe the buildings' physical attributes, external and interior. Life cycle tag keys such year of construction provide value to various studies, e.g. refurbishment and material stock estimations [120].

4. Methodology

4.1. Overview

Our study focuses on the following three quality elements for building attributes, which comprehensively capture their quality and enable

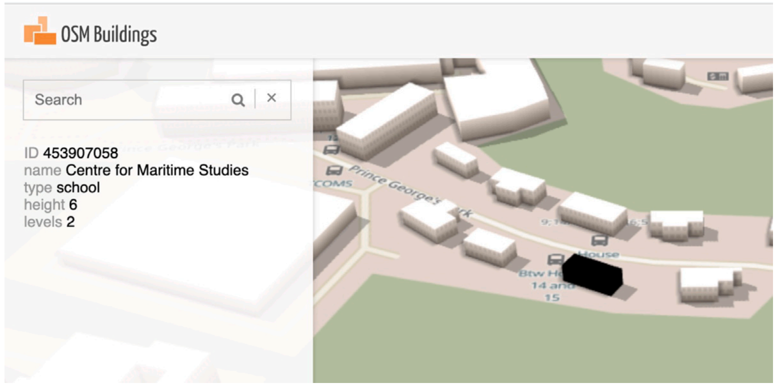


Fig. 2. Example of a rendered 3D building model provided by the web service OSM Buildings, generated using the same building information shown in Fig. 1. Map and geo data (c) OpenStreetMap contributors & 3D OSM Building.

understanding the fit for purpose: (1) completeness, (2) consistency, and (3) accuracy. These elements follow the frameworks and definitions established in the GIScience literature [121–123] (and see Section 2.2 for further references) — completeness evaluates the presence of attributes in an area (e.g. the percentage of buildings for which their type is mapped); consistency refers to adhering to logical rules of data structure and attribute rules (e.g. whether attribute values that are supposed to be numerical, such as number of storeys, are indeed expressed using numbers only); and accuracy evaluates whether the value of an attribute is correct. The approaches for the first two are intrinsic and their analyses cover the entire dataset, while the third one uses an external data source and it focuses on a subset of buildings since it cannot be scaled.

4.2. General principles

For this study, we obtained a complete copy of the OpenStreetMap database from the Planet OSM service,[3] and loaded into a local database. To organise the data and understand the spatial distribution of the quality and association with socio-economic aspects, we have used two additional datasets: GADM (the Database of Global Administrative Areas, a high-resolution database of multi-level administrative areas) and WorldPop (a global gridded dataset with population estimates) [124]. While all quality elements can be examined at the global level, such analysis gives only a general sense of the OSM quality, and an analysis at country and higher administrative division levels could potentially offer more detailed and stratified perspectives into the OSM quality of different states, provinces and/or counties/cities, and comparison among them. Thus, in addition to the overall global overview, thanks to the rich administrative data in GADM, the analysis was performed at the country level, and at the scale of three levels of administrative division to provide results that are easier and more meaningful to interpret. Further, coupling the OSM data with administrative information aids balancing the analysis so it covers the world evenly. The hierarchical administrative division levels adopted from GADM that we used in the study are 1, 2, and 3, which nearly all countries have. For example, Switzerland is divided into 26 cantons (level 1), 137 districts (level 2), and 2197 communes/municipalities (level 3), and each building has been assigned each of the matching subdivisions.

The results will show that analysing them by jurisdiction helps interpreting them better, e.g. understanding the completeness of attributes in a particular administrative unit such as city, which also helped us to identify best and worst instances of data at the urban scale in terms of quality. However, as administrative units are of irregular shape and size, and may not always be directly compared across countries (e.g. in the USA, level 1 corresponds to a state, which are often larger than many entire units at the higher level i.e. countries), the data has also been grouped according to a regular grid — we use a 1 × 1 km global grid adopted from WorldPop. This scale corresponds to a district or precinct scale, thus, it is suitable to consider in the context of use cases that operate at such scales, and it facilitates a global comparative analysis. This approach is also in line with studies that aggregate building information at the level of a cell, such as urban morphology and energy studies [43,125,126].

Because it may not be meaningful to analyse areas with a very small number of buildings, and because those may indicate poor building completeness, we analyse only areas with a certain number of buildings. The minimum building count thresholds were set to be 4000, 3000, 2000 and 1000 for the country, administrative divisions 1, 2 and 3, respectively. Due to the variations in the urban form around the world and different definitions of built-up areas [127], setting such a global threshold may be subjective, but it is an acceptable trade-off to filter out areas that may not be considered as built-up.

Among the attributes we analyse, we give particular focus to the building height, as our results will demonstrate that it is one of the most frequent ones, and it is an attribute found across many domains and use cases related to the built environment [128,129]. Because the number of floors is often used as a proxy for the height [130–132], we include it in the analysis as well. Further, this information is important for many use cases, e.g. estimating the floor space for building energy simulations and population estimations and estimating the volume [133–137]. By focusing on this aspect, our study also gives an understanding of the potential of OSM to be used for 3D building model generation around the world. Coupled with administrative data thanks to GADM, we are also able to analyse such potential for particular cities and derive what are the cities or districts with most potential.

4.3. Methodology for quality element 1: Completeness of attributes

In the first part of the study, we examine the completeness of building attribute tag keys per building. This part of the analysis is focused on understanding what is the frequency and spatial pattern of attributes that are most commonly associated to buildings. This is a relatively straightforward part of the study as it includes querying the database and analysing the results by the spatial subdivisions.

4.4. Methodology for quality element 2: Consistency of attributes

Consistency and validity of attribute information is instrumental to their interoperability and usability. For example, values of the types

[3] https://planet.openstreetmap.org/

of buildings may differ and not follow a consistent practice (e.g. using different values for the same information, such as 'residence' and 'residential'). As OSM only gives guidelines, and it does not strictly enforce rules and constrain such tags and values, this is an important consideration when understanding its quality. Technically, any combination is not incorrect, thus, we had to develop an own approach to determine whether a tag or value is valid.

For the study of consistency and validity of the building tag keys and tag values, we focus on building tag keys and values that turned out to have the highest number of discouraged keys and values that were in invalid format. The validity of the OSM tag keys was determined by referring to the OSM documentation for tag keys with `approved` or `de facto` tag status. These indicate the tag keys are generally accepted and supported by the community (but others are not barred). A tag key with an `in use` tag status is considered valid as long as it is not among the strongly discouraged list of tag keys.

The consistency and validity of the tag values were determined by comparing the commonly used values (for textual values) and the acceptable format(s) (for numeric data) documented in the OSM guidelines. A textual tag value is considered consistent if the lower-cased value is among the lower-cased common values used; whereas a numeric tag value is considered valid if it matches the acceptable format verified with regular expressions.

4.5. Methodology for quality element 3: accuracy of attributes

The aforedescribed completeness and consistency analyses will give a solid overview of how prevalent attribute information of buildings are and what is their integrity. However, those analyses do not touch upon the veracity of the information, which is the final quality element we analysed and describe in this section, and an element of considerable importance to understanding the usability and reliability of OSM data. For example, a residential building may be wrongly tagged as an office building, or its number of storeys may be inaccurate.

To verify the accuracy and correctness of the data, we used street-level imagery. This was used as the ground truth for the following reasons: (i) is a data source readily and freely available in scores of countries worldwide; (ii) enables us to inspect a large range of relevant building information due to its visual representation; and (iii) has been used frequently for many tasks in the geospatial domain [8]. However, its disadvantage is that it requires manual work when used for this purpose, its coverage is not fully global, and not all buildings are covered (imagery is usually taken from cars on driveable roads, not providing insights in buildings and parts of buildings not visible from them) [138]. Street-level imagery has been used previously to extract such building information [139]. There have been several recent efforts to do so automatically [30,140–146], but such research is still embryonic and experimental — it does not result in easy-to-use libraries. Next, the efforts are fragmented and narrow as papers tend to focus on deriving a single building property. Further, the prediction models are localised as they are usually trained in a single city, so they may not be transferable elsewhere. Thus, we resorted to a manual approach — locating a building on Google Street View (GSV), inspecting its image(s), and noting down its properties, which was relatively trivial in most cases (e.g. the number of storeys was evident in most cases thanks to counting rows of windows in residential buildings). This approach limits the number of buildings that we can inspect, but it (i) ensures a high level of accuracy (which is essential when assessing data quality); (ii) gives us insights into almost any building aspect that can be inferred visually from a pedestrian point of view, which is in line with the set of attributes usually recorded by contributors in OSM (revealed later in the results; see Section 5.1); (iii) as a building may be available in multiple street-level images with variable quality [147], it allows us to pick the best image in GSV; and (iv) enables us to understand challenges in obtaining and harmonising information on buildings [148].

4.6. System set up

Finally, we describe technical details. For the database, we used PostgreSQL, which was set up on a Dell T480 server with 8 processors and 128 GB RAM. Subsequently, PostGIS 3.2, PostGIS raster, hstore and fuzzystrmatch extensions were installed on the database, to support spatial data and operations. Loading the 68 GB OSM file to the database took up 1.3TB storage and it lasted 12 days. After loading the GADM and WorldPop data, and recasting the OSM data according to them, the database took up 1.5TB of disk space. For the analysis, we developed a series of SQL scripts that analyse the quality aspects, and after running them, the final database took up 3TB of storage.

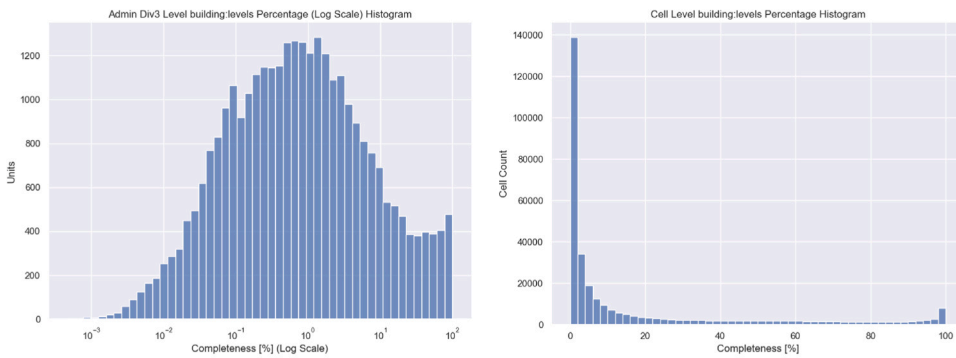
5. Results

5.1. Results for quality element 1: Completeness of attributes

Overall, after analysing all the buildings in the database, among the tag keys relevant to physical and structural attributes, 19.5% buildings where tagged with the type of the building. Next, 4.6% of buildings were tagged with `building:levels` (number of storeys), and only 2.9% had the `height` (height of a building) tag. There are 7% of them where at least either of the two are available. The completeness of location tag keys were slightly higher than other tags, e.g. for 11.8% buildings the street name is available. The year of construction of a building was tagged in 2.9% instances. All other attributes are available for less than 1% of buildings, and are thus, not considered further. The use of indoor tag keys were the least prevalent, most with below 0.01%, suggesting that OSM was not yet imbued with activities on indoor mapping, which are also conceptually supported. In general, buildings tend to be poorly described semantically, with only 1.2% of buildings having five tags or more.

It is clear that more often than not, buildings lack attribute information, and as such, the data may not appear to be suitable for analyses that require such information on buildings. However, as the continuation of the section will demonstrate, when the results are analysed spatially, our study affirms the findings of related work (Section 2.2) that the completeness around the world is highly heterogeneous, with a large range of degrees of data availability, from fully incomplete areas to those where some attributes are available for all buildings (Fig. 3).

With the vertical extent of a building being the most important attribute of buildings in many domains, we further look into `building:levels` and `height` — we dived deeper into the completeness of these 'vertical' tags at country, administrative division levels 1 through 3, and at the cell level, to understand their prevalence. In general, moving from the global scale and its low completeness overall, `building:levels` and `height` completeness at country level remain low. However, we found that there are concentrated areas from all over the world with very high completeness. Table 2 outlines countries and administrative divisions 1, 2, 3 with the most, while Table 3 lists the ones with the least completeness of `building:levels` tag keys. Tables 4 and 5 offer the same insights for the `height` tag. It is important to note that these results should be viewed in relative values, e.g. some areas may have more buildings tagged than other more complete areas because of the larger number of buildings, but ultimately, completeness is measured in relative (percentage) terms. First, the results underline that there is a large variation in recording building characteristics, with some countries having only a fraction of data, while others more than a third populated. Second, while we find that no country has a large degree of completeness, there are many lower level divisions such as counties that offer full or near-full completeness of these vital attributes. For example, we find that there are 443 administrative units at the third level (e.g. cities, municipalities) that have completeness of the number of storeys higher than 80%. At the district scale, there are 22,710 of them around the world, suggesting the high potential of the use of OSM, despite the lacking



(a) Administrative unit (level 3). (NB: log scale)   (b) Cells (1x1 km).

Fig. 3. Completeness of the number of storeys by different scale of analysis. The availability of such building information remains very low overall, but there are thousands of districts around the world that are complete with one or more attributes such as building height or type.

Table 2
Administrative units with the highest `building:levels` completeness.

	Administrative unit	Building count	Tagged	% tagged
Top level	Mexico	2,542,184	1,352,881	53.22%
	Czechia	4,945,588	2,266,515	45.83%
	Bolivia	281,946	115,113	40.83%
	Hong Kong	75,315	24,898	33.06%
	Azerbaijan	97,992	31,668	32.32%
Level 1	Mexico, Baja California	1,244,939	1,233,037	99.04%
	Mayotte, Bandraboua	4,126	3,976	96.36%
	Turkey, Bayburt	5,203	4,882	93.83%
	Argentina, San Juan	98,468	89,223	90.61%
	Botswana, Chobe	26,579	23,740	89.32%
Level 2	Brazil, Rio Grande do Sul, Santo Antônio das Missões	4,574	4,573	99.98%
	Colombia, Valle del Cauca, Caicedonia	6,140	6,134	99.90%
	Colombia, Meta, Granada	10,911	10,882	99.73%
	Algeria, Bouira, Mezdour	3,940	3,928	99.70%
Level 3	Brazil, Rio Grande do Sul, Santo Antônio das Missões	4,573	4,573	100.00%
	Brazil, Rio Grande do Sul, Bossoroca, Bossoroca	3,247	3,244	99.91%
	Tanzania, Dar es Salaam, Kinondoni, Hananasifu	2,984	2,979	99.83%

Table 3
Administrative units with the lowest `building:levels` completeness.

	Administrative unit	Building count	Tagged	% tagged
Top level	American Samoa	18,212	0	0.00%
	Benin	818,347	95	0.0116%
	Gambia	284,342	58	0.0204%
	Togo	1,001,556	212	0.0212%
	Mauritania	336,584	75	0.0223%
Level 1	Nigeria, Bauchi	815,211	38	0.00466%
	Nigeria, Sokoto	641,839	44	0.00686%
	Madagascar, Fianarantsoa	1,802,900	127	0.00704%
	Democratic Republic of the Congo, Sud-Kivu	1,079,786	77	0.00713%
	Democratic Republic of the Congo, Tanganyika	96,996	7	0.00722%
Level 2	India, West Bengal, South 24 Parganas	365,753	257	0.0703%
	Brazil, São Paulo, São Paulo	2,182,944	1872	0.0858%
	Nepal, Central, Janakpur	818,814	756	0.0923%
	Nepal, West, Lumbini	508,825	480	0.0943%
	Indonesia, Jawa Barat, Bogor	434,056	518	0.119%
Level 3	India, Karnataka, Bangalore, Bangalore	341,775	1666	0.487%
	India, NCT of Delhi, West, Delhi	232,961	1391	0.597%
	France, Nouvelle-Aquitaine, Gironde, Bordeaux	572,923	3732	0.651%
	United Kingdom, England, Northumberland, Northumberland	126,035	872	0.692%
	France, Bretagne, Morbihan, Vannes	321,480	2238	0.696%

completeness at the global scale. While many of the highly complete areas are likely a result of data imports, there are massive discrepancies in the interest in contributing with such information. Therefore, it may be more meaningful to consider such aspect at the district scale, which mirrors the scale of many built environment studies.

The number of floors is more frequent than the building height. The key reason is that it is visually discernible, while the height is often sourced from measurements to which contributors have no access or limited public information (e.g. the heights of landmark buildings are often publicised). It is relevant to note that, according to the OSM

Table 4
Administrative units with the highest `height` completeness.

	Administrative unit	Building count	Tagged	% tagged
Top level	Mexico	2,542,184	1,296,144	50.99%
	Brazil	7,118,436	2,616,895	36.76%
	Mayotte	70,554	19,803	28.07%
	Eritrea	29,569	4,901	16.57%
	United States	53,276,415	7,958,494	14.94%
Level 1	Mexico, Baja California	1,244,939	1,232,305	98.99%
	Mayotte, Bandraboua	4,126	3,976	96.36%
	Italy, Friuli-Venezia Giulia	622,969	596,833	95.80%
	Turkey, Bayburt	5,203	4,881	93.81%
	Puerto Rico, San Juan	102,656	95,253	92.79%
Level 2	Mexico, México, Cuautitlán	19,327	19,224	99.47%
	United States, New York, Kings	342,980	340,952	99.41%
	Mexico, Baja California, Tijuana	1,239,298	1,231,684	99.39%
	United States, New York, Queens	464,807	459,831	98.93%
	Romania, Suceava, Ipotesti	2,434	2,406	98.85%
Level 3	Brazil, São Paulo, São Paulo, Artur Alvim	25,183	25,139	99.83%
	Brazil, São Paulo, São Paulo, Ponte Rasa	37,408	37,325	99.78%
	Brazil, Ceará, Fortaleza, Antonio Bezerra	74,187	74,017	99.77%
	Italy, Lombardia, Lodi, Graffignana	1,812	1,807	99.72%
	Brazil, São Paulo, São Paulo, Cidade Lider	42,319	42,196	99.71%

Table 5
Administrative units with the lowest `height` completeness.

	Administrative unit	Building count	Tagged	% tagged
Top level	Barbados	161,994	0	0%
	Brunei	46,347	0	0%
	Grenada	45,436	0	0%
	Saint Lucia	44,070	0	0%
	Antigua and Barbuda	43,406	0	0%
Level 1	Madagascar, Fianarantsoa	1,802,900	3	0.0002%
	Madagascar, Antsiranana	953,106	2	0.0002%
	Tanzania, Singida	462,333	1	0.0002%
	Tanzania, Tabora	901,757	2	0.0002%
	Democratic Republic of the Congo, Nord-Kivu	1,322,186	3	0.0002%
Level 2	Ukraine, Kharkiv, Kupians'kyi	13,974	1	0.0072%
	India, Telangana, Ranga Reddy	521,173	75	0.0144%
	Indonesia, Jawa Barat, Bogor	434,056	70	0.0161%
	Bangladesh, Dhaka, Dhaka	926,498	171	0.0185%
	United States, Arizona, Maricopa	1,184,801	222	0.0187%
Level 3	Côte d'Ivoire, Abidjan, Abidjan, Abidjan	354,116	30	0.0085%
	India, Telangana, Ranga Reddy, n.a. (1728)	381,169	60	0.0157%
	India, Maharashtra, Pune, n.a. (1612)	226,173	47	0.0208%
	France, Pays de la Loire, Maine-et-Loire, Angers	300,121	72	0.0240%
	France, Grand Est, Meurthe-et-Moselle, Nancy	223,704	57	0.0255%

guidelines, the number of levels refers to the number of floors above ground including the ground floor, while underground levels are not included.

Drilling down the results, we find that there are different patterns in completeness. For only 1.4% of buildings for which the type is not known, the number of storeys is available, in a stark contrast with 17.4% of those where the type is available. This finding suggests that in many cases either no building characteristics tend to be acquired or when contributors acquire building properties they consider a few key of them together, which benefits use cases. There is also a large difference by type of building. For example, a break down by building function reveals that the number of storeys is available for 40% buildings that are tagged residential and 13.6% schools. These disparate results may also reveal challenges in collecting such data for various building typologies.

The spatial variation of the completeness among different entities at the same level has also been analysed visually. In Fig. 4, we illustrate the completeness of this attribute at the administrative level 2, for a balanced set of countries. Completeness is highly heterogeneous, with some regions in the same country reaching nearly 100% completeness, and being suitable for a set of use cases that require such information. Czechia is an example of relatively homogeneous (but not full) nationwide completeness, suggesting coordinated efforts or data imports from authoritative sources where it is allowed to do so, as the country has a nation-wide open government database with a set of building information but partial completeness [149]. Fig. 5 provides a similar overview, but at the different level of analysis — grids within cities. The insights maintain heterogeneity, indicating that cities that do not have complete data, may have districts and their larger parts fully complete in terms of some attributes. Such data give the means to many applications that do not require city-scale analysis, such as generating 3D city models of districts and neighbourhoods useful for microclimate and energy studies.

5.2. Results for quality element 2: Consistency of attributes

5.2.1. Tag keys

Table 6 lists discouraged tag keys, their frequencies and percentage of usage. This aspect is of high quality — the frequency of discouraged tags appears to be low despite the lack of enforcement around the usage of tag keys. Further analysis on the top three discouraged tag keys: `building:units`, `building:roof` and `building:age`, indicate where these undesirable tag keys were used at various administrative levels. For `building:units`, used to record the number of residential units (flats, apartments) in a building, they were most widely used

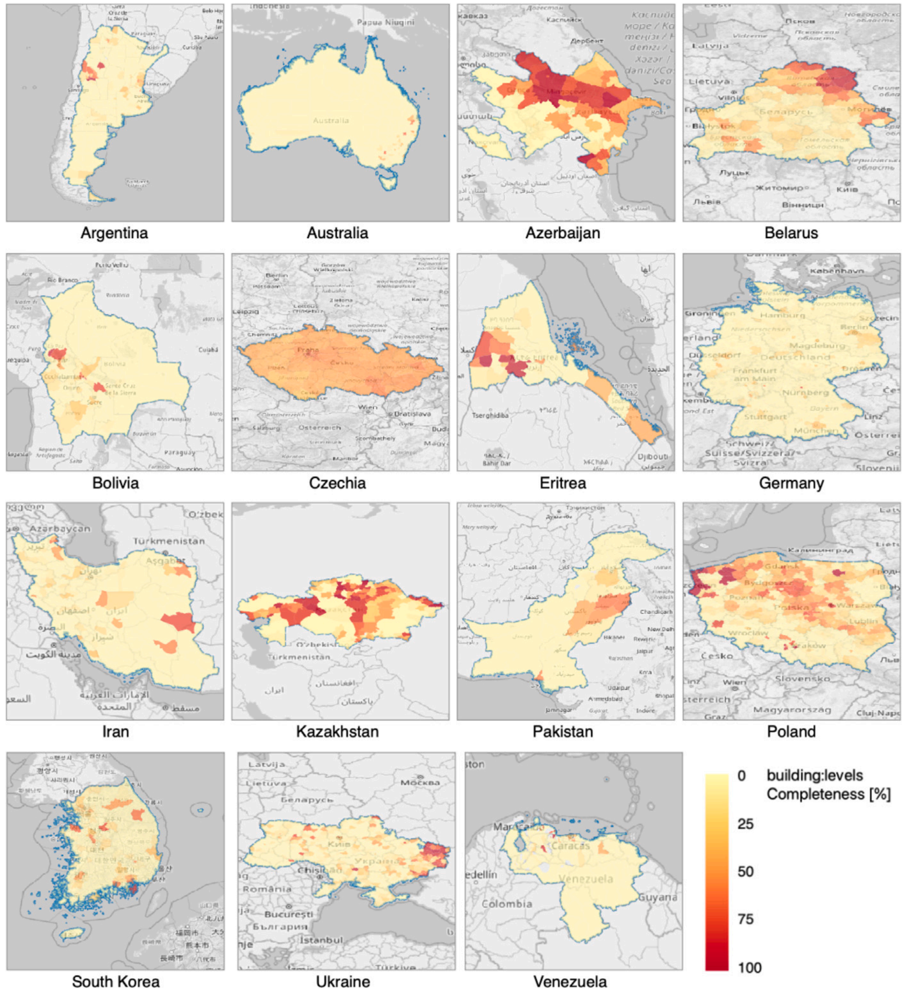


Fig. 4. The spatial distribution of the completeness of `building:levels` of a selected set of countries at the second level of administrative division (sorted by alphabet). The results affirm the heterogeneous discourse in OSM data quality analyses.

Table 6
`discouraged_tag` frequencies and percentage.

discouraged_tags	% corresponding correct tag	count tagged	% tagged
building:units	building:flats	2,835,628	0.5390
building:roof	roof:material	444,748	0.0845
building:age	start_date	251,994	0.0479
building:cladding	building:material	37,089	0.0071
building:roof:shape	roof:shape	33,602	0.0064
levels	level	26,071	0.0050
building:level	building:levels	17,503	0.0033

in the USA, likely because the term *units*, instead of *flats*, is more commonly used to refer to them in American English. This example suggests the local cultural and language influence in the choice of tag keys, and it may impair interoperability in comparative studies involving multiple cities. However, OSM is relatively consistent in the use of English for tag keys around the world (an exception are local names for features such as streets and neighbourhoods), and the occasional incorrect uses of tag keys rarely affects building information.

Further, the `building:roof` tag, which is supposed to be `roof:material`, are mostly observed in Chobe, Botswana, in 86.3% of its buildings, indicating that such issues are mostly isolated. Similar examples may be found for other tag keys. Observing these handful areas with high count and concentration of inconsistent tag keys, we suspect that these could be the result of organised mapping events by the local communities before the recommended tag keys were introduced. Caution needs to be exercised when using such data and account for possible variations of the nomenclature of attributes.

5.2.2. Tag values

Tag values are even less constrained than the tag keys. We referred to the OSM documentation for a list of commonly used tag values for describing buildings — `building:material`, `roof:shape`, `roof:material` and others. As for the seemingly numeric fields such as `height` and `building:levels`, there are some formats that the community deems acceptable. For example, the value of a building that is seven metres tall could be tagged with `height=7` or `height=7.0`, without a measurement unit (metre is considered default); or a unit can be added: `height=7 m`, and the value and unit are separated by a space. Heights in feet and inches are also possible and they are accepted in the format of `height=8'10''`, without space between the feet and inches symbols. As for building levels, any integer or numeric with decimals are acceptable. Subsequently, we verified the tag values of `height` and `building:levels` against these guidelines by matching the values using a series of regular expression (RegEx) we have developed for acceptable patterns.

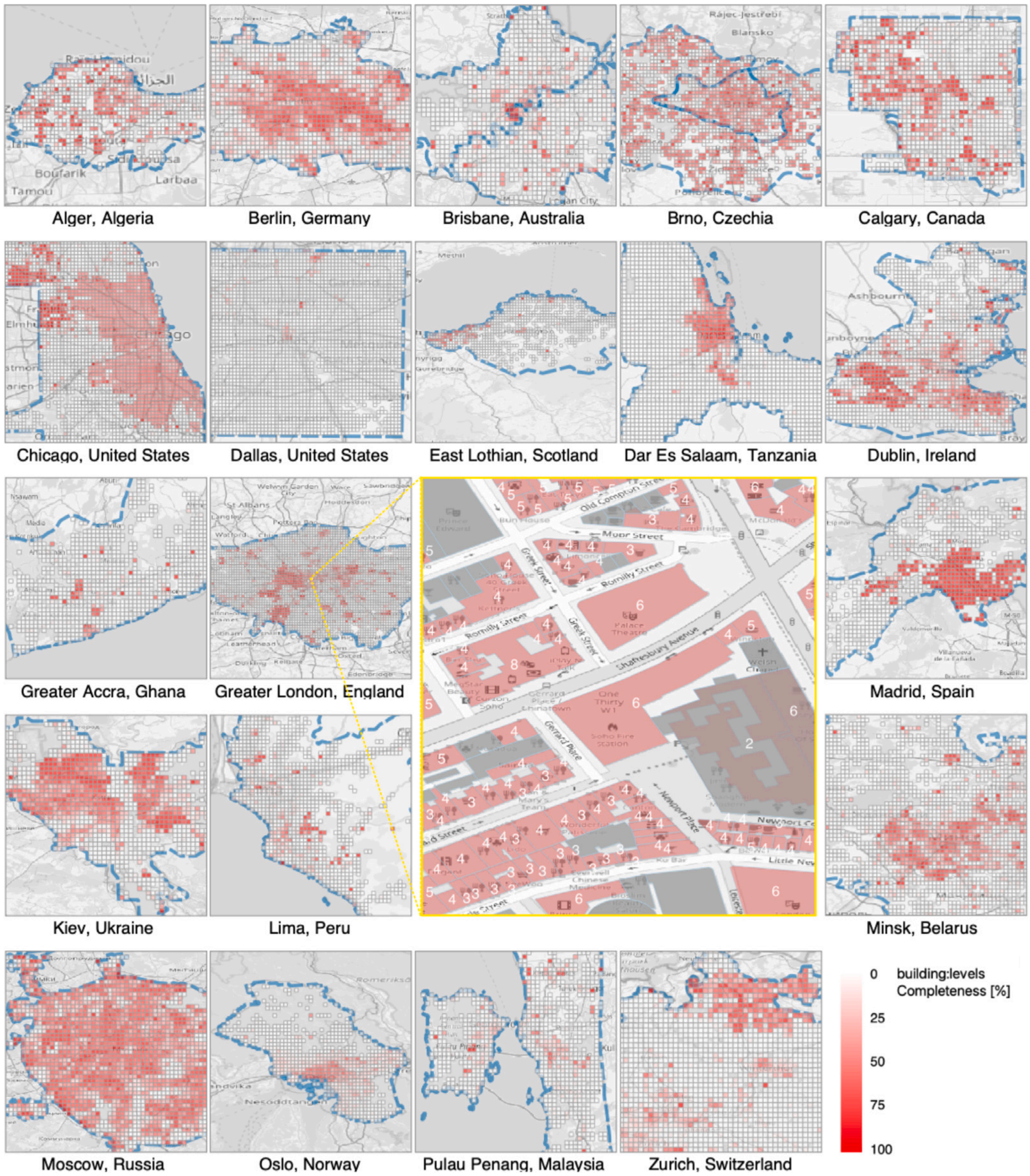


Fig. 5. The spatial distribution of the completeness of `building:levels` of a selected set of cities at the grid scale (1 × 1 km) (sorted by alphabet). The inset in the centre shows a zoomed part of London that has a high but not full level of completeness (the red polygons represent buildings that have the information on the number of storeys available, with the values stated in white). Such regions are still of use, as the data gaps may be filled automatically or manually, and the existing data may be sufficiently representative of the built form of the area. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Date fields such as `start_date` follow a set of much more complicated acceptance rules, inhibiting their validation. When verifying the consistency and validity of these fields, RegEx was developed to match patterns that were approved according to the lengthy and extensive documentation on the OSM wiki page.

An observation from the analysis is that – despite the high degree of freedom – all the tags whose values are numeric and measurable, particularly number of floors and height have a high share of validity with nearly all values being consistent (99% and above). The tags whose values are texts, e.g. roof shape and building material, have relatively lower validity but some of them have a very high level of validity (from 85% to 100%).

Some examples follow. For values where numerical information is expected (with optionally associated units), we have observed a variety of inconsistent and evidently wrong instances. For example, for `building:levels` there are values such as: '`G, 1,2,3,4`', '`Ground Level`', '`'1`', '`-1`', '`yes`', and '`1s`'. Examples for `height` include values: '`110m`', '`Unit 3`', '`5;41.99`', '`Brick`', '∼ 10'. These indicate that contributors may make typos, misunderstand the guidelines or not be aware of them, or misinterpret the purpose of a tag. However, there is no mechanism to strictly enforce a set of values. For descriptive text, consequently, it is no surprise that the validity is lower, as more flexibility is allowed and text can be more ambiguous than numerical values and subject to interpretation and judgement of the contributors. For example, a contributor may not only make typos (e.g. tag a building as '`residental`' instead of '`residential`'), but the same value can mean different things in different countries, e.g. in most countries, a dormitory refers to a shared building intended for college students, but in others such as in the UK it may refer to a shared room for multiple occupants, and another term (e.g. hall) is used instead. Next, free-standing small residential buildings tend to be labelled as '`house`' in most of the world. However, due to different semantics around the world, in certain countries, the same classes of buildings are tagged as '`detached`' reflecting the local practice of calling such buildings detached houses.

Thus, the tagging freedom that OSM accords is both a boon and a bane, and therefore, conducting consistency analyses is marred by such flexibility. A value is technically not incorrect (as OSM allows any; guidelines are standardised recommendations rather than being constrained), but it certainly causes issues in interoperability and usability, and users may need to invest effort into cleaning and harmonising such values before using them.

5.3. Results for quality element 3: Accuracy of attributes

Designing a balanced sample around the world (covering a share of buildings in each country), we selected 6578 buildings in OSM to inspect their street-level imagery in GSV. Since the majority of buildings has no attributes (Section 5.1), in this set we selected only those that have existing attributes. We first started with understanding whether GSV is available for the location. If so, we further examine if the building is visible from GSV or not (e.g. it may be obstructed by other structures or vegetation). If the building is visible, we then proceed to inspect a handful of visually verifiable attributes.

We found that 55% (3625) of the sampled buildings are located in areas that has coverage in GSV. Where the building in OSM had its type available, and where it was possible to infer the same from the street-level image, we find that 84.4% of them are correct, which is relatively accurate given the crowdsourced provenance. However, we must keep in mind that the evaluation of these values may be subjective and there is a blurry boundary in some values as they are very difficult to standardise all over the world. During the validation, we followed the OSM guidelines as strictly as possible, For example, we found cases where the value was ‘`commercial`’, but where ‘`retail`’ (used for a commercial building that houses primarily shops) could be more appropriate, a difference that for some use cases may not play a major role. Given such nuances, and considering that there are dozens of recommended values that contributors may have difficulties discerning between many ambiguous cases in practice (which ultimately may not significantly affect downstream analyses), the achieved accuracy may be interpreted as rather high.

Moving on to number of floors, we find that 72.2% buildings has a correct value. However, when tolerating uncertainty of one level (which may be within an acceptable margin of error), we find that there are 93.3% such values.

As for other descriptive attributes — the shape of the roof is correct in 82.8% of cases. A positive finding is that flat roofs are rarely labelled wrongly. Akin to the previously described building type errors, the wrong labels in this case are the similar roof shapes (e.g. gabled vs hipped). We will revisit this topic in the discussion. Next, the material is rarely recorded, and when available is wrong in more than half of cases. Such insight indicates that there is a lack of awareness of what the proper values used in OSM are, and that recording certain attributes remains a difficult challenge due to ambiguities in giving a single value per building (e.g. a building may have a mix of materials).

For none of the attributes we have encountered substantial regional variation in their accuracy. For example, the number of floors is correct for 73% buildings in Europe, 71% buildings in Asia, and also 71% in the Americas.

Overall, these results suggest variable accuracy of data, which may affect use cases (especially those in the energy and life cycle assessment domains that require materials [150]), and we describe them further in the next section.

6. Discussion

6.1. Overall findings

The fragmented local studies on OSM data quality agree on the highly heterogeneous quality. Our study affirms such findings at the global scale, with a comprehensive and integrated overview across multiple scales and aspects that reveal a huge degree of inequality globally.

Overall, most buildings in OSM have no attributes at all, and only their geometry (2D footprint) is available. Such quality of the data reflects the driving mechanisms in mapping them — mostly the location seems to be important in these, but not the attributes. For example, there are many use cases that rely solely on the 2D (footprint) building geometry, without any attributes [151], and there may be a lack of awareness among contributors about the use and importance of such information. Even though OSM may be the most comprehensive database on building information, it is still largely incomplete.

That said, a key result of our work is that OSM building data in many places is of sufficient quality for a variety of use cases (this matter will be discussed further in Section 6.3). There are sizeable spatial extents around the world for which information such as building use and number of storeys is available for nearly all buildings (cf. Fig. 5). Unsurprisingly, the most frequent tags are the ones that can be collected ‘visually’, such as function of a building.

Even though in the vast majority of areas the completeness of the attributes is insufficient for most analyses, that does not mean that OSM does not have some use even in such areas. First, government datasets may be outdated, and OSM could be used to supplement them with newer information or to detect changes [12]. Second, data from the authorities may also be incomplete. For example, a recent study has found that there are European countries where governments also have only partial completeness of attributes [152]. Thus, OSM could be used in a symbiosis with them to mitigate each others’ gaps.

Because of the scale of our analysis, it is not possible to obtain a government dataset as reference, typically used in related work (Section 2.2). However, that is an advantage, because in studies that use government data as ground truth, a question may arise whether some of such data has been adopted from such datasets, a possibility unbeknown if not documented. Thus, areas where government data is available may not be representative of other areas. Our study does not suffer from such bias. However, we acknowledge the inability to verify the recorded accuracy of all buildings worldwide as a potential limitation that is inevitable with such scale of research.

6.2. High variation of quality

The highly heterogeneous OSM building quality across the world prompts the question what affects the quality of the OSM data at the different aggregated levels. Take `building:levels` as an example — one of the most tagged building attribute we studied, most countries exhibited very low completeness overall, with less than 10% of their building tagged with `building:levels` (Fig. 4 and Table 2). Large swaths of land in the vast majority of countries have zero or close to zero percentage of `building:levels` completeness, with selected pockets with relatively higher completeness percentage.

The heterogeneous quality is not just observed at the country level, it is also very much observed at the more granular administrative division levels. We studied the `building:levels` completeness of a handful of administrative division 2 areas which are some moderately to highly populated cities, illustrated in Fig. 5. In these administrative division 2 areas, the cells around the city area generally showed higher `building:levels` completeness percentage whereas the surrounding areas show slightly lower completeness percentage. Zooming into a cell in London we could observe the sporadic pattern in which the `building:levels` being tagged.

The descriptive tag keys with text tag values have even higher variance, resulting in lower consistency in their tag values. We suspect there could be multiple reasons for this pattern. One could be because the spectrum of word choices could be more broader than what could be completely documented on the OSM wiki. Besides, the same material could be referred differently, subject to local cultural and language differences. In addition, based on the distribution of the number of buildings mapped per mapper, we could conclude that most mappers tend to map less than 10 buildings. These mappers may lack the

Fig. 6. Correlations of external factors and the percentages of buildings populated with particular tags. This table refers to values at the country level.
Source: The World Bank (2021).

knowledge about aspects such as building materials; and may have not checked the guidelines in detail or received ample training and guidance that are provided during organised mapping events. It is also not uncommon for any building to be constructed with multiple types of materials and have multiple roof shapes and/or materials. Thus, when tagging these properties, the choice of tag values is highly subjected to personal opinion of the mapper.

6.3. Implications for use cases

Understanding whether these results are favourable for certain use cases depends on their data requirements. For some, certain areas will provide sufficient information, but not for others. Error propagation studies have to be conducted (e.g. see [153–156]) for specific use cases and specific attributes, to understand the sensitivity of an analysis for input data errors, and provide further interpretation of the results.

The availability of some tags is poor everywhere, and will inhibit use cases that mandate them. For example, building material is useful for understanding outdoor thermal comfort [157], but our study exposes the very low completeness of this building property, with very few areas that have a level of completeness worth mentioning. On the other hand, the information on the height of a building is available for a large number of cities (i.e. administrative divisions at the level 3), likely satisfying a range of use cases.

While fully complete areas are scarce, OSM can still be a valuable data source for many use cases. First, data gaps can be filled automatically based on surrounding buildings [9,46,158,159] (see Fig. 5 — buildings closer to each other tend to have similar values); and manually, e.g. in a solar potential estimation study relying on 3D building models from OSM [42], the height values of a few buildings missing such attribute have been collected manually, not requiring particular effort (see the same map of London — if one was conducting a district-scale analysis requiring the heights of buildings, they might simply manually fill such information for the few buildings missing it from other sources). We also find that some gaps could be filled easily. For example, buildings of type garage are one of the most commonly tagged, but for only 7.5% of them, the building storeys are available. Still, because garages (especially those of small footprint and next to houses) tend to have only one storey, such gap could be bridged easily with basic heuristics. Further, recent work developing unconventional means to acquire building information, such as predicting the number of floors from other attributes using machine learning [160], and inferring the type of buildings based on surrounding carparks and other features [161,162] and from real estate portals [163,164], may be worth considering to fill such data gaps. Existing information in OSM may be sufficient to serve as training and validation data. Second, many use cases can be conducted at the district scale and do not require city-wide data [3,153], and for such, OSM can provide data of sufficient quality for many areas around the world. Third, there are use cases that do not require data of very high quality. For example, many studies in the built environment domain [165–167] infer the general built form in neighbourhoods, an application that does not require data on the height of all buildings in an area, as the ones that are available may be sufficiently representative of the rough form of the entire neighbourhood.

6.4. Association of quality with other factors

Having affirmed the diversity of OSM building data quality, we investigated if there are any associations between the quality elements and other factors. There is almost no correlation observed in nearly all the variables, thus, explaining patterns remains difficult and it seems arbitrary, which is not surprising given the large number of diverse contributors. Also, we find that there is no correlation among the quality elements, e.g. between the completeness and validity of `building:levels` and `height` tag values. Because of the large number of combinations and scales of analysis, we describe a subset of them.

Examining external factors that may potentially explain the completeness and validity of the attributes in an area, we considered predictors such as the area, population, street network morphology (e.g. total road length, sinuosity) and the national normalised GDP. At the country level, we show a correlation table (Fig. 6) that analyses the association of the completeness of a set of tags with a series of external factors. None of the pairs exhibits even a moderate amount of correlation. The results at other scales and other quality aspects are similar. There may be multiple reasons for the lack of such associations, e.g. with wealth and living standard. First, much of mapping efforts are humanitarian focusing on less developed regions that are experiencing disasters and are often contributed from overseas [116], and may receive a disproportionate amount of attention from mappers. Second, the mapping inequalities may be caused also by political oppression, legislation or conflicts, e.g. in some relatively developed regions such as China, there are restrictions on mapping (i.e. crowdsourced mapping efforts are illegal), rendering their quality much lower than other countries at a similar level of development [63,168,169].

6.5. Street-level image inspection for accuracy

To facilitate future studies, we provide more insight into the third quality element. Although the process of inspecting buildings and

(a) A building that may be residential or a school. (b) A building whose function is difficult to infer.

Fig. 7. Ambiguity in understanding the function of a building from street-level imagery.
Source of the imagery: Google Street View.

inferring their characteristics such as the number of floors may sound straightforward in theory, it is not so in practice. The images from GSV do not always seamlessly reveal building information and may be warped in some instances, and they are not always clear enough to facilitate the process and come up with a conclusive decision.

Despite GSV being an established service with a global focus, there are still many parts around the world that are not covered by GSV yet. Also, some of the sampled buildings are not visible in the images, as they were blocked by other buildings, fences, trees or are too far down at the end of the streets that it could not be fully visible. Further, not all attributes can be inferred from imagery (e.g. the age of a building), and would require authoritative data sources, which are not available at such scale. Therefore, while GSV is in many cases an excellent data source with many advantages such as offering multiple viewpoints of the same building and making it easy to discern many building features, it cannot be considered a complete and global solution to quality assurance (as it is the case for any other potential reference data source). Lastly, as with any manual data collection, in the same way as data acquisition in OSM, the labelling for building, building material, `roof:shape`, and `roof:material` tags are subject to the person's knowledge of possible values of each of those shapes, and his/her personal judgement. For the buildings that are visible on GSV, the high subjectivity in `building`, `building:material`, `roof:shape` values makes it challenging to verify accuracy. These challenges permeate various lineages and stakeholders in charge of building data, including governments and companies, where despite detailed standards and trained staff, there are always ambiguous situations and a degree of subjectivity. In OSM, this issue is compounded further due to flexible rules and contributors that have a range of understanding of the guidelines and conscientiousness, thus, the results of our study are not surprising and mirror related efforts of collecting data on the function of the building stock [146].

Fig. 7 includes example images to illustrate the ambiguity of building functions and indecision one could face when choosing a building type value when mapping. Therefore, it might make sense that many buildings are not associated with its type. Fig. 8 shows some examples for `building:material`. On the left image, two different mappers gave two different building materials for the same building, demonstrating the subjectivity of personal opinion; and on the right image, masonry and stone could mean the very same thing, but these different choice of words showed how non-exhaustive the spectrum of texts could be to refer the same object as well as overlaps in possible values. These challenges make it hard and impossible to have a standardised way to quantify the accuracy and correctness of these values. Finally, Fig. 9 demonstrates more examples for the ambiguity and challenge when it comes to recording the type of the roof.

6.6. Recommendations for OpenStreetMap

The OSM community has made impressive advancements in the past several years, including mapping buildings, which are now used widely in studies. We outline a few potential points of improvement that may lead to increasing the quality of building information further, and consequently, the usability in research.

First, it might be worthwhile to attempt to motivate mappers to input attributes of buildings, i.e. when mapping them or to add them to existing buildings. Buildings seem to be a popular feature type to be mapped after roads, but it appears that the benefit of recording building properties is not equally favoured as they are still largely deficient in completeness. Second, unlike mapping roads, corporate contributors do not seem to have a particular interest in mapping buildings. Future efforts could focus on understanding business uses of such data to incentivise companies to include buildings in their mapping campaigns. Third, although OSM is built based on the premise of being receptive and open, and has a well documented set of guidelines, with its growing contributors from various backgrounds and applications in various domains, it could benefit from being more stringent with tag keys and values. Fourth, we deem that a point of improvement is enhancing guidelines on the privacy and safety aspects when mapping buildings, and their enforcement. OSM provides guidelines on limitations on mapping private information[4] and it appears that violations of the privacy of people are minimised, but most of the guidelines are general not focusing specifically on buildings, they are not binding, and they do not always have a strong consensus. For example, there is an agreement in matters such as that the individual ownership of features such as buildings should not be recorded, that it is not allowed to add the names of inhabitants in dwellings, and that mapping certain information such as the location of safe houses for victims of domestic violence should not be allowed. Such agreements apply globally. On the other hand, it remains less conclusive in aspects that may vary globally due to cultural and legal differences, such as mapping private backyards and features associated to buildings (e.g. private swimming pools, sheds), and it defers to local laws about particular aspects, such as mapping military buildings, which may be challenging to enforce and track.

6.7. Limitations and future work

We believe that our study provides a solid and overarching understanding the quality of OSM building information at the global scale. While a study such as ours has the benefit of a holistic overview, that comes at the expense of not being able to elaborate on all the attributes and geographical areas.

Our study focuses on attributes, so it does not account for global building completeness. That is, our work may suggest that in a city the building attribute completeness is very high with almost all buildings having certain tags, but in reality it can happen that not all the buildings are actually mapped. Unfortunately, OSM building completeness studies tend to be limited in scale. However, very recent efforts [170,171] may change that, e.g. the latter preprint suggests that

[4] https://wiki.openstreetmap.org/wiki/Limitations_on_mapping_private_information

(a) Brick or concrete? (b) Masonry or stone?

Fig. 8. Ambiguity in building material information.
Source of the imagery: Google Street View.

(a) Gabled or gambrel? (b) Gabled or skillion?

Fig. 9. Ambiguity in type of the roof.
Source of the imagery: Google Street View.

globally, 21% buildings are mapped, and for more than a thousand cities worldwide, building completeness exceeds 80%. We hope that in the future our results can be connected with such studies and examined together.

7. Conclusion

Geospatial building information sourced from OpenStreetMap, both geometric and descriptive attributes, have gained a foothold in multiple domains across the built environment thanks to its liberal licence, growth in completeness of the building stock mapped, and raising awareness of this crowdsourced platform of geospatial data. However, quality remains a concern, and no studies understanding the attribute content have been conducted at the global scale. As the volume of buildings mapped in OSM has increased dramatically all over the world, and the studies making use of such information are multiplying, we presented a timely global study on the quality of semantic building information in OpenStreetMap examining multiple dimensions of relevance to studying the building stock and the built environment. It is the first such study to the extent of our knowledge, and we believe that it will be found of interest to researchers and practitioners in the built environment community, from those using building data for developing decarbonisation strategies and understanding life cycle assessment to those conducting urban energy and urban climate studies at the district and urban scale, which are increasingly relying on OpenStreetMap. Other contributions are that it has a variety of use cases in focus and that analyses attributes, which have been somewhat overlooked in OSM data quality assessments, and that it brings attention to the critical topic of spatial data quality in this community. This topic is important across multiple domains, not only from the user point of view, but also as crowdsourcing has been gaining interest in domains such as building energy and environmental research [172,173] and as the topic of open data is becoming increasingly important [174].

The key result is in line with localised studies — the quality of OpenStreetMap greatly varies across countries and administrative divisions, thus, the answer whether the building data is good enough depends on the geographical extent, which specific spatial data quality element, which set of information (i.e. specific building attributes), and the purpose (for what use case the data would be used). Building information remains scarce and fragmented, but in many of them, especially smaller units, the quality is sufficiently good for a number of use cases, such as microclimate, urban morphology, and energy modelling, which require basic information on buildings such as height, and may be conducted at the precinct or district scale. For example, we found that more than 20 thousand built-up 1 × 1 km cells have information on the heights of at least 80% buildings (cf. Fig. 5). In addition, some aspects exhibit consistently high accuracy — the numeric tag values are of high validity.

Another valuable lesson from this study is that it is impractical and unscaleable to comprehensively measure the accuracy and correctness of the OSM tag values by comparing them to street view imagery showing the buildings in reality. Nevertheless, we believe that the large and well balanced sample of buildings gives a good indication about the general accuracy of the data and variations across different regions. The accuracy of the text-based tag values is also hard to quantify because it is highly subjective to personal knowledge and opinion. In conclusion, scaling spatial data quality assessments and conducting them across many countries with different architectures and morphologies remains a challenge, but we hope that this study brings this research line a step forward and set the scene for subsequent work.

While new buildings continue to be mapped and existing ones enhanced with information on a daily basis, we believe that the findings of our study will remain valid for several years as many of the elaborated advantages, trends, and challenges will persist.

CRediT authorship contribution statement

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

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is sourced from OpenStreetMap, which is available openly.

Acknowledgements

We thank the OpenStreetMap community for their valuable and impressive work on mapping the world, and the members of the NUS Urban Analytics Lab for the discussions. We highly appreciate the pertinent comments by the five anonymous reviewers, which have improved the paper. We gratefully acknowledge Patrick Janssen for providing the hardware on which this study was run. 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.

References

[1] A. Malhotra, J. Bischof, A. Nichersu, K.-H. Häfele, J. Exenberger, D. Sood, J. Allan, J. Frisch, C. van Treeck, J. O'Donnell, G. Schweiger, Information modelling for urban building energy simulation—A taxonomic review, Build. Environ. 208 (2022) 108552, http://dx.doi.org/10.1016/j.buildenv.2021.108552.

[2] C. Wang, M. Ferrando, F. Causone, X. Jin, X. Zhou, X. Shi, Data acquisition for urban building energy modeling: A review, Build. Environ. 217 (2022) 109056, http://dx.doi.org/10.1016/j.buildenv.2022.109056.

[3] T. Novosel, M. Grozdek, J. Domac, N. Duić, Spatial assessment of cooling demand and district cooling potential utilizing public data, Sustainable Cities Soc. 75 (2021) 103409, http://dx.doi.org/10.1016/j.scs.2021.103409.

[4] N. Szarka, F. Biljecki, Population estimation beyond counts—Inferring demographic characteristics, PLoS One 17 (4) (2022) e0266484, http://dx.doi.org/10.1371/journal.pone.0266484.

[5] H. Ning, Z. Li, X. Ye, S. Wang, W. Wang, X. Huang, Exploring the vertical dimension of street view image based on deep learning: a case study on lowest floor elevation estimation, Int. J. Geogr. Inf. Sci. 36 (7) (2022) 1317–1342, http://dx.doi.org/10.1080/13658816.2021.1981334.

[6] C. Zhang, H. Fan, G. Kong, VGI3D: an interactive and low-cost solution for 3D building modelling from street-level VGI images, J. Geovisualization Spat. Anal. 5 (2) (2021) 18, http://dx.doi.org/10.1007/s41651-021-00086-7.

[7] Y. Xie, J. Cai, R. Bhojwani, S. Shekhar, J. Knight, A locally-constrained YOLO framework for detecting small and densely-distributed building footprints, Int. J. Geogr. Inf. Sci. 34 (4) (2019) 1–25, http://dx.doi.org/10.1080/13658816.2019.1624761.

[8] F. Biljecki, K. Ito, Street view imagery in urban analytics and GIS: A review, Landsc. Urban Plan. 215 (2021) 104217, http://dx.doi.org/10.1016/j.landurbplan.2021.104217.

[9] N. Milojevic-Dupont, N. Hans, L.H. Kaack, M. Zumwald, F. Andrieux, D. de Barros Soares, S. Lohrey, P.-P. Pichler, F. Creutzig, Learning from urban form to predict building heights, PLoS One 15 (12) (2020) e0242010, http://dx.doi.org/10.1371/journal.pone.0242010.

[10] C. Meng, Y. Song, J. Ji, Z. Jia, Z. Zhou, P. Gao, S. Liu, Automatic classification of rural building characteristics using deep learning methods on oblique photography, Build. Simul. 15 (6) (2022) 1161–1174, http://dx.doi.org/10.1007/s12273-021-0872-x.

[11] O.M. Garbasevschi, J.E. Schmiedt, T. Verma, I. Lefter, W.K.K. Altes, A. Droin, B. Schiricke, M. Wurm, Spatial factors influencing building age prediction and implications for urban residential energy modelling, Comput. Environ. Urban Syst. 88 (2021) 101637, http://dx.doi.org/10.1016/j.compenvurbsys.2021.101637.

[12] F. Biljecki, L.Z.X. Chew, N. Milojevic-Dupont, F. Creutzig, Open government geospatial data on buildings for planning sustainable and resilient cities, 2021, http://dx.doi.org/10.48550/ARXIV.2107.04023, URL http://arxiv.org/abs/2107.04023.

[13] J. Yuan, P.K.R. Chowdhury, J. McKee, H.L. Yang, J. Weaver, B. Bhaduri, Exploiting deep learning and volunteered geographic information for mapping buildings in Kano, Nigeria, Sci. Data 5 (1) (2018) 180217, http://dx.doi.org/10.1038/sdata.2018.217.

[14] Z. Zhang, Z. Qian, T. Zhong, M. Chen, K. Zhang, Y. Yang, R. Zhu, F. Zhang, H. Zhang, F. Zhou, J. Yu, B. Zhang, G. Lü, J. Yan, Vectorized rooftop area data for 90 cities in China, Sci. Data 9 (1) (2022) 66, http://dx.doi.org/10.1038/s41597-022-01168-x.

[15] W. Sirko, S. Kashubin, M. Ritter, A. Annkah, Y.S.E. Bouchareb, Y. Dauphin, D. Keysers, M. Neumann, M. Cisse, J. Quinn, Continental-scale building detection from high resolution satellite imagery, 2021, arXiv:2107.12283.

[16] X. Huang, C. Wang, Estimates of exposure to the 100-year floods in the conterminous United States using national building footprints, Int. J. Disaster Risk Reduct. 50 (2020) 101731, http://dx.doi.org/10.1016/j.ijdrr.2020.101731.

[17] R. Peters, B. Dukai, S. Vitalis, J. van Liempt, J. Stoter, Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the netherlands, Photogramm. Eng. Remote Sens. 88 (3) (2022) 165–170, http://dx.doi.org/10.14358/pers.21-00032r2.

[18] B. Dukai, R. Peters, T. Wu, T. Commandeur, H. Ledoux, T. Baving, M. Post, V. van Altena, W. van Hinsbergh, J. Stoter, Generating, storing, updating, and disseminating a country-wide 3D model, ISPRS - Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLIV-4/W1-2020 (2020) 27–32, http://dx.doi.org/10.5194/isprs-archives-xliv-4-w1-2020-27-2020.

[19] Y. Yan, C.-C. Feng, W. Huang, H. Fan, Y.-C. Wang, A. Zipf, Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience, Int. J. Geogr. Inf. Sci. 34 (9) (2020) 1–27, http://dx.doi.org/10.1080/13658816.2020.1730848.

[20] M. Hacar, Analyzing the behaviors of OpenStreetMap volunteers in mapping building polygons using a machine learning approach, ISPRS Int. J. Geo-Inf. 11 (1) (2022) 70, http://dx.doi.org/10.3390/ijgi11010070.

[21] C.I. Nievas, M. Pilz, K. Prehn, D. Schorlemmer, G. Weatherill, F. Cotton, Calculating earthquake damage building by building: the case of the city of Cologne, Germany, Bull. Earthq. Eng. (2022) 1–47, http://dx.doi.org/10.1007/s10518-021-01303-w.

[22] M. Cerri, M. Steinhausen, H. Kreibich, K. Schröter, Are OpenStreetMap building data useful for flood vulnerability modelling? Nat. Hazards Earth Syst. Sci. 21 (2) (2020) 643–662, http://dx.doi.org/10.5194/nhess-21-643-2021.

[23] C. Westrope, R. Banick, M. Levine, Groundtruthing OpenStreetMap building damage assessment, Procedia Eng. 78 (2014) 29–39, http://dx.doi.org/10.1016/j.proeng.2014.07.035.

[24] D. Paprotny, H. Kreibich, O. Morales-Nápoles, P. Terefenko, K. Schröter, Estimating exposure of residential assets to natural hazards in Europe using open data, Nat. Hazards Earth Syst. Sci. 20 (1) (2020) 323–343, http://dx.doi.org/10.5194/nhess-20-323-2020.

[25] J. Schiefelbein, J. Rudnick, A. Scholl, P. Remmen, M. Fuchs, D. Müller, Automated urban energy system modeling and thermal building simulation based on OpenStreetMap data sets, Build. Environ. 149 (2019) 630–639, http://dx.doi.org/10.1016/j.buildenv.2018.12.025.

[26] A. Alhamwi, W. Medjroubi, T. Vogt, C. Agert, GIS-based urban energy systems models and tools: Introducing a model for the optimisation of flexibilisation technologies in urban areas, Appl. Energy 191 (2017) 1–9, http://dx.doi.org/10.1016/j.apenergy.2017.01.048.

[27] C. Wang, S. Wei, S. Du, D. Zhuang, Y. Li, X. Shi, X. Jin, X. Zhou, A systematic method to develop three dimensional geometry models of buildings for urban building energy modeling, Sustainable Cities Soc. 71 (2021) 102998, http://dx.doi.org/10.1016/j.scs.2021.102998.

[28] A. Alhamwi, W. Medjroubi, T. Vogt, C. Agert, OpenStreetMap data in modelling the urban energy infrastructure: a first assessment and analysis, Energy Procedia 142 (2017) 1968–1976, http://dx.doi.org/10.1016/j.egypro.2017.12.397.

[29] J. Valdes, S. Wöllmann, A. Weber, G. Klaus, C. Sigl, M. Prem, R. Bauer, R. Zink, A framework for regional smart energy planning using volunteered geographic information, Adv. Geosci. 54 (2020) 179–193, http://dx.doi.org/10.5194/adgeo-54-179-2020.

[30] K. Mayer, L. Haas, T. Huang, J. Bernabé-Moreno, R. Rajagopal, M. Fischer, Estimating building energy efficiency from street view imagery, aerial imagery, and land surface temperature data, Appl. Energy 333 (2023) 120542, http://dx.doi.org/10.1016/j.apenergy.2022.120542.

[31] R. Ma, T. Wang, Y. Wang, J. Chen, Tuning urban microclimate: A morpho-patch approach for multi-scale building group energy simulation, Sustainable Cities Soc. 76 (2022) 103516, http://dx.doi.org/10.1016/j.scs.2021.103516.

[32] M. Mortezazadeh, L.L. Wang, Solving city and building microclimates by fast fluid dynamics with large timesteps and coarse meshes, Build. Environ. 179 (2020) 106955, http://dx.doi.org/10.1016/j.buildenv.2020.106955.

[33] C.C. Fonte, P. Lopes, L. See, B. Bechtel, Using OpenStreetMap (OSM) to enhance the classification of local climate zones in the framework of WUDAPT, Urban Clim. 28 (2019) 100456, http://dx.doi.org/10.1016/j.uclim.2019.100456.

[34] X. Li, B. Yang, F. Liang, H. Zhang, Y. Xu, Z. Dong, Modeling urban canopy air temperature at city-block scale based on urban 3D morphology parameters– A study in Tianjin, North China, Build. Environ. 230 (2023) 110000, http://dx.doi.org/10.1016/j.buildenv.2023.110000.

[35] J. Schilling, J. Tränckner, Estimation of wastewater discharges by means of OpenStreetMap data, Water 12 (3) (2020) 628, http://dx.doi.org/10.3390/w12030628.

[36] R. Braun, R. Padsala, T. Malmir, S. Mohammadi, U. Eicker, Using 3D CityGML for the modeling of the food waste and wastewater generation—A case study for the city of Montréal, Front. Big Data 4 (2021) 662011, http://dx.doi.org/10.3389/fdata.2021.662011.

[37] A.N. Wu, F. Biljecki, Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability, Landsc. Urban Plan. 214 (2021) 104167, http://dx.doi.org/10.1016/j.landurbplan.2021.104167.

[38] D. Feldmeyer, C. Meisch, H. Sauter, J. Birkmann, Using OpenStreetMap data and machine learning to generate socio-economic indicators, ISPRS Int. J. Geo-Inf. 9 (9) (2020) 498, http://dx.doi.org/10.3390/ijgi9090498.

[39] J. Kim, M. Kent, K. Kral, T. Dogan, Seemo: A new tool for early design window view satisfaction evaluation in residential buildings, Build. Environ. 214 (2022) 108909, http://dx.doi.org/10.1016/j.buildenv.2022.108909.

[40] N. Zhang, Z. Luo, Y. Liu, W. Feng, N. Zhou, L. Yang, Towards low-carbon cities through building-stock-level carbon emission analysis: a calculating and mapping method, Sustainable Cities Soc. 78 (2022) 103633, http://dx.doi.org/10.1016/j.scs.2021.103633.

[41] Y. Chen, J. Yang, R. Yang, X. Xiao, J.C. Xia, Contribution of urban functional zones to the spatial distribution of urban thermal environment, Build. Environ. 216 (2022) 109000, http://dx.doi.org/10.1016/j.buildenv.2022.109000.

[42] A. Palliwal, S. Song, H.T.W. Tan, F. Biljecki, 3D city models for urban farming site identification in buildings, Comput. Environ. Urban Syst. 86 (2021) 101584, http://dx.doi.org/10.1016/j.compenvurbsys.2020.101584.

[43] F. Biljecki, Y.S. Chow, Global building morphology indicators, Comput. Environ. Urban Syst. 95 (2022) 101809, http://dx.doi.org/10.1016/j.compenvurbsys.2022.101809.

[44] X. Deng, W. Nie, X. Li, J. Wu, Z. Yin, J. Han, H. Pan, C.K.C. Lam, Influence of built environment on outdoor thermal comfort: A comparative study of new and old urban blocks in Guangzhou, Build. Environ. (2023) 110133, http://dx.doi.org/10.1016/j.buildenv.2023.110133.

[45] J. León, M. Vicuña, A. Ogueda, S. Guzmán, A. Gubler, C. Mokrani, From urban form analysis to metrics for enhancing tsunami evacuation: Lessons from twelve Chilean cities, Int. J. Disaster Risk Reduct. 58 (2021) 102215, http://dx.doi.org/10.1016/j.ijdrr.2021.102215.

[46] A. Bandam, E. Busari, C. Syranidou, J. Linssen, D. Stolten, Classification of building types in Germany: A data-driven modeling approach, Data 7 (4) (2022) 45, http://dx.doi.org/10.3390/data7040045.

[47] K. Bhuyan, C.V. Westen, J. Wang, S.R. Meena, Mapping and characterising buildings for flood exposure analysis using open-source data and artificial intelligence, Nat. Hazards (2022) 1–31, http://dx.doi.org/10.1007/s11069-022-05612-4.

[48] M. Over, A. Schilling, S. Neubauer, A. Zipf, Generating web-based 3D city models from OpenStreetMap: The current situation in Germany, Comput. Environ. Urban Syst. 34 (6) (2010) 496–507, http://dx.doi.org/10.1016/j.compenvurbsys.2010.05.001.

[49] A.N. Wu, F. Biljecki, GANmapper: geographical data translation, Int. J. Geogr. Inf. Sci. 36 (2022) 1394–1422, http://dx.doi.org/10.1080/13658816.2022.2041643.

[50] M. Naghavi, A.A. Alesheikh, F. Hakimpour, M.H. Vahidnia, A. Vafaeinejad, VGI-based spatial data infrastructure for land administration, Land Policy 114 (2022) 105969, http://dx.doi.org/10.1016/j.landusepol.2021.105969.

[51] N. Milojevic-Dupont, F. Creutzig, Machine learning for geographically differentiated climate change mitigation in urban areas, Sustainable Cities Soc. 64 (2021) 102526, http://dx.doi.org/10.1016/j.scs.2020.102526.

[52] S.A. Nitoslawski, N.J. Galle, C.K.V.D. Bosch, J.W. Steenberg, Smarter ecosystems for smarter cities? A review of trends, technologies, and turning points for smart urban forestry, Sustainable Cities Soc. 51 (2019) 101770, http://dx.doi.org/10.1016/j.scs.2019.101770.

[53] M. Ahmad, M.S.H. Khayal, A. Tahir, Analysis of factors affecting adoption of volunteered geographic information in the context of national spatial data infrastructure, ISPRS Int. J. Geo-Inf. 11 (2) (2022) 120, http://dx.doi.org/10.3390/ijgi11020120.

[54] S. Basiouka, C. Potsiou, E. Bakogiannis, OpenStreetMap for cadastral purposes: an application using VGI for official processes in urban areas, Surv. Rev. 47 (344) (2015) 333–341, http://dx.doi.org/10.1179/1752270615y.0000000011.

[55] F. Biljecki, Exploration of open data in Southeast Asia to generate 3D building models, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. VI-4/W1-2020 (2020) 37–44, http://dx.doi.org/10.5194/isprs-annals-vi-4-w1-2020-37-2020.

[56] L. Lucks, L. Klingbeil, L. Plümer, Y. Dehbi, Improving trajectory estimation using 3D city models and kinematic point clouds, Trans. GIS 25 (1) (2021) 238–260, http://dx.doi.org/10.1111/tgis.12719.

[57] A. Komadina, Z. Mihajlovic, Automated 3D urban landscapes visualization using open data sources on the example of the city of Zagreb, KN - J. Cartogr. Geogr. Inf. 72 (2) (2022) 139–152, http://dx.doi.org/10.1007/s42489-022-00102-w.

[58] H. Alsaad, M. Hartmann, R. Hilbel, C. Voelker, The potential of facade greening in mitigating the effects of heatwaves in central European cities, Build. Environ. 216 (2022) 109021, http://dx.doi.org/10.1016/j.buildenv.2022.109021.

[59] Y. Fang, L. Zhao, Assessing the environmental benefits of urban ventilation corridors: A case study in Hefei, China, Build. Environ. 212 (2022) 108810, http://dx.doi.org/10.1016/j.buildenv.2022.108810.

[60] M. Goetz, Towards generating highly detailed 3D CityGML models from OpenStreetMap, Int. J. Geogr. Inf. Sci. 27 (5) (2013) 845–865, http://dx.doi.org/10.1080/13658816.2012.721552.

[61] A. Scalas, D. Cabiddu, M. Mortara, M. Spagnuolo, Potential of the geometric layer in urban digital twins, ISPRS Int. J. Geo-Inf. 11 (6) (2022) 343, http://dx.doi.org/10.3390/ijgi11060343.

[62] M. Haklay, P. Weber, OpenStreetMap: User-generated street maps, IEEE Pervasive Comput. 7 (4) (2008) 12–18, http://dx.doi.org/10.1109/mprv.2008.80.

[63] W. So, F. Duarte, Cartographers of North Korea: Who are they and what are the technical, political, and social issues involved in mapping North Korea, Geoforum 110 (2020) 147–156, http://dx.doi.org/10.1016/j.geoforum.2020.02.008.

[64] G. Quattrone, L. Capra, P.D. Meo, There's no such thing as the perfect map, in: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work, ACM, 2015, http://dx.doi.org/10.1145/2675133.2675235.

[65] K. Moreri, Volunteer reputation determination in crowdsourcing projects using latent class analysis, Trans. GIS 25 (2) (2021) 968–984, http://dx.doi.org/10.1111/tgis.12713.

[66] R. Witt, L. Loos, A. Zipf, Analysing the impact of large data imports in OpenStreetMap, ISPRS Int. J. Geo-Inf. 10 (8) (2021) 528, http://dx.doi.org/10.3390/ijgi10080528.

[67] F. Botta, M. Gutiérrez-Roig, Modelling urban vibrancy with mobile phone and OpenStreetMap data, PLoS One 16 (6) (2021) e0252015, http://dx.doi.org/10.1371/journal.pone.0252015.

[68] Y. Xu, Z. Chen, Z. Xie, L. Wu, Quality assessment of building footprint data using a deep autoencoder network, Int. J. Geogr. Inf. Sci. 31 (10) (2017) 1–23, http://dx.doi.org/10.1080/13658816.2017.1341632.

[69] C. Kunze, R. Hecht, Semantic enrichment of building data with volunteered geographic information to improve mappings of dwelling units and population, Comput. Environ. Urban Syst. 53 (2015) 4–18, http://dx.doi.org/10.1016/j.compenvurbsys.2015.04.002.

[70] D. Sarkar, J.T. Anderson, Corporate editors in OpenStreetMap: Investigating co-editing patterns, Trans. GIS 26 (4) (2022) 1879–1897, http://dx.doi.org/10.1111/tgis.12910.

[71] J. Panek, L. Sobotova, Community mapping in urban informal settlements: Examples from Nairobi, Kenya, Electron. J. Inf. Syst. Dev. Ctries. 68 (1) (2015) 1–13, http://dx.doi.org/10.1002/j.1681-4835.2015.tb00487.x.

[72] S. Soman, A. Beukes, C. Nederhood, N. Marchio, L. Bettencourt, Worldwide detection of informal settlements via topological analysis of crowdsourced digital maps, ISPRS Int. J. Geo-Inf. 9 (11) (2020) 685, http://dx.doi.org/10.3390/ijgi9110685.

[73] B. Bechtel, M. Demuzere, P. Sismanidis, D. Fenner, O. Brousse, C. Beck, F.V. Coillie, O. Conrad, I. Keramitsoglou, A. Middel, G. Mills, D. Niyogi, M. Otto, L. See, M.-L. Verdonck, Quality of crowdsourced data on urban morphology—The human influence experiment (HUMINEX), Urban Sci. 1 (2) (2017) 15, http://dx.doi.org/10.3390/urbansci1020015.

[74] H. Ledoux, val3dity: validation of 3D GIS primitives according to the international standards, Open Geospatial Data Softw. Stand. 3 (2018) 1, http://dx.doi.org/10.1186/s40965-018-0043-x.

[75] F.J. Aguilar, J.P. Mills, Accuracy assessment of lidar-derived digital elevation models, Photogramm. Rec. 23 (122) (2008) 148–169, http://dx.doi.org/10.1111/j.1477-9730.2008.00476.x.

[76] S. de Bruin, G.B.M. Heuvelink, J.D. Brown, Propagation of positional measurement errors to agricultural field boundaries and associated costs, Comput. Electron. Agric. 63 (2) (2008) 245–256, http://dx.doi.org/10.1016/j.compag.2008.03.005.

[77] Y. Hou, F. Biljecki, A comprehensive framework for evaluating the quality of street view imagery, Int. J. Appl. Earth Obs. Geoinf. 115 (2022) 103094, http://dx.doi.org/10.1016/j.jag.2022.103094.

[78] J.-F. Girres, G. Touya, Quality assessment of the french OpenStreetMap dataset, Trans. GIS 14 (4) (2010) 435–459, http://dx.doi.org/10.1111/j.1467-9671.2010.01203.x.

[79] ISO, ISO 19157:2013 – Geographic Information – Data Quality, No. 19157, Tech. Rep., 2013, p. 146.

[80] H. Senaratne, A. Mobasheri, A.L. Ali, C. Capineri, M.M. Haklay, A review of volunteered geographic information quality assessment methods, Int. J. Geogr. Inf. Sci. 31 (1) (2017) 139–167, http://dx.doi.org/10.1080/13658816.2016.1189556.

[81] A. Basiri, M. Haklay, G. Foody, P. Mooney, Crowdsourced geospatial data quality: challenges and future directions, Int. J. Geogr. Inf. Sci. 33 (8) (2019) 1–6, http://dx.doi.org/10.1080/13658816.2019.1593422.

[82] Y. Zhao, W. Yang, Y. Liu, Z. Liao, Discovering transition patterns among OpenStreetMap feature classes based on the Louvain method, Trans. GIS (2021) http://dx.doi.org/10.1111/tgis.12843.

[83] G. Yeboah, J.P.d. Albuquerque, R. Troilo, G. Tregonning, S. Perera, S.A.K.S. Ahmed, M. Ajisola, O. Alam, N. Aujla, S.I. Azam, K. Azeem, P. Bakibinga, Y.-F. Chen, N.N. Choudhury, P.J. Diggle, O. Fayehun, P. Gill, F. Griffiths, B. Harris, R. Iqbal, C. Kabaria, A.K. Ziraba, A.Z. Khan, P. Kibe, L. Kisia, C. Kyobutungi, R.J. Lilford, J.J. Madan, N. Mbaya, B. Mberu, S.F. Mohamed, H. Muir, A. Nazish, A. Njeri, O. Odubanjo, A. Omigbodun, M.E. Osuh, E. Owoaje, O. Oyebode, V. Pitidis, O. Rahman, N. Rizvi, J. Sartori, S. Smith, O.J. Taiwo, P. Ulbrich, O.A. Uthman, S.I. Watson, R. Wilson, R. Yusuf, Analysis of OpenStreetMap data quality at different stages of a participatory mapping process: Evidence from slums in Africa and Asia, ISPRS Int. J. Geo-Inf. 10 (4) (2021) 265, http://dx.doi.org/10.3390/ijgi10040265.

[84] R.C. Sundaram, E. Naghizade, R. Borovica-Gajic, M. Tomko, Can you fixme? An intrinsic classification of contributor-identified spatial data issues using topic models, Int. J. Geogr. Inf. Sci. 36 (1) (2022) 1–30, http://dx.doi.org/10.1080/13658816.2021.1893323.

[85] H. Wu, A. Lin, K.C. Clarke, W. Shi, A. Cardenas-Tristan, Z. Tu, A comprehensive quality assessment framework for linear features from volunteered geographic information, Int. J. Geogr. Inf. Sci. 35 (9) (2021) 1826–1847, http://dx.doi.org/10.1080/13658816.2020.1832228.

[86] C. Barrington-Leigh, A. Millard-Ball, The world's user-generated road map is more than 80% complete, PLoS One 12 (8) (2017) e0180698 – 20, http://dx.doi.org/10.1371/journal.pone.0180698.

[87] D. Zacharopoulou, A. Skopeliti, B. Nakos, Assessment and visualization of OSM consistency for European cities, ISPRS Int. J. Geo-Inf. 10 (6) (2021) 361, http://dx.doi.org/10.3390/ijgi10060361.

[88] F. Balducci, Is OpenStreetMap a good source of information for cultural statistics? the case of Italian museums, Environ. Plan. B Urban Anal. City Sci. 48 (3) (2019) 503–520, http://dx.doi.org/10.1177/2399808319876949.

[89] J. Yamashita, T. Seto, N. Iwasaki, Y. Nishimura, Quality assessment of volunteered geographic information for outdoor activities: an analysis of OpenStreetMap data for names of peaks in Japan, Geo-Spat. Inf. Sci. (2022) 1–13, http://dx.doi.org/10.1080/10095020.2022.2085188.

[90] Q. Zhou, S. Wang, Y. Liu, Exploring the accuracy and completeness patterns of global land-cover/land-use data in OpenStreetMap, Appl. Geogr. 145 (2022) 102742, http://dx.doi.org/10.1016/j.apgeog.2022.102742.

[91] H. Fan, A. Zipf, Q. Fu, P. Neis, Quality assessment for building footprints data on OpenStreetMap, Int. J. Geogr. Inf. Sci. 28 (4) (2014) 700–719, http://dx.doi.org/10.1080/13658816.2013.867495.

[92] Y. Zhang, Q. Zhou, M.A. Brovelli, W. Li, Assessing OSM building completeness using population data, Int. J. Geogr. Inf. Sci. 36 (7) (2022) 1443–1466, http://dx.doi.org/10.1080/13658816.2021.2023158.

[93] M.A. Brovelli, G. Zamboni, A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints, ISPRS Int. J. Geo-Inf. 7 (8) (2018) 289, http://dx.doi.org/10.3390/ijgi7080289.

[94] Y. Liu, W. Shi, H. Zhang, M. Zhang, A multilevel stratified spatial sampling approach based on terrain knowledge for the quality assessment of OpenStreetMap dataset in Hong Kong, Trans. GIS 27 (1) (2023) 290–318, http://dx.doi.org/10.1111/tgis.13026.

[95] G.M. Foody, L. See, S. Fritz, M. van der Velde, C. Perger, C. Schill, D.S. Boyd, A. Comber, Accurate attribute mapping from volunteered geographic information: Issues of volunteer quantity and quality, Cartogr. J. 52 (4) (2015) 336–344, http://dx.doi.org/10.1080/00087041.2015.1108658.

[96] H. Du, N. Alechina, M. Jackson, G. Hart, A method for matching crowdsourced and authoritative geospatial data, Trans. GIS 21 (2) (2017) 406–427, http://dx.doi.org/10.1111/tgis.12210.

[97] H. Dorn, T. Törnros, A. Zipf, Quality evaluation of VGI using authoritative data—A comparison with land use data in Southern Germany, ISPRS Int. J. Geo-Inf. 4 (3) (2015) 1657–1671, http://dx.doi.org/10.3390/ijgi4031657.

[98] T. Ullah, S. Lautenbach, B. Herfort, M. Reinmuth, D. Schorlemmer, Assessing completeness of OpenStreetMap building footprints using MapSwipe, ISPRS Int. J. Geo-Inf. 12 (4) (2023) 143, http://dx.doi.org/10.3390/ijgi12040143.

[99] S. Borkowska, K. Pokonieczny, Analysis of OpenStreetMap data quality for selected counties in Poland in terms of sustainable development, Sustainability 14 (7) (2022) 3728, http://dx.doi.org/10.3390/su14073728.

[100] R. Hecht, C. Kunze, S. Hahmann, Measuring completeness of building footprints in OpenStreetMap over space and time, ISPRS Int. J. Geo-Inf. 2 (4) (2013) 1066–1091, http://dx.doi.org/10.3390/ijgi2041066.

[101] G. Salvucci, L. Salvati, Official statistics, building censuses, and OpenStreetMap completeness in Italy, ISPRS Int. J. Geo-Inf. 11 (1) (2021) 29, http://dx.doi.org/10.3390/ijgi11010029.

[102] H. Li, B. Herfort, S. Lautenbach, J. Chen, A. Zipf, Improving OpenStreetMap missing building detection using few-shot transfer learning in sub-Saharan Africa, Trans. GIS 26 (8) (2022) 3125–3146, http://dx.doi.org/10.1111/tgis.12941.

[103] Goldblatt, Jones, Mannix, Assessing OpenStreetMap completeness for management of natural disaster by means of remote sensing: A case study of three small island states (Haiti, Dominica and St. Lucia), Remote Sens. 12 (1) (2020) 118, http://dx.doi.org/10.3390/rs12010118.

[104] M. Minghini, F. Frassinelli, OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date? Open Geospatial Data Softw. Stand. 4 (1) (2019) 9, http://dx.doi.org/10.1186/s40965-019-0067-x.

[105] S.S. Sehra, J. Singh, H.S. Rai, Assessing OpenStreetMap data using intrinsic quality indicators: An extension to the QGIS processing toolbox, Future Internet 9 (2) (2017) 15, http://dx.doi.org/10.3390/fi9020015.

[106] M. Minghini, M.A. Brovelli, F. Frassinelli, An open source approach for the intrinsic assessment of the temporal accuracy, up-to-dateness and lineage of OpenStreetMap, ISPRS - Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-4/W8 (2018) 147–154, http://dx.doi.org/10.5194/isprs-archives-xlii-4-w8-147-2018.

[107] K.T. Jacobs, S.W. Mitchell, OpenStreetMap quality assessment using unsupervised machine learning methods, Trans. GIS 24 (5) (2020) 1280–1298, http://dx.doi.org/10.1111/tgis.12680.

[108] C. Barron, P. Neis, A. Zipf, A comprehensive framework for intrinsic OpenStreetMap quality analysis, Trans. GIS 18 (6) (2013) 877–895, http://dx.doi.org/10.1111/tgis.12073.

[109] J. Almendros-Jiménez, A. Becerra-Terón, Analyzing the tagging quality of the spanish OpenStreetMap, ISPRS Int. J. Geo-Inf. 7 (8) (2018) 323, http://dx.doi.org/10.3390/ijgi7080323.

[110] M. Goetz, A. Zipf, OpenStreetMap in 3D – Detailed insights on the current situation in Germany, in: Proceedings of the AGILE'2012 International Conference on Geographic Information Science, 2012, pp. 288–292.

[111] G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C.C. Fonte, V. Antoniou (Eds.), Mapping and the Citizen Sensor, Ubiquity Press, 2017, http://dx.doi.org/10.5334/bbf.

[112] Z. Wang, L. Niu, A data model for using OpenStreetMap to integrate indoor and outdoor route planning, Sensors 18 (7) (2018) 2100, http://dx.doi.org/10.3390/s18072100.

[113] I. Martinez, J.L. Bruse, A.M. Florez-Tapia, E. Viles, I.G. Olaizola, ArchABM: An agent-based simulator of human interaction with the built environment. CO2 and viral load analysis for indoor air quality, Build. Environ. 207 (2022) 108495, http://dx.doi.org/10.1016/j.buildenv.2021.108495.

[114] D. Zielstra, H.H. Hochmair, P. Neis, Assessing the effect of data imports on the completeness of OpenStreetMap – A United States case study, Trans. GIS 17 (3) (2013) 315–334, http://dx.doi.org/10.1111/tgis.12037.

[115] L. Juhász, H.H. Hochmair, OSM data import as an outreach tool to trigger community growth? A case study in Miami, ISPRS Int. J. Geo-Inf. 7 (3) (2018) 113, http://dx.doi.org/10.3390/ijgi7030113.

[116] B. Herfort, S. Lautenbach, J.P.d. Albuquerque, J. Anderson, A. Zipf, The evolution of humanitarian mapping within the OpenStreetMap community, Sci. Rep. 11 (1) (2021) 3037, http://dx.doi.org/10.1038/s41598-021-82404-z.

[117] Y. Feng, X. Huang, M. Sester, Extraction and analysis of natural disaster-related VGI from social media: review, opportunities and challenges, Int. J. Geogr. Inf. Sci. 36 (7) (2022) 1275–1316, http://dx.doi.org/10.1080/13658816.2022.2048835.

[118] A.Y. Grinberger, M. Schott, M. Raifer, A. Zipf, An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap, Trans. GIS 25 (2) (2021) 622–641, http://dx.doi.org/10.1111/tgis.12746.

[119] J. Anderson, D. Sarkar, L. Palen, Corporate editors in the evolving landscape of OpenStreetMap, ISPRS Int. J. Geo-Inf. 8 (5) (2019) 232, http://dx.doi.org/10.3390/ijgi8050232.

[120] A. Stephan, A. Athanassiadis, Quantifying and mapping embodied environmental requirements of urban building stocks, Build. Environ. 114 (2017) 187–202, http://dx.doi.org/10.1016/j.buildenv.2016.11.043.

[121] K. Brassel, F. Bucher, E.-M. Stephan, A. Vckovski, Completeness, in: Elements of Spatial Data Quality, Elsevier, 1995, pp. 81–108, http://dx.doi.org/10.1016/b978-0-08-042432-3.50012-4.

[122] W. Kainz, Logical consistency, in: Elements of Spatial Data Quality, Elsevier, 1995, pp. 109–137, http://dx.doi.org/10.1016/b978-0-08-042432-3.50013-6.

[123] M.F. Goodchild, Attribute accuracy, in: Elements of Spatial Data Quality, Elsevier, 1995, pp. 59–79, http://dx.doi.org/10.1016/b978-0-08-042432-3.50011-2.

[124] C.T. Lloyd, H. Chamberlain, D. Kerr, G. Yetman, L. Pistolesi, F.R. Stevens, A.E. Gaughan, J.J. Nieves, G. Hornby, K. MacManus, P. Sinha, M. Bondarenko, A. Sorichetta, A.J. Tatem, Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets, Big Earth Data 3 (2) (2019) 108–139, http://dx.doi.org/10.1080/20964471.2019.1625151.

[125] M.P. Heris, N.L. Foks, K.J. Bagstad, A. Troy, Z.H. Ancona, A rasterized building footprint dataset for the United States, Sci. Data 7 (1) (2020) 207, http://dx.doi.org/10.1038/s41597-020-0542-3.

[126] M. Wurm, A. Droin, T. Stark, C. Geiß, W. Sulzer, H. Taubenböck, Deep learning-based generation of building stock data from remote sensing for urban heat demand modeling, ISPRS Int. J. Geo-Inf. 10 (1) (2021) 23, http://dx.doi.org/10.3390/ijgi10010023.

[127] Y. Li, Q. Sun, X. Ji, L. Xu, C. Lu, Y. Zhao, Defining the boundaries of urban built-up area based on taxi trajectories: a case study of Beijing, J. Geovisualization Spat. Anal. 4 (1) (2020) http://dx.doi.org/10.1007/s41651-020-00047-6.

[128] M. Varentsov, T. Samsonov, M. Demuzere, Impact of urban canopy parameters on a Megacity's modelled thermal environment, Atmosphere 11 (12) (2020) 1349, http://dx.doi.org/10.3390/atmos11121349.

[129] G.J. Bruyns, C.D. Higgins, D.H. Nel, Urban volumetrics: From vertical to volumetric urbanisation and its extensions to empirical morphological analysis, Urban Stud. 58 (5) (2021) 922–940, http://dx.doi.org/10.1177/0042098020936970.

[130] L. Cheng, F. Zhang, S. Li, J. Mao, H. Xu, W. Ju, X. Liu, J. Wu, K. Min, X. Zhang, M. Li, Solar energy potential of urban buildings in 10 cities of China, Energy 196 (2020) 117038, http://dx.doi.org/10.1016/j.energy.2020.117038.

[131] H. Usui, Comparison of precise and approximated building height: Estimation from number of building storeys and spatial variations in the Tokyo metropolitan region, Environ. Plan. B Urban Anal. City Sci. 50 (2023) 487–499, http://dx.doi.org/10.1177/23998083221116117.

[132] Y. Liu, C. Chen, J. Li, W.-Q. Chen, Characterizing three dimensional (3-D) morphology of residential buildings by landscape metrics, Landsc. Ecol. 35 (11) (2020) 2587–2599, http://dx.doi.org/10.1007/s10980-020-01084-8.

[133] R. Boeters, K. Arroyo Ohori, F. Biljecki, S. Zlatanova, Automatically enhancing CityGML LOD2 models with a corresponding indoor geometry, Int. J. Geogr. Inf. Sci. 29 (12) (2015) 2248–2268, http://dx.doi.org/10.1080/13658816.2015.1072201.

[134] R. Buffat, A. Froemelt, N. Heeren, M. Raubal, S. Hellweg, Big data GIS analysis for novel approaches in building stock modelling, Appl. Energy 208 (2017) 277–290, http://dx.doi.org/10.1016/j.apenergy.2017.10.041.

[135] X. Yang, M. Hu, N. Heeren, C. Zhang, T. Verhagen, A. Tukker, B. Steubing, A combined GIS-archetype approach to model residential space heating energy: A case study for the Netherlands including validation, Appl. Energy 280 (2020) 115953, http://dx.doi.org/10.1016/j.apenergy.2020.115953.

[136] B. Yu, H. Liu, J. Wu, Y. Hu, L. Zhang, Automated derivation of urban building density information using airborne LiDAR data and object-based method, Landsc. Urban Plan. 98 (3–4) (2010) 210–219, http://dx.doi.org/10.1016/j.landurbplan.2010.08.004.

[137] A. Janowski, M. Renigier-Biłozor, M. Walacik, A. Chmielewska, Remote measurement of building usable floor area – Algorithms fusion, Land Policy 100 (2021) 104938, http://dx.doi.org/10.1016/j.landusepol.2020.104938.

[138] L.Y. Gaw, S. Chen, Y.S. Chow, K. Lee, F. Biljecki, Comparing street view imagery and aerial perspectives in the built environment, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-4/W3-2022 (2022) 49–56, http://dx.doi.org/10.5194/isprs-annals-X-4-W3-2022-49-2022.

[139] J. von Platten, C. Sandels, K. Jörgensson, V. Karlsson, M. Mangold, K. Mjörnell, Using machine learning to enrich building databases—Methods for tailored energy retrofits, Energies 13 (10) (2020) 2574, http://dx.doi.org/10.3390/en13102574.

[140] S. Zou, L. Wang, Mapping individual abandoned houses across cities by integrating VHR remote sensing and street view imagery, Int. J. Appl. Earth Obs. Geoinf. 113 (2022) 103018, http://dx.doi.org/10.1016/j.jag.2022.103018.

[141] H. Fan, G. Kong, C. Zhang, An interactive platform for low-cost 3D building modeling from VGI data using convolutional neural network, Big Earth Data 5 (1) (2021) 49–65, http://dx.doi.org/10.1080/20964471.2021.1886391.

[142] Y. Yan, B. Huang, Estimation of building height using a single street view image via deep neural networks, ISPRS J. Photogramm. Remote Sens. 192 (2022) 83–98, http://dx.doi.org/10.1016/j.isprsjprs.2022.08.006.

[143] M. Sun, C. Han, Q. Nie, J. Xu, F. Zhang, Q. Zhao, Understanding building energy efficiency with administrative and emerging urban big data by deep learning in glasgow, Energy Build. 273 (2022) 112331, http://dx.doi.org/10.1016/j.enbuild.2022.112331.

[144] M. Sun, F. Zhang, F. Duarte, C. Ratti, Understanding architecture age and style through deep learning, Cities 128 (2022) 103787, http://dx.doi.org/10.1016/j.cities.2022.103787.

[145] J. Kang, M. Körner, Y. Wang, H. Taubenböck, X.X. Zhu, Building instance classification using street view images, ISPRS J. Photogramm. Remote Sens. 145 (2018) 44–59, http://dx.doi.org/10.1016/j.isprsjprs.2018.02.006, arXiv:1802.09026.

[146] S.P. Ramalingam, V. Kumar, Automatizing the generation of building usage maps from geotagged street view images using deep learning, Build. Environ. 235 (2023) 110215, http://dx.doi.org/10.1016/j.buildenv.2023.110215.

[147] H.E. Pang, F. Biljecki, 3D building reconstruction from single street view images using deep learning, Int. J. Appl. Earth Obs. Geoinf. 112 (2022) 102859, http://dx.doi.org/10.1016/j.jag.2022.102859.

[148] C. León-Sánchez, G. Agugiaro, J. Stoter, Creation of a CityGML-based 3D city model testbed for energy-related applications, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLVIII-4/W5-2022 (2022) 97–103, http://dx.doi.org/10.5194/isprs-archives-xlviii-4-w5-2022-97-2022.

[149] P. Tobiáš, J. Cajthaml, Models of cultural heritage buildings in a procedurally generated geospatial environment, Trans. GIS 25 (2) (2021) 1104–1122, http://dx.doi.org/10.1111/tgis.12727.

[150] W. Pei, F. Biljecki, R. Stouffs, Dataset for urban scale building stock modelling: Identification and review of potential data collection approaches, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-4/W2-2022 (2022) 225–232, http://dx.doi.org/10.5194/isprs-annals-X-4-W2-2022-225-2022.

[151] A. Leonard, S. Wheeler, M. McCulloch, Power to the people: Applying citizen science and computer vision to home mapping for rural energy access, Int. J. Appl. Earth Obs. Geoinf. 108 (2022) 102748, http://dx.doi.org/10.1016/j.jag.2022.102748.

[152] N. Milojevic-Dupont, F. Wagner, F. Nachtigall, J. Hu, G.B. Brüser, M. Zumwald, F. Biljecki, N. Heeren, L.H. Kaack, P.-P. Pichler, F. Creutzig, EUBUCCO v0.1: European building stock characteristics in a common and open database for 200+ million individual buildings, Sci. Data 10 (1) (2023) 147, http://dx.doi.org/10.1038/s41597-023-02040-2.

[153] E. Prataviera, J. Vivian, G. Lombardo, A. Zarrella, Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis, Appl. Energy 311 (2022) 118691, http://dx.doi.org/10.1016/j.apenergy.2022.118691.

[154] C. Ellul, M. Adjrad, P. Groves, The impact of 3D data quality on improving GNSS performance using city models initial simulations, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. IV-2-W1 (2016) 171–178, http://dx.doi.org/10.5194/isprs-annals-iv-2-w1-171-2016.

[155] Y. Li, A.J. Brimicombe, M.P. Ralphs, Spatial data quality and sensitivity analysis in GIS and environmental modelling: the case of coastal oil spills, Comput. Environ. Urban Syst. 24 (2) (2000) 95–108, http://dx.doi.org/10.1016/s0198-9715(99)00048-4.

[156] J. Beekhuizen, G.B.M. Heuvelink, A. Huss, A. Bürgi, H. Kromhout, R. Vermeulen, Impact of input data uncertainty on environmental exposure assessment models: A case study for electromagnetic field modelling from mobile phone base stations, Environ. Res. 135 (2014) 148–155, http://dx.doi.org/10.1016/j.envres.2014.05.038.

[157] H.A.S. Othman, A.A. Alshboul, The role of urban morphology on outdoor thermal comfort: The case of Al-Sharq City – Az Zarqa, Urban Clim. 34 (2020) 100706, http://dx.doi.org/10.1016/j.uclim.2020.100706.

[158] F. Biljecki, H. Ledoux, J. Stoter, Generating 3D city models without elevation data, Comput. Environ. Urban Syst. 64 (2017) 1–18, http://dx.doi.org/10.1016/j.compenvurbsys.2017.01.001.

[159] C.C. Fonte, M. Minghini, V. Antoniou, J. Patriarca, L. See, Classification of building function using available sources of VGI, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-4 (2018) 209–215, http://dx.doi.org/10.5194/isprs-archives-xlii-4-209-2018.

[160] E. Roy, M. Pronk, G. Agugiaro, H. Ledoux, Inferring the number of floors for residential buildings, Int. J. Geogr. Inf. Sci. (2022) 1–25, http://dx.doi.org/10.1080/13658816.2022.2160454.

[161] M. Kutrzyński, Z. Telec, B. Trawiński, H.C. Dac, An approach to estimation of residential housing type based on the analysis of parked cars, in: Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science, Vol. 11432, fourth ed., Springer Berlin Heidelberg, 2019, pp. 280–289.

[162] K.S. Atwal, T. Anderson, D. Pfoser, A. Züfle, Predicting building types using OpenStreetMap, Sci. Rep. 12 (1) (2022) 19976, http://dx.doi.org/10.1038/s41598-022-24263-w.

[163] K. Hopf, Mining volunteered geographic information for predictive energy data analytics, Energy Inform. 1 (1) (2018) 1–21, http://dx.doi.org/10.1186/s42162-018-0009-3.

[164] X. Chen, F. Biljecki, Mining real estate ads and property transactions for building and amenity data acquisition, Urban Inform. 1 (2022) 12, http://dx.doi.org/10.1007/s44212-022-00012-2.

[165] G. Manoli, S. Fatichi, M. Schläpfer, K. Yu, T.W. Crowther, N. Meili, P. Burlando, G.G. Katul, E. Bou-Zeid, Magnitude of urban heat islands largely explained by climate and population, Nature 573 (7772) (2019) 55–60, http://dx.doi.org/10.1038/s41586-019-1512-9.

[166] M. Li, E. Koks, H. Taubenböck, J. van Vliet, Continental-scale mapping and analysis of 3D building structure, Remote Sens. Environ. 245 (2020) 111859, http://dx.doi.org/10.1016/j.rse.2020.111859.

[167] Y. Li, S. Schubert, J.P. Kropp, D. Rybski, On the influence of density and morphology on the urban heat Island intensity, Nature Commun. 11 (1) (2020) 2647, http://dx.doi.org/10.1038/s41467-020-16461-9.

[168] W. Lin, Volunteered geographic information constructions in a contested terrain: A case of OpenStreetMap in China, Geoforum 89 (2018) 73–82, http://dx.doi.org/10.1016/j.geoforum.2018.01.005.

[169] C. Bittner, OpenStreetMap in Israel and palestine – 'Game changer' or reproducer of contested cartographies? Political Geogr. 57 (2017) 34–48, http://dx.doi.org/10.1016/j.polgeo.2016.11.010.

[170] Q. Zhou, Y. Zhang, K. Chang, M.A. Brovelli, Assessing OSM building completeness for almost 13,000 cities globally, Int. J. Digit. Earth 15 (1) (2022) 2400–2421, http://dx.doi.org/10.1080/17538947.2022.2159550.

[171] B. Herfort, S. Lautenbach, J.P. de Albuquerque, J. Anderson, A. Zipf, Investigating the digital divide in OpenStreetMap: spatio-temporal analysis of inequalities in global urban building completeness, 2022, http://dx.doi.org/10.21203/rs.3.rs-1913150/v1.

[172] C. Miller, P. Arjunan, A. Kathirgamanathan, C. Fu, J. Roth, J.Y. Park, C. Balbach, K. Gowri, Z. Nagy, A.D. Fontanini, J. Haberl, The ASHRAE great energy predictor III competition: Overview and results, Sci. Technol. Built Environ. 26 (10) (2020) 1427–1447, http://dx.doi.org/10.1080/23744731.2020.1795514.

[173] J. Picaut, N. Fortin, E. Bocher, G. Petit, P. Aumond, G. Guillaume, An open-science crowdsourcing approach for producing community noise maps using smartphones, Build. Environ. 148 (2019) 20–33, http://dx.doi.org/10.1016/j.buildenv.2018.10.049.

[174] J.P. Wilson, K. Butler, S. Gao, Y. Hu, W. Li, D.J. Wright, A five-star guide for achieving replicability and reproducibility when working with GIS software and algorithms, Ann. Am. Assoc. Geogr. 111 (5) (2021) 1–7, http://dx.doi.org/10.1080/24694452.2020.1806026.