

Continental-scale mapping and analysis of 3D building structure

Mengmeng Li^{a,*}, Elco Koks^a, Hannes Taubenböck^{b,c}, Jasper van Vliet^a



^a Institute for Environmental Studies, VU University Amsterdam, De Boelelaan 1085, 1081, HV, Amsterdam, the Netherlands

^b German Aerospace Center, German Remote Sensing Data Center, Oberpfaffenhofen 82234, Germany

^c Institute for Geography and Geology, Julius-Maximilians-Universität Würzburg, Würzburg 97074, Germany

ARTICLE INFO

Keywords:

Urban form
Land use intensity
Building height
Urban density
SAR

ABSTRACT

Urban land use is often characterized based on the presence of built-up land, while the land use intensity of different locations is ignored. This narrow focus is at least partially due to a lack of data on the vertical dimension of urban land. The potential of Earth observation data to fill this gap has already been shown, but this has not yet been applied at large spatial scales. This study aims to map urban 3D building structure, i.e. building footprint, height, and volume, for Europe, the US, and China using random forest models. Our models perform well, as indicated by R^2 values of 0.90 for building footprint, 0.81 for building height, and 0.88 for building volume, for all three case regions combined. In our multidimensional input variables, we find that built-up density derived from the Global Urban Footprint (GUF) is the most important variable for estimating building footprint, while backscatter intensity of Synthetic Aperture Radar (SAR) is the most important variable for estimating building height. A combination of the two is essential to estimate building volume. Our analysis further highlights the heterogeneity of 3D building structure across space. Specifically, buildings in China tend to be taller on average (10.35 m) compared to Europe (7.37 m) and the US (6.69 m). At the same time, the building volume per capita in China is lowest, with 302.3 m^3 per capita, while Europe and the US show estimates of 404.6 m^3 and 565.4 m^3 , respectively. The results of this study (3D building structure data for Europe, the US, and China) are publicly available, and can be used for further analysis of urban environment, spatial planning, and land use projections.

1. Introduction

Urban development is manifested differently in different world regions, both horizontally and vertically. For example, Singapore has built numerous high-rise and compact apartments to accommodate its growing population (Grace Wong, 2004). A recent study on selected cities finds that urban development in the United States is dominated by decentralized-sprawl patterns, while central-compact patterns are typically found in Europe and China (Dong et al., 2019). Moreover, urban expansion in the Global South is often characterized by the proliferation of low-rise slums (Badmos et al., 2018; Kusno, 2019; Wang et al., 2019a). The structure of urban areas has large impacts on both the biophysical and socioeconomic conditions of urban areas (Connors et al., 2012; Engelfriet and Koomen, 2017; Hudeček et al., 2019). For example, compact urban structure contributes to reducing greenhouse gas (GHG) emissions on the one hand (e.g. Glaeser (2011)), but it could also worsen the urban environment through the urban heat island effect on the other hand (Berger et al., 2017). Other studies have shown the impacts of urban structure on landscape aesthetics, urban climate,

health aspects, or energy consumption (Güneralp et al., 2017; Lin et al., 2018; Miles et al., 2012; Stewart and Oke, 2012), among others.

Urban structure involves both the horizontal and vertical configurations of urban land and infrastructure (Wentz et al., 2018). Monitoring the horizontal aspect, i.e. urban extent, has been prominent in earth-observation studies for decades, resulting in various products available from local to global scales (Carlson and Sanchez-Azofeifa, 1999; Gong et al., 2020; Mertes et al., 2015; Schneider et al., 2009; Taubenböck et al., 2012). These urban extent products are crucial for environmental assessments to address sustainability challenges such as food insecurity, biodiversity loss, and risk exposure (Angel et al., 2011; Du et al., 2018; van Vliet, 2019). Moreover, urban extent products have also been used for better characterization of the terrestrial biosphere, for instance using landscape mosaics, anthropomes, and land systems (Ellis and Ramankutty, 2008; Messerli et al., 2009; van Asselen and Verburg, 2012). However, because the impacts of different types of urban development vary, there is a need to characterize urban development beyond two-dimensional spatial patterns, in order to comprehensively assess urban sustainability. To date, only a few studies have

* Corresponding author.

E-mail address: mengmeng.li@vu.nl (M. Li).

analyzed the vertical dimension of urban structure, either at a small scale (He et al., 2016; Kedron et al., 2019), or for selected (mega)cities across the globe (e.g. Frolking et al. (2013), Straka and Sodoudi (2019), and Zhang et al. (2018)).

The significance and urgency of mapping urban structure in the horizontal as well as in the vertical dimension (hereafter referred to as the 3D building structure) are further highlighted in a recent review on urban remote sensing (Zhu et al., 2019). Yet, compared to the identification of building extent, retrieval of a building's vertical profile based on remote sensing is a more complex process. There is a growing body of literature trying to extract building height (Bagheri et al., 2018; Liasis and Stavrou, 2016; Weissgerber et al., 2017), but most of them are devoted to local and regional scales. A large number of remote sensing based data sources are available to retrieve building height, which can generally be categorized into four categories: conventional optical images, stereo optical images, Light Detection And Ranging (LiDAR), and Radar. LiDAR is widely acknowledged as the most robust source. However, applications of LiDAR-derived data are highly constrained by their coverage, as data are scarce, expensive, and scattered. Recently, hybrid data have been used to characterize 3D building structure. For example, Geiß et al. (2019) present a multistep approach to estimate 3D building structure based on TanDEM-X and optical Sentinel-2 data. Nonetheless, large-scale (or even continental-scale) estimates of 3D building structure are thus far still lacking. Case studies of building volume estimates derived from LiDAR and Radar (both scatterometer and SAR) respectively reveal that these two source datasets are highly consistent (Bagheri et al., 2018; Mathews et al., 2019), suggesting that current fine-resolution Radar data could contribute to the estimation of 3D building structure at a larger scale.

From a land use perspective, the combination of horizontal and vertical urban structure can be considered as an expression of urban land use intensity. Urban land use intensity can be interpreted as the equivalent of agricultural land use intensity, as it expresses the density or intensity of the use of agricultural land in a location. Consistently, urban land use intensity can be characterized in different ways, and it is not clear a priori what measure is preferable (see e.g. Kuemmerle et al. (2013) for a discussion on quantifying agricultural land use intensity, and Dovey and Pafka (2013) for a discussion on measuring urban density). Recent studies for example include population density (van Vliet et al., 2019), or a spatial characterization of urban structure (Susaki et al., 2014; Xia et al., 2020). This study aims to complement these data by developing the first continental-scale data on 3D building structure, i.e. building footprint, height, and volume, where continental-scale refers to complete continents, like Europe, or areas that are comparable in size, like the US and China. Based on the reference data collected from various sources, we train random forest models to estimate 3D building structure using a large number of explanatory variables. In the following, Section 2 describes the methodological approach for mapping 3D building structure in more detail. Section 3 presents the results of these models, as well as an analysis of how building structure differs between our study regions and an elaborate analysis of the model accuracy and uncertainty. In Section 4 we further discuss these results, and reflect on the contribution of these data for sustainable settlement development.

2. Materials and method

2.1. Overview

In this study we estimate building footprint, building height, and building volume at a 1 km² resolution for Europe, the US, and China. The US and China refer to the conterminous United States and mainland China (including Hong Kong and Macao), respectively. We choose a 1 km² resolution because the aim of this study is to characterize urban areas as a land use type, which can be used for further analysis of land use, land use changes, as well as their impacts. As a result, we do not

characterize individual buildings, but instead focus on the characterization of the general building structure within larger spatial units (pixels), which can be considered as a characterization of urban land use intensity. For these analyses individual buildings are of little interest as the related phenomena act at a coarser scale (e.g. van Vliet et al. (2019), Stewart and Oke (2012), and Wang et al. (2019b)). Building footprint denotes the share of each 1 km² pixel that is occupied by buildings (therefore expressed as m² per m²). We use the term *building footprint* rather than *building density*, because *building density* has also been used to denote the building floor space per unit area thus including vertical aspects as well, and we want to avoid such confusion. Building height denotes the average height of all buildings in a pixel, weighted by the area of each building. Building volume is the total volume within each pixel taken by buildings. Conceptually, building volume is the building footprint multiplied by the average building height in a pixel, although all three properties are predicted independently in our study.

We train random forest models to estimate building footprint, height, and volume using reference data for different locations in the study areas. We subsequently use these trained models to estimate 3D building structure based on the same variables for all other locations within our study areas. These study areas are Europe, the US, and China, which were selected based on the availability of reference data. Fig. 1 illustrates the overall approach of our study. This approach consists of four parts: 1) the collection and preprocessing of spatial data that are used as *explanatory variables* in our models, using the Google Earth Engine (GEE). GEE is a cloud-based platform for geospatial analysis at a planetary scale, which also consists of various ready-to-use datasets, co-located within a high-performance, intrinsically parallel computation service (Gorelick et al., 2017); 2) collection and preprocessing of *reference data*, including both readily available 3D building data and manual interpretation of 3D building structure based on Very High Resolution (VHR) satellite/aerial imagery and street view imagery; 3) training, optimizing, and validating *random forest models* to produce maps of 3D building structure; 4) *spatial analysis of building properties* in the three study regions and the differences between these regions.

2.2. Spatial data for explanatory variables

We estimate 3D building structure using a large number of spatial data sets as explanatory variables. These variables are selected based on four criteria: *First*, they should be expected to provide information on building height. *Second*, the data for each variable should be close to the year 2015, for temporal consistency, as the uncertainty increases when data is recorded further away from the dates at which the reference data was collected. *Third*, data for each variable should be available for all three regions, thus allowing cross-region comparison. In practice, this means we used datasets with a global coverage. *Fourth*, the data for each variable should be based on direct measurements rather than being downsampled, to ensure independence. We further group explanatory variables into four classes according to their sources or imaging modes: *optical RS*, *SAR*, *RS-derived*, and *others*.

Optical RS data include all available spectral bands representing surface reflectance from Landsat 8 for the year 2015, covering Europe and the US under cloud-free conditions. We include optical RS data, because previous studies have shown that reflectance values reveal information on the urban environment (Lee and Kim, 2013; Yuan and Bauer, 2007). For China, the whole territory is not fully covered in 2015 only, thus we include data for the period 2014–2016. As shown in Fig. 2, for each Landsat band, we first compute the median of all cloud-free and shadow-free images for each pixel at the original resolution, to generate the representative values for this period and to exclude extreme values. Consequently, we make a spatial aggregation of corresponding representative values into 1-km cells using a mean function.

We use 10-meter resolution Sentinel-1 SAR images, which ideally

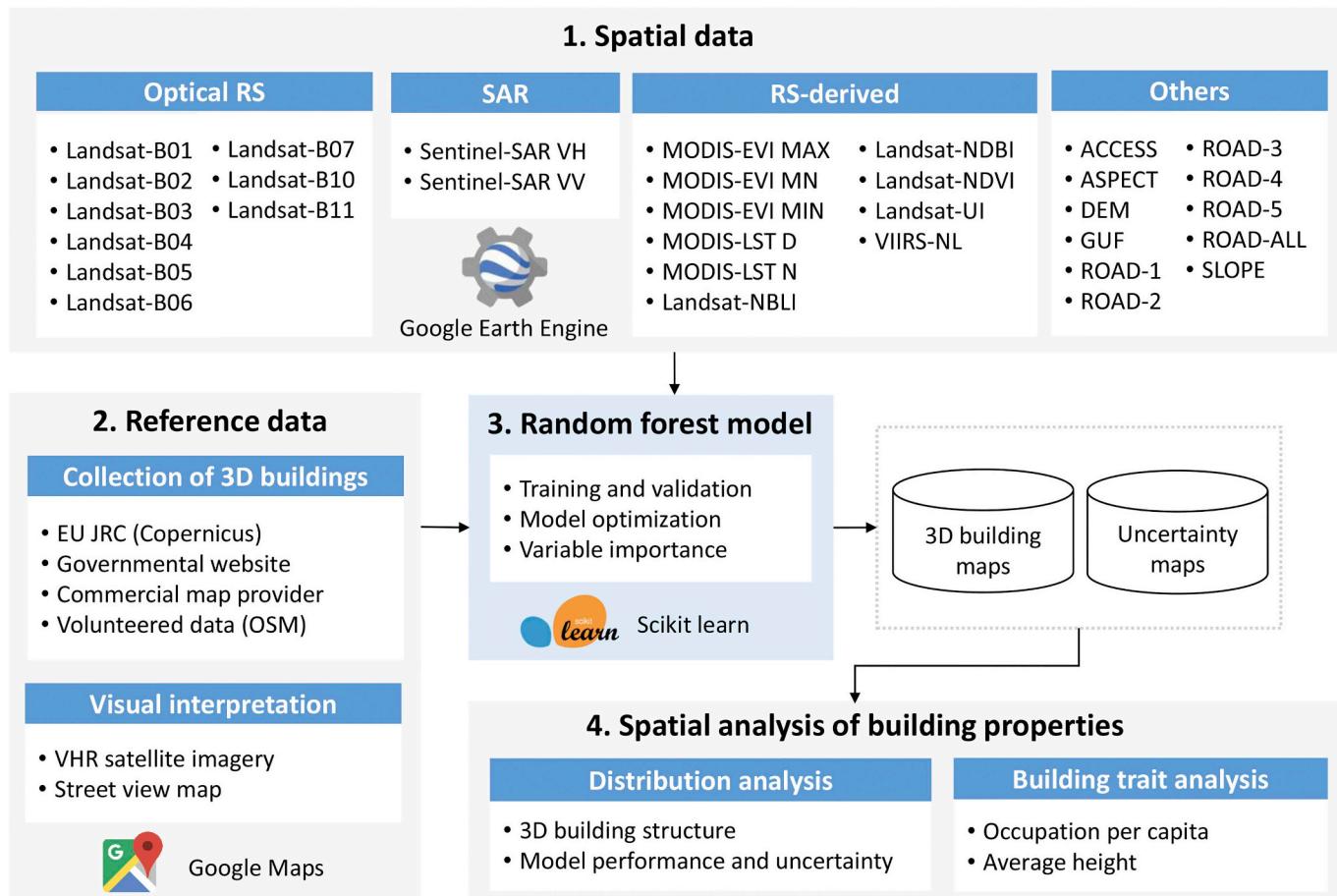


Fig. 1. Methodological flowchart for the estimation of 3D building structure. The abbreviations for spatial data are referred to in Table 1.

have a global coverage for every 12 days (Malenovský et al., 2012). SAR data are responsive to surface roughness and therefore we expect that these data are especially relevant as explanatory variable for building height and volume (Brunner et al., 2010; Li et al., 2020; Soergel et al., 2009). Besides buildings, other objects such as trees are also sensitive to backscatter coefficients (x_{bc}) of SAR. Therefore, we selected SAR images during two winter seasons around the year 2015, i.e. 1st December

2014–31st March 2015, and 1st December 2015–31st March 2016. However, we added information from adjacent years in areas that were not fully covered by the data from the winters in 2015. We do not differentiate between orbit directions, i.e. ascending or descending, as exploratory data analysis reveals that our case study regions are not fully covered within one single orbit direction. All available SAR images are processed, calibrated, and geo-rectified with the Sentinel-1 Toolbox

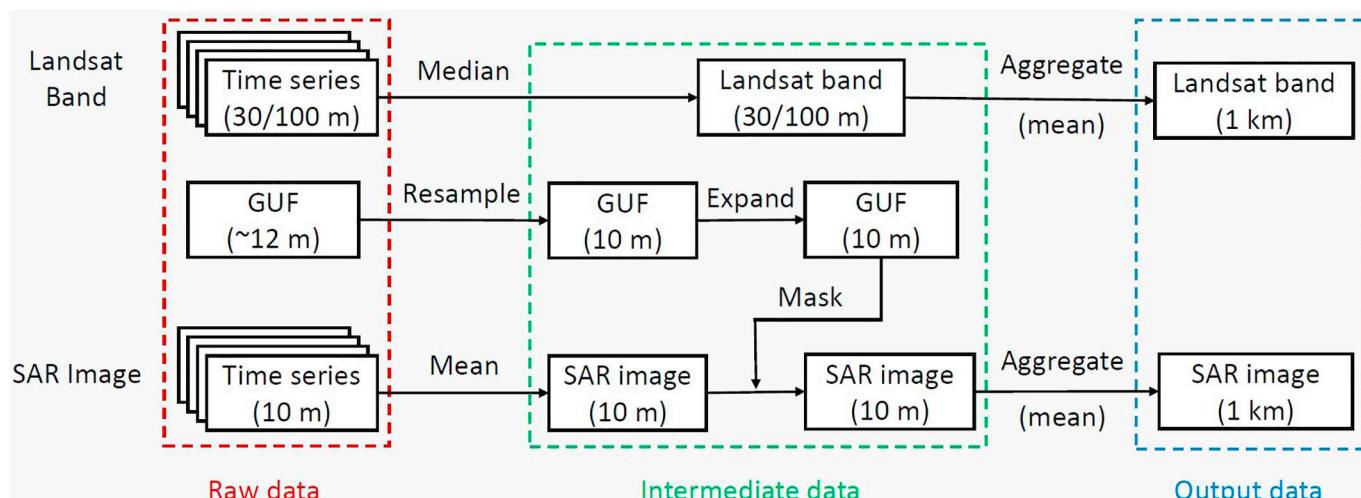


Fig. 2. Reduction and aggregation of time-series cloud-free Landsat and Sentinel-1 SAR data. Note: SAR data provided in GEE is log-scaled, we transform the scaled SAR into backscatter coefficient before further operations are applied. Algorithms used to reduce time-series of Landsat and SAR collections are suggested by GEE officials (<https://developers.google.com/earth-engine/>), but in further aggregation operation we additionally mask SAR cell values based on the expanded GUF.

(ESA, 2019). As illustrated in Fig. 2, for each 10-m pixel we firstly average all backscatter coefficients (x_{bc}) available in the study period for VV (vertical transmission and vertical reception) and VH (vertical transmission and horizontal reception) polarization modes separately, and then the averaged x_{bc} for each image is aggregated to a 1×1 km resolution using the mean of only the values within the built environment. For this spatial restriction we use the built environment as mapped by the Global Urban Footprint (GUF) (Esch et al., 2017). We use this as a mask to reduce the influence of objects such as trees and topographic relief outside the built environment. In addition, our exploratory data analysis shows that a large number of buildings (especially the higher ones) are displaced due to the side-looking SAR measurement, thus the GUF mask is buffered with a distance of 2 pixels, i.e., 20 m.

RS-derived data consist of Enhanced Vegetation Index (EVI), land surface temperature (LST), relevant built-up indices derived from Landsat, and nighttime light intensity (VIIRS). We expect that vegetation indices could be relevant explanatory variables as they correct for surface roughness recorded in SAR data caused by vegetation, while we expect that the other indices (VIIRS, LST, and UI) provide information about building structure themselves, as they have been used to characterize urban land use intensity in previous studies (Ma et al., 2014; Wellmann et al., 2018; Zhang and Huang, 2015). EVI is available for every 16-day period from MODIS products. We process all these data throughout the year 2015 into three variables using maximum, mean, and minimum functions separately. LST data are provided by the MOD11A2 V6 product, which is a simple average of all the corresponding MOD11A1 LST cells collected within every 8-day period, where daytime and nighttime are independently stored (Wan et al., 2015). We average all LST data throughout the study period for daytime and nighttime, respectively. Normalized Difference Built-up Index (NDBI), Normalized Difference Bare Land index (NBLI), Normalized Difference Vegetation Index (NDVI) and Urban Index (UI) are also used as the explanatory variables, which are all derived from Landsat images. To have a systematic understanding of these indices, readers are referred to Mushore et al. (2017). Nighttime light intensity data are derived from stray-light corrected VIIRS nighttime light (Butler et al., 2013), which are provided as monthly composites at 500-m scale. We combine all these monthly data available in GEE for the year 2015 into annual nighttime light intensity using a maximum function, and spatially aggregate them into 1-km data using an average function. The maximum function is used to remove cloud shadow effects in night light images. Since other light sources such as wildfire and water bodies reflecting moonlight or anthropogenic light can appear in non-built-up area, we also apply the GUF mask in order to exclude these areas.

In addition to remote sensing imagery, we use a series of other data, including urban footprint, accessibility, roads, and topography as explanatory variables (see Table 1). We expect that urban footprint, road density, and accessibility could indirectly provide information on urban land use intensity, while we expect that topography could provide a correction on the signal from SAR backscatter, because also topography could lead to surface roughness recorded in SAR data (van der Wal et al., 2005). Urban footprint recorded in the GUF shows impervious surface which we expect to relate strongly with building footprint. Built-up density is calculated based on the GUF, which is a global binary settlement layer created by the German Aerospace Center using satellite images from TerraSAR-X and TanDEM-X (Esch et al., 2013). Data from these sensors are not included in our model otherwise, in order to avoid double-counting or circularity. Based on a comparison of estimates for Central Europe, GUF comes out as the most reliable map of urban extent datasets in terms of resolution and accuracy (Klotz et al., 2016). However, it is generated using images during 2011–2013. We assume that other explanatory variables for the year 2015 and the short time interval are sufficient to compensate such defect. Accessibility-to-cities data by Weiss et al. (2018) represent land-based travel time to the nearest densely-populated area for the nominal year 2015.

Vector road data from Meijer et al. (2018) are used to generate five hierarchical road density maps including highways, primary roads, secondary roads, tertiary roads, and local roads. In addition, we also add a density map for all roads, which embodies unclassified roads. DEM, slope, and aspect are all derived from Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010).

2.3. Reference data

Reference data are collected using publicly available datasets from various sources for the three case study areas. Specifically, for Europe we use gridded building height data of 25 cities (<https://land.copernicus.eu/>), representing the year 2012, complemented by building footprint layers from OpenStreetMap (OSM, access date: 11 January 2019). To reduce the negative effects caused by null values in building height data, we only consider areas where the proportion of buildings with valid height values exceeds 80% of all the building footprint area. This threshold is set to exclude locations where a large share of buildings has been built after the gridded building height data have been produced. For the US, we employ data that are publicly available from the websites of local governments for the nominal year 2015 (including occasional updates published in the ArcGIS Hub <http://hub.arcgis.com/>, see Table S2 for details). These datasets demonstrate vector building footprints with vertical properties for 27 urban areas. These datasets include areas ranging from megacities like New York and Los Angeles to counties that only include small villages in remote areas. Thus, these datasets include the full variability with respect to the combination of building footprint and building height. Building height data for China, expressed as floor number, are available for 24 selected large cities nominally for the year 2015 (<https://www.amap.com>). In this paper, for all building height expressed as floor numbers, we assume that each floor is 3-m high (Leichtle et al., 2019; Zhou et al., 2014). It is worth noting that relatively low model performance was observed for China in our preliminary evaluation, which was ultimately explained by a substantial number of missing buildings in some areas of Chinese cities when compared with VHR satellite imagery from Google Maps. Therefore, we removed all the data points (i.e. 1 km² pixels) that we suspected contained such omissions.

Available reference data is biased towards large urban regions. Therefore, we complement these data with empirical data for smaller settlements, which are classified manually. For this, we use Google Maps Static API to randomly download VHR satellite images outside large urban regions (travel time to cities > 10 min, built-up density > 0). Each image represents a 1×1 km landscape at a 0.25 m resolution, which we assume is sufficiently detailed for building footprint detection. During the visual interpretation process, a fishnet layer with 50×50 regular squared grids is used for specifying grid numbers, as well as Google Street View for the estimation of building height. These estimations are all based on visual interpretation of VHR satellite images and streetscapes provided by Google Maps. Together, from all 1×1 km grid cells that contain built-up land according to the GUF, we randomly select 1146 images from the US, 2573 images from Europe, and 2445 images from China to complement our reference data. Because of the scarcity of street view maps in mainland China, building height is not estimated manually there. We exclude images that are invalid due to high cloud coverage or image inaccessibility. See Fig. S1 for an example of valid imageries. The methodology for estimating building height is further illustrated in Fig. S2 and Table S1. For other locations where no street view map exists, we specify building height by interpreting similar adjacent places where street view maps are available. In total, our reference data contained data for 55,656, 47,639, and 47,553 pixels of 1×1 km for building footprint, height, and volume, respectively (Table 2).

To examine the reliability of our visual interpretation approach, we digitize building footprints based on 100 randomly selected VHR images. Because of the high amount of detail in this VHR imagery

Table 1

Spatial data used as explanatory variables in this study.

| Datasets | | Original resolution | Time | Source | Abbreviation |
|------------|----------------------|---------------------|-------|---|-----------------|
| Optical RS | Landsat Band 1 | 30 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B01 |
| | Landsat Band 2 | 30 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B02 |
| | Landsat Band 3 | 30 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B03 |
| | Landsat Band 4 | 30 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B04 |
| | Landsat Band 5 | 30 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B05 |
| | Landsat Band 6 | 30 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B06 |
| | Landsat Band 7 | 30 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B07 |
| | Landsat Band 10 | 100 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B10 |
| | Landsat Band 11 | 100 m | 2015 | https://landsat.gsfc.nasa.gov/ | Landsat-B11 |
| | Sentinel-1 VH | 10 m | 2015 | https://sentinel.esa.int/ | Sentinel-SAR VH |
| | Sentinel-1 VV | 10 m | 2015 | https://sentinel.esa.int/ | Sentinel-SAR VV |
| RS-derived | EVI maximum | 1 km | 2015 | MODIS/006/MYD13A2 | MODIS-EVI MAX |
| | EVI mean | 1 km | 2015 | MODIS/006/MYD13A2 | MODIS-EVI MN |
| | EVI minimum | 1 km | 2015 | MODIS/006/MYD13A2 | MODIS-EVI MIN |
| | LST day | 1 km | 2015 | MOD11A2 | MODIS-LST D |
| | LST night | 1 km | 2015 | MOD11A2 | MODIS-LST N |
| | NBLI | 30 m | 2015 | Landsat | Landsat-NBLI |
| Others | NDBI | 30 m | 2015 | Landsat | Landsat-NDBI |
| | NDVI | 30 m | 2015 | Landsat | Landsat-NDVI |
| | UI | 30 m | 2015 | Landsat | Landsat-UI |
| | Night time light | 1 km | 2015 | VIIRS | VIIRS-NL |
| | Accessibility | 1 km | 2015 | Weiss et al. (2018) | ACCESS |
| | Aspect | ~250 m | 2010 | GMTED2010 | ASPECT |
| | DEM | ~250 m | 2010 | GMTED2010 | DEM |
| | GUF | ~12 m | ~2012 | German Aerospace Center (DLR) | GUF |
| | Highways | Vector | ~2015 | Meijer et al. (2018) | ROAD-1 |
| | Primary roads | Vector | ~2015 | Meijer et al. (2018) | ROAD-2 |
| Others | Secondary roads | Vector | ~2015 | Meijer et al. (2018) | ROAD-3 |
| | Tertiary roads | Vector | ~2015 | Meijer et al. (2018) | ROAD-4 |
| | Local roads | Vector | ~2015 | Meijer et al. (2018) | ROAD-5 |
| | All (+ unclassified) | Vector | ~2015 | Meijer et al. (2018) | ROAD-ALL |
| | Slope | ~250 m | 2010 | GMTED2010 | SLOPE |

relative to the information that is coded, and because these data are collected independently from the RF models, it is found acceptable for generating reference data. The comparison shows very high reliability (see Fig. S3). Abandoned buildings and temporary structures are all included, due to the fact that we are not able to differentiate building types for specific purposes from Google Earth images. As a consequence, the total building footprint area provided here could exceed the actual footprint of 'permanent buildings' or 'under roof' measurements published elsewhere. As shown in studies testing positional accuracy of Google Earth images (Mohammed et al., 2013; Pulighe et al., 2015), error in the horizontal planimetric accuracy (the correct longitudinal and latitudinal placement of a feature on the Earth's surface) is expected to be less than 1.6 m, which we consider sufficiently accurate for our 1 km resolution analysis.

We combine the available reference data with the manually classified data derived from Google maps to obtain the full set of reference data for training the model. Fig. 3 shows the distribution of reference data as a function of footprint and height, in which only reference data where both footprint and height are valid are shown. Reference data

points (i.e. 1×1 km cells) are unevenly distributed within one region, but show complementarity across the three case regions. Specifically, the US has more reference data in medium-footprint and low-height compared to Europe and China, while China has more reference data in medium-footprint and medium-height than the other regions.

2.4. Model development and evaluation

As mentioned, we estimate three parameters for each pixel: 1) building footprint (m^2/m^2), 2) building height (m), and 3) building volume (m^3/km^2). For classification, we first selected only 1×1 km pixels that have built-up land > 0 according to GUF, primarily to improve the computational efficiency. Therefore, valid reference data for building footprint, building height, and building volume, described in part 2.3 account for 1.17%, 1.00%, and 1.00%, respectively, of the total area included in the model (also see Table 2).

The ensemble regression random forest (RF) approach is used for estimating building footprint, building height, and building volume. This is an efficient prediction method, especially when observations are

Table 2

Number of 1×1 km cells in the reference data as well as in the predicted data. Collected data refers to data that was collected from multiple sources, while Interpreted data refers to data that was manually interpreted from Google Maps.

| Predicted data points | Reference data points | | | | | |
|-----------------------|-----------------------|----------------|-----------|----------------|-----------|----------------|
| | Footprint | | Height | | Volume | |
| | Collected | Interpreted | Collected | Interpreted | Collected | Interpreted |
| Europe | 1,681,014 | 13,728 | 2469 | 13,731 | 2466 | 13,731 |
| The US | 1,447,489 | 30,350 | 1091 | 24,551 | 1091 | 24,465 |
| China | 1,632,283 | 5814 | 2204 | 5800 | 0 | 5800 |
| Total | 4,760,786 | 49,892 | 5764 | 44,082 | 3557 | 43,996 |
| | | 55,656 (1.17%) | | 47,639 (1.00%) | | 47,553 (1.00%) |

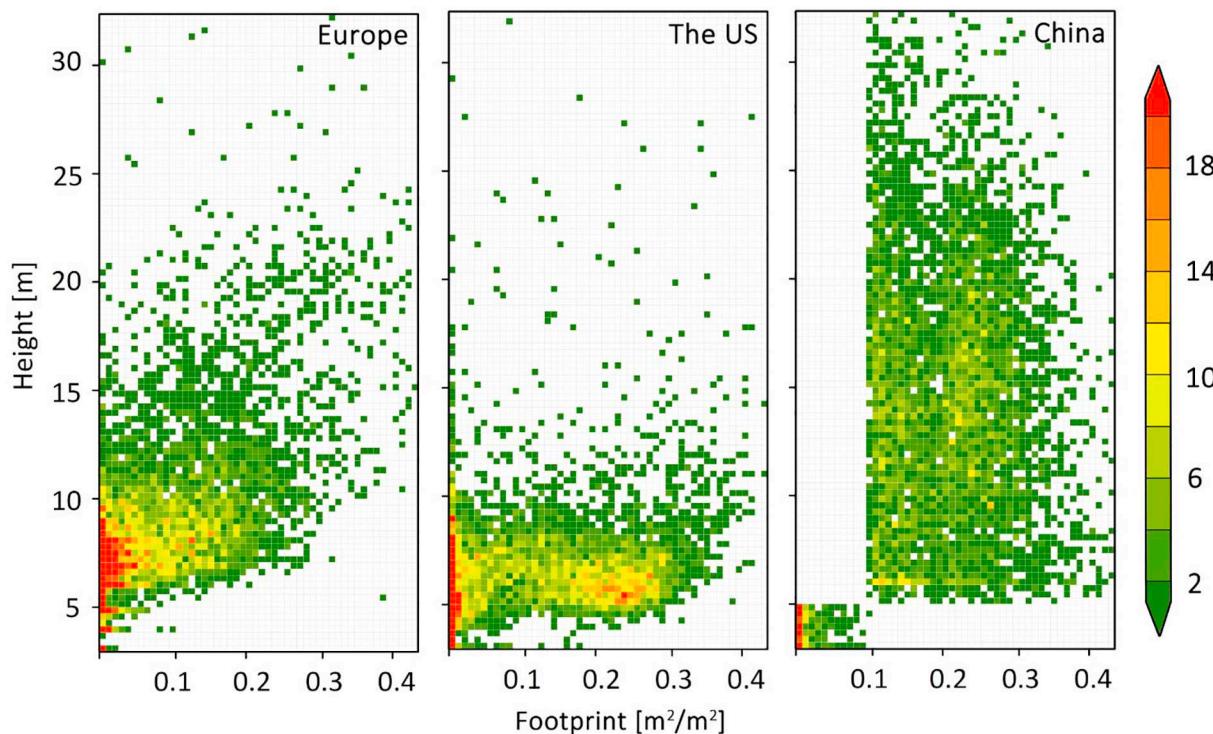


Fig. 3. Distribution of reference data (1×1 km cells) for mapping 3D building structure. For China, we removed uncertain data points from the reference data, which are identified by “footprint $< 0.1 \text{ m}^2/\text{m}^2$ and height $> 5 \text{ m}$ ” or “footprint $> 0.1 \text{ m}^2/\text{m}^2$ and height $< 5 \text{ m}$ ” because our preliminary evaluation of these data showed large inaccuracies.

Table 3

Population data collected for the analysis of building occupation per capita in different case study regions.

| Region | Time | Source |
|---------------------|------|---|
| Europe | 2015 | https://population.un.org/ |
| The US | 2015 | https://www.census.gov/ |
| China | 2015 | http://www.stats.gov.cn/ |
| Hong Kong and Macao | 2015 | https://www.worldbank.org/ |

much scarcer compared to the predictors (Svetnik et al., 2003). The RF model is trained and applied for each of the three variables and for each case region separately, as well as for all regions combined. RF combines several decision trees, built on different combinations of explanatory variables, and produces the mean prediction of the individual trees. This strategy is beneficial to alleviate the overfitting problem of simple decision trees (Pelletier et al., 2016; Tramontana et al., 2015). The primary property of tree-models is a partitioning of space into smaller regions to manage phenomena characterized by very complex interactions among variables. In particular, in tree models, partitioning is recursive. The phenomena occur when the subdivisions are divided again until the partitioning reduces the appropriate cost function. Recursive partitioning is terminated when the cost function cannot be further minimized. Hence, a simple model, usable only for the partitioned sub-region, can be estimated. For each observation, the output of a RF model is the average of the outputs of the trees. Therefore, RF models typically yield a reduced bias in the estimations and in general good accuracies (Tramontana et al., 2015). More technical details on the applied RF algorithm can be found in Breiman (2001).

We develop the RF models using scikit-learn, a machine learning package in Python (Pedregosa et al., 2011). To some extent, more trees yield better results. However, the improvement decreases as the number of trees increases, and at a certain point the benefit in prediction performance from including more trees will not be worth the extra computation resources. Therefore, after initial tuning experiments

we maximize the number of trees to 150, whereas the minimum number of samples required at a leaf node is fixed to 5. The importance of explanatory variables is measured by the Gini decrease in node impurity measure, which is computed by permuting the explanatory variables with the out-of-bag data in the RF validation approach (Breiman, 2001).

After training the RF models to estimate building footprint, height, and volume using the reference data, we apply the trained models to the entire study regions. For each pixel and for each of the three characteristics the RF model initially estimates 100 values, corresponding with the 100 trees in the RF-model, and the mean of these values is used as eventual outcome for that pixel. For each of the three building properties, the reliability of our model is evaluated by a ten-fold cross-validation method as well as an uncertainty analysis. The independent validation dataset is built by a random selection of 10% of the reference data in these three regions, while the other reference data (90%) are used as training data. This process is repeated 100 times, and for each run we calculate the Pearson's correlation coefficients (R^2) to express the agreement between predicted and observed values. In addition, we quantify uncertainty as the range of values generated by all trees in the RF model for a specific pixel. A large range indicates that individual models differ widely, which we interpret as an uncertain estimate. Conversely, a small range is interpreted as agreement in the trees of the RF model and thus a relatively certain estimate. Specifically, for the 100 predicted values in each cell, we calculate its coefficient of variation (CV) as the indicator for uncertainty, see Eq. (1):

$$CV = \frac{\sigma}{\mu} \quad (1)$$

where σ and μ refer to the standard deviation and mean value of a corresponding cell for these 100 runs, respectively, where μ is also the final predicted value as defined here.

To further asses model performance, we calculate the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Systematic Error (SE). To ensure independency, we calculate RMSE, MAE, and SE

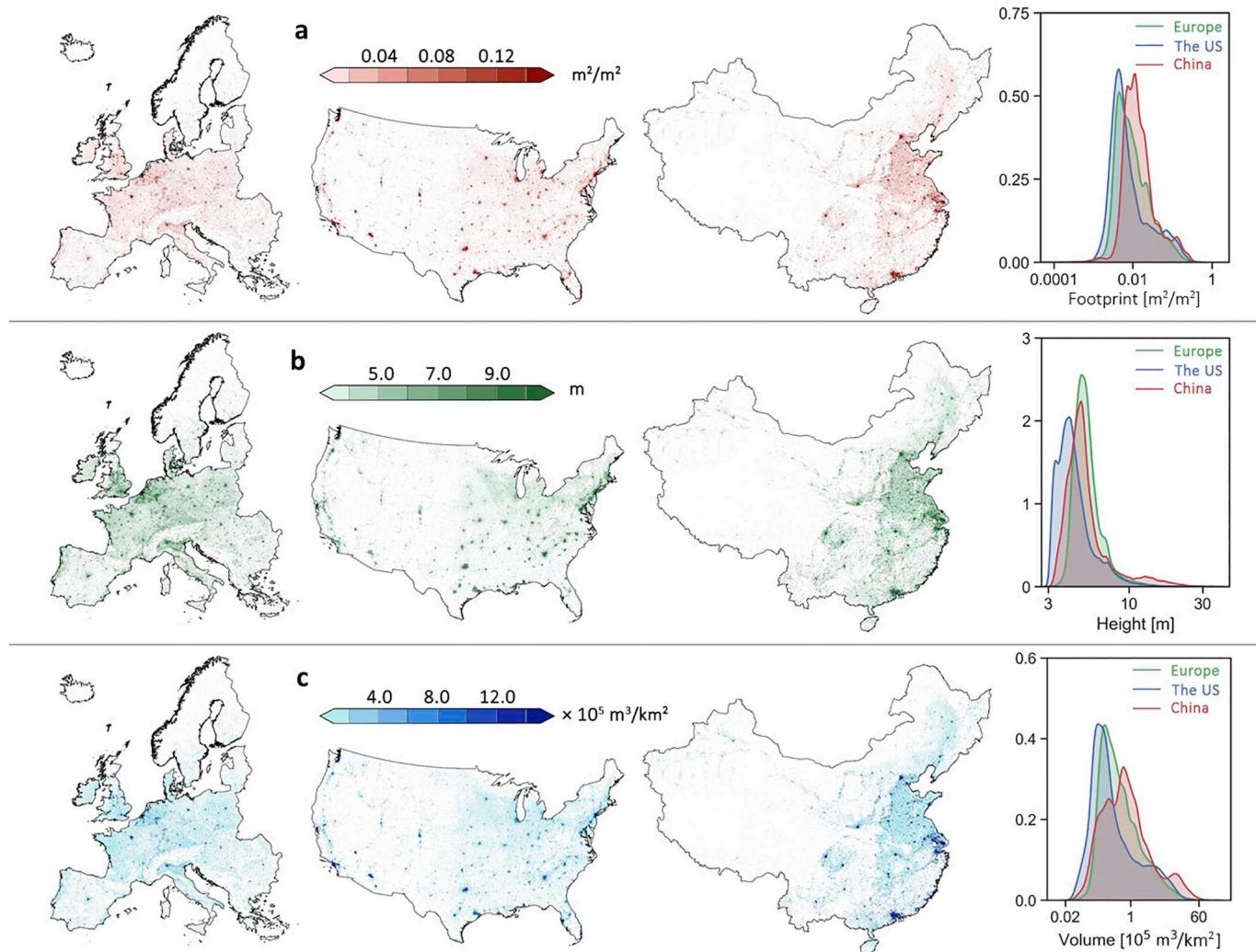


Fig. 4. Distribution of building structure in the three study regions. a) building footprint; b) building height; c) building volume. The graphs on the right show the kernel density estimations, of which the x-axis is scaled using a logarithmic function. The area under the curves is normalized to 1 to facilitate the comparison of distributions across continents when using the logarithm transformed value of x-axis.

based only on the data points in our reference data that were not used for training the model, following Eq. (2), Eq. (3), Eq. (4), and Eq. (5).

$$B_{\text{pred},j} = \frac{\sum_{i=1}^t B_{\text{pred},j,i}}{t} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^s (B_{\text{pred},j} - B_{\text{test},j})^2}{s}} \quad (3)$$

$$\text{MAE} = \frac{\sum_{j=1}^s |B_{\text{pred},j} - B_{\text{test},j}|}{s} \quad (4)$$

$$\text{SE} = \frac{\sum_{j=1}^s (B_{\text{pred},j} - B_{\text{test},j})}{s} \quad (5)$$

where $B_{\text{pred},j}$ is the predicted value of endmember j , while $B_{\text{pred},j,i}$ refers to the predicted value of endmember j in the i th model, and t refers to the total number of independent predictions that the endmember j is included. $B_{\text{test},j}$ is the reference value of endmember j , and s is the total number of unique endmembers in all test collections for independent predictions in the 100 models.

Finally, we examine variable importance of the best-fitted runs as identified by their R^2 values. The core principle of variable importance is to calculate the degradation of model performance if such variable is permuted randomly while keeping other input variables constant,

which allows for evaluating the relevance of one variable for model output (Zhao et al., 2019).

2.5. Analysis of the building structure

We estimate 3D building structure for three case study regions, and compare the results to analyze structural differences. To do so, we calculate the average as well as the distribution of all three variables in all three regions and for each European country, American state, and Chinese province in these regions, using population data from multiple sources (Table 3). Subsequently, we create distribution curves to measure the distribution of building footprint, height, and volume for each region, based on 100,000 randomly $1 \times 1 \text{ km}$ grid cells for which results are estimated. These distributions are subsequently compared across regions. Moreover, we analyze the correlation coefficients between the building properties in the case regions as well as the combined region.

3. Results

3.1. Characterization of 3D building structure in Europe, the US, and China

The distribution of building footprint, height, and volume corresponds largely between regions: high values for all three variables are,

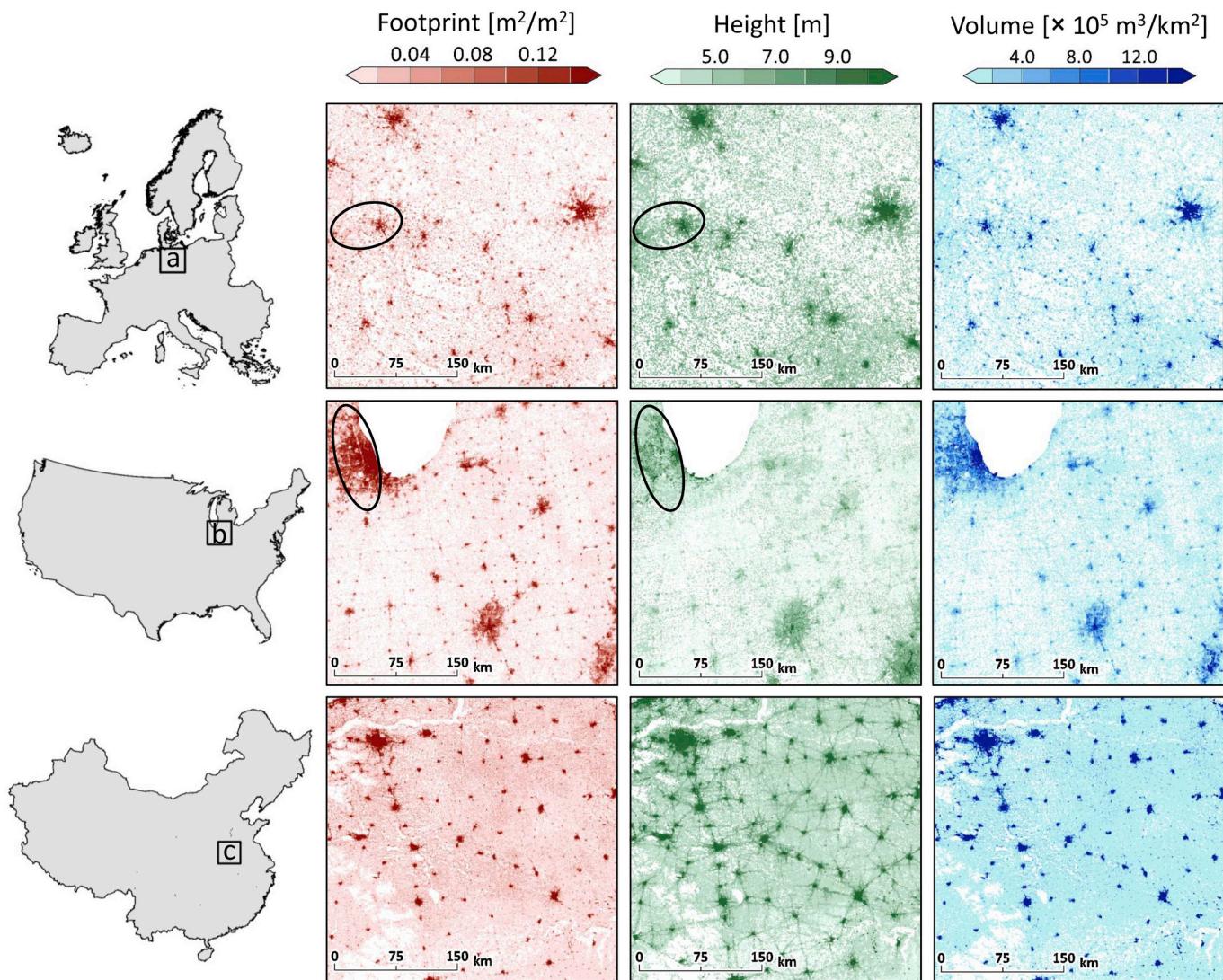


Fig. 5. Building structure in three densely populated areas located within predominantly agricultural plains. a) around Berlin; b) around Chicago; c) around Zhengzhou in the province Henan. Encircled areas indicate regions with smaller footprints but higher buildings in 5a (around Hannover), and higher footprint but lower buildings in 5b (around Chicago). These three areas are selected to show typical settlement patterns that are dominantly shaped by human activities, rather than natural or biophysical constraints such as topography.

as expected, concentrated in and around the larger urban areas of the three regions, such as Paris, New York, and Shanghai (See Fig. 4). Yet, there are notable differences across the three study regions, which are visible from the distribution curves of all values per continent in the right of Fig. 4. For example, China has more pixels with a relatively large building footprint as well as a high building height, while the US has more pixels with a low building height, typical for suburban sprawl. Specifically, China has the highest average building height at 10.35 m. In Europe, the average building height is 7.37 m, and in the US this is 6.69 m. Consistently, China has more areas that have a very high building volume, while the opposite is true for the US.

A more detailed inspection of 3D building structure highlights the different spatial configurations of buildings in different regions (Fig. 5). For example, building footprints in the Chinese agricultural plain (Fig. 5c) are rather dense, as compared to rural areas in Europe and the US. A large urban footprint is often associated with high-rise buildings, especially for China. Yet, this seems not appropriate for many locations in the US, as is illustrated in the area encircled in Fig. 5b. Conversely, we also find some areas with a relatively sparse footprint value and a large height value (e.g. around the city of Hannover in Europe, encircled in Fig. 5a). The detailed results in Fig. 6c also highlight a

particular phenomenon in China, where buildings tend to be taller along main roads that connect large cities, much more than these in Europe and the US.

Further quantitative analysis shows that building footprint, height, and volume are correlated, but this correlation is well below 1 (Fig. 6). This indicates the need to analyze the three properties independently. The correlation coefficient between footprint and height ranges from 0.55 in the US to 0.74 in Europe. The correlation coefficients between volume and height as well as volume and footprint are higher ranging from 0.69 in the US to 0.93 in Europe. It is not unexpected that the correlation between footprint and height is lower than the other two correlation coefficients, as volume is by definition the product of footprint and height, and thus at least partially related to both of these properties. Nonetheless, all three properties are estimated independently in this study, and therefore this correlation is not trivial from the setup of the study.

The average building footprint per capita is only 29.2 m^2 in China, which is about one third of that in the US (84.5 m^2), and about a half of that in Europe (54.9 m^2). Building volume per capita in China is 302.3 m^3 , while it is 565.4 m^3 in the US, and 404.6 m^3 in Europe. These results indicate that settlements in the US have a higher land take per

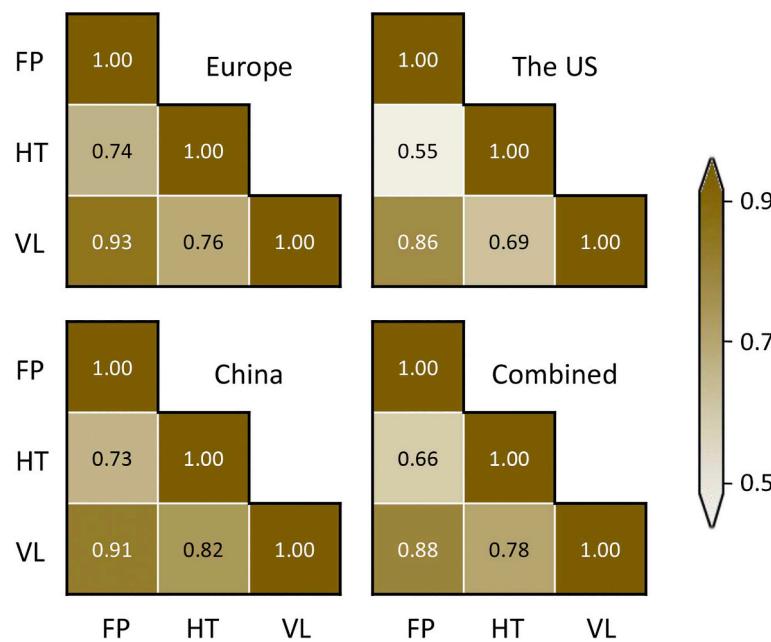


Fig. 6. Correlation coefficients for building footprint, height, and volume, which are referred to as FP, HT, and VL, respectively, for each of the three case study areas separately, as well as for all areas combined.

person as well as a higher space consumption per person, in comparison to the other regions.

The spatial distribution of 3D building structure also differs between sub-regions (Fig. 7). For example, the values for building footprint per capita vary much more across US states than across European countries and Chinese provinces, and especially high values are observed in several predominantly rural states such as North Dakota, South Dakota, Wyoming, Iowa, and Montana (Fig. 7a). At the same time, in Sichuan and Guizhou, two rural sub-regions of China, building footprint per capita is lower than most other equally-developed sub-regions (Fig. 7a). Building height, on the other hand, varies most across Chinese provinces and much less across EU countries and US states (Fig. 7b). Buildings tend to be lower in the inland rural states of the US. Conversely, buildings are much higher in developed sub-regions of China, most of which are coastal sub-regions. The distribution of building volume per person is mostly consistent with the distribution of building footprint per person, with relatively high variation in the US and relatively low variation in Europe and China (Fig. 7c). In the US, sub-regions of which have a large building volume per capita are mostly located in rural inland states with high values in building footprint per capita, despite their moderate height. In China, regions that have large building volume per capita are mostly located in urbanized coastal sub-regions such as Jiangsu and Zhejiang, characterized by high buildings but not necessarily by a large building footprint per capita.

3.2. Model performance and uncertainty

The RF models yield high accuracies for building footprint, height, and volume, as indicated by R^2 values for the three regions combined all larger than 0.80, either for the separated models or the combined models (Fig. 8). When models are run for each case region separately, building footprint is most accurately predicted for the US. As for the building volume, results for Europe and the US are more accurate than for China. When models are run for all case regions combined, there is no significant improvement compared with the separated models. Among the three properties, building footprint is most accurately predicted, especially for the US. As for the building volume, results for

Europe and the US are more accurate than for China.

To further characterize the accuracy of our estimates, we assessed the RMSE, MAE, and SE, based on the independent validation data for each model. It should be noted that the training data have on average higher values of building footprint, height, and volume, thus also leading to higher values for RMSE, MAE, and SE than can be expected for the complete estimated data set. For the combined models, the RMSE values of building footprint, height, and volume for the three regions combined is $0.03 \text{ m}^2/\text{m}^2$, 2.69 m , and $6.03 \times 10^5 \text{ m}^3/\text{km}^2$, respectively. Correspondingly, MAE values of the three building properties are only $0.02 \text{ m}^2/\text{m}^2$, 1.36 m , and $2.55 \times 10^5 \text{ m}^3/\text{km}^2$. While SE values for the three building properties are all close to 0, suggesting that there is no lateral overestimation or underestimation in general.

The accuracy of separate models and one combined model for all regions is somehow comparable, but combining reference data for all case regions into one model yields a decrease in uncertainty, relative to models trained on one region only (Fig. 9). Therefore, analyses in the above section are based on the results generated by the “combined” model. Specifically, when trained with data from all regions together the model for building footprint shows a large decrease in uncertainty in areas with a low building footprint ($< 0.1 \text{ m}^2/\text{m}^2$), which accounts for a large proportion of the study area. Uncertainty of predicted building height shows a decreasing trend over a wider range of values compared to building footprint, especially for the US and China. Moreover, uncertainty is distributed unevenly over different combinations of building footprint and building height (Fig. 10). Notably, the uncertainty in building footprint was found mostly in areas with either a rather sparse footprint (around $0.04\text{--}0.1 \text{ m}^2/\text{m}^2$), or at a rather dense building footprint (around $0.3 \text{ m}^2/\text{m}^2$). Overall, uncertainty of building height is largest in areas with small values for building footprint and high values for building height. In particular, we find some scattered pixels with large uncertainty in some mountainous areas of southern China, which can potentially be explained by interference from other landscape elements, such as trees and rocks. We also find that building height is prone to large uncertainty in areas that are not covered by SAR data, for example, a diamond-shaped area in Sichuan province of China, and a small square area in Milwaukee city of the US. The largest

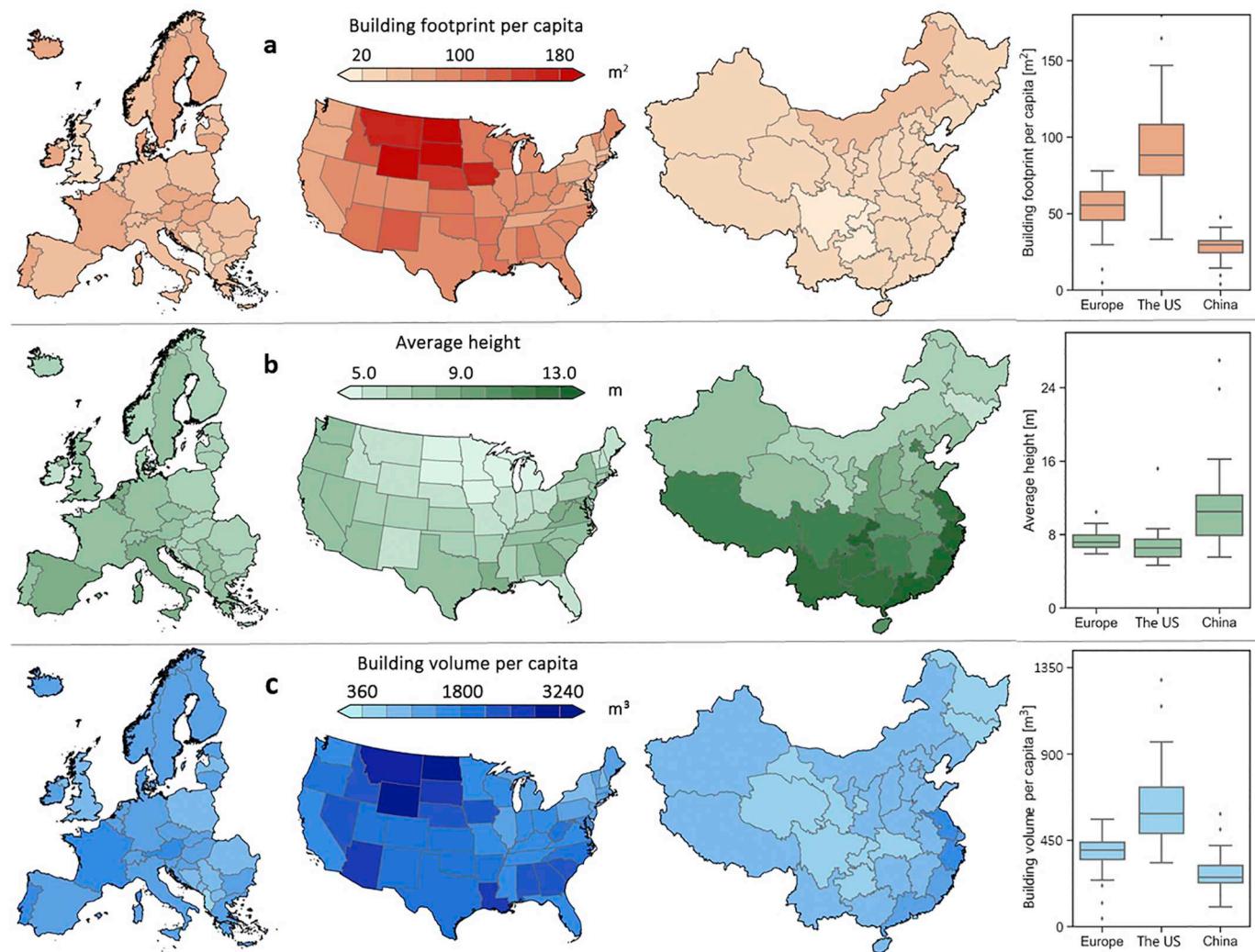


Fig. 7. Analysis of 3D building structure at sub-regional scale. The boxplots on the right are plotted based on all sub-regions for each study area, of which y-axes are capped to enhance interpretation. The boxes represent the interquartile ranges (25%–75%), and the lines represent the ensemble-median values.

uncertainty in building volume is found in areas with a low building footprint and a medium-high building height, as well as in some locations with a high building footprint. Possibly, this uncertainty is also explained by disturbance of other objects, especially in places with a low building footprint.

Zooming in on individual cities further demonstrates the superior performance of the combined model over the models for separate regions (Fig. 11). For building footprint there is no visible difference between the separate models and the combined model. However, the separate models generally overestimate building height for Paris, Los Angeles, and Shanghai.

The best performing models for each of the three characteristics of building structure are selected for further analysis of the variable importance. This analysis reveals that built-up density derived from GUF, in general, is most valuable for estimating building footprint, while backscatter intensity of SAR has little influence (Fig. 12). The opposite is true for building height estimation, as backscatter intensity has the largest importance, while built-up density is of course of little influence. Compared to other variables, both built-up density and backscatter intensity are important to predict building volume. In addition, we find a trend shift of VH/VV variable importance when reference data in these three regions are combined. A further analysis indicates that VH and VV are complements when explaining height and volume (Figs. S4 and S5).

4. Discussion

4.1. Mapping 3D building structure at a continental scale

This study shows that the combination of various remote sensing data sets and other spatial data allows estimating building footprint, height, and volume at a continental scale with high accuracy. Hence our models make it possible to map the built landscape in three dimensions, to analyze the structural specifics for certain regions (e.g. rural vs. urban) and to analyze differences across geographical characteristics. The proposed RF models yield high accuracies (R^2 values larger than 0.80 for all regions combined), outperforming other models for estimating 3D building structure at large spatial scales, such as the Bayesian Network-based model developed by Paprotny et al. (2020) in terms of R^2 , RMSE, and MAE, which are mutually reported in the two studies.

We find that buildings in China are the highest on average (10.35 m), compared with the other regions (7.37 m for Europe, and 6.69 m for the US). Higher values for building height are especially found on the urbanized east-coast of China, suggesting that the recent and rapid urban development characterized by multi-story buildings affects the country average (Mahatta et al., 2019). Conversely, while Europe is known for relatively compact development, urban expansion in the US is for a large part manifested as suburban sprawl (Barrington-Leigh and Millard-Ball, 2015; Dong et al., 2019). This type of

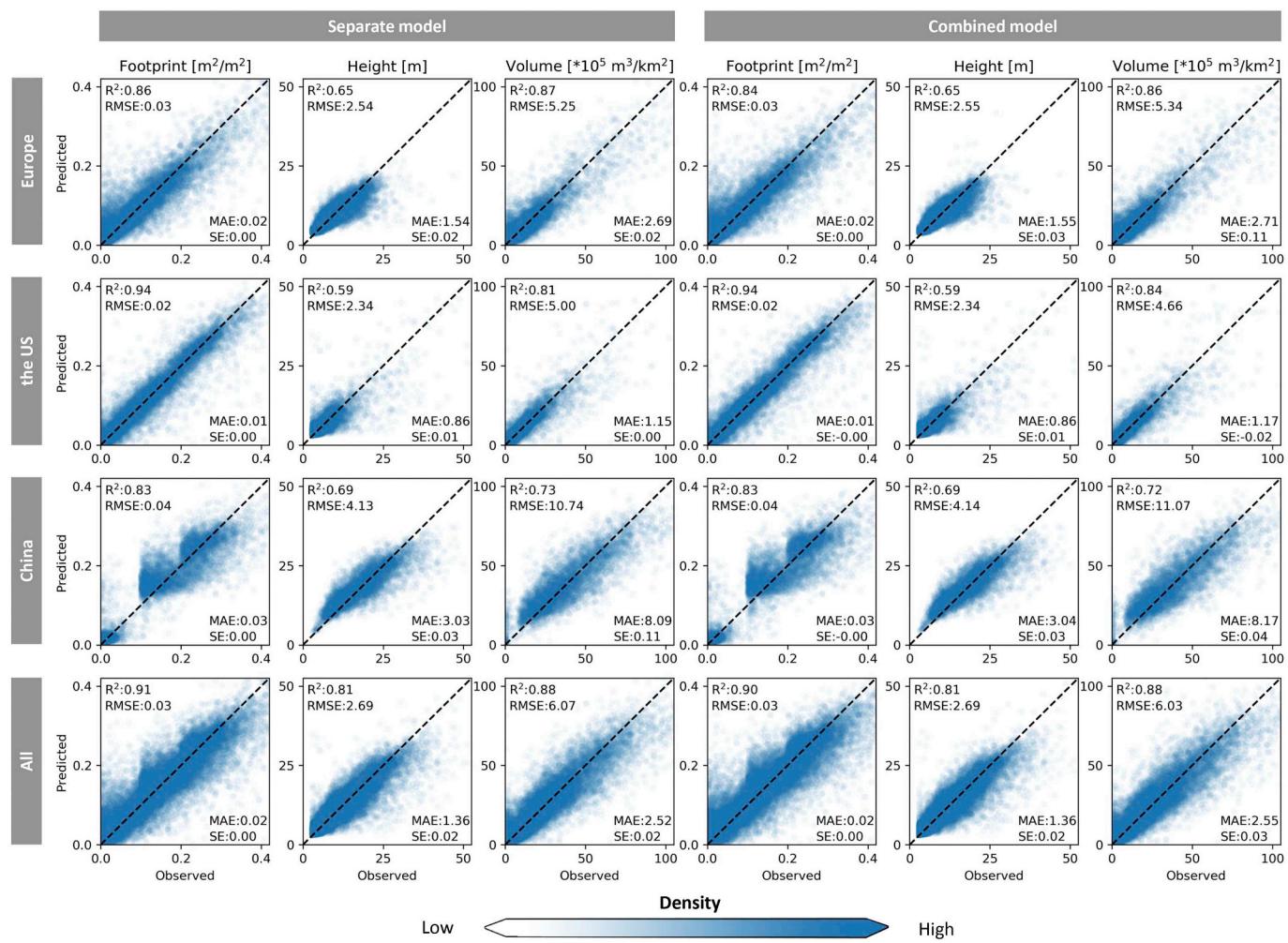


Fig. 8. Scatterplots of the observed values and predicted values for building footprint, height, and volume. Predicted values represent the mean of all independently predicted values for each location in the reference data.

development is not only characterized by a relatively low density of buildings, but also by relatively low buildings, as is reflected in the lower average height for the US. Notably, the large building occupation per capita in some rural states of the US could be related to the abundance of agricultural buildings such as barns for livestock (Harun and Ogneva-Himmelberger, 2013). At the same time, both building footprint per capita and building volume per capita are the smallest for China. Some studies have found that urban land per capita is driven by biophysical and socioeconomic conditions such as terrain characteristics, wealth, price of gasoline, and planning strictness (Angel et al., 2011; Taubenböck et al., 2018). These characteristics could at least partly explain the observed differences. For example, GDP per person is higher in the US than in Europe, on average, which is again higher than in China. Conversely, urban planning is rather strict in China, due to national planning regulations (Liu et al., 2014), while it is weakest in the US. Yet, existing analyses of urban building structure have mainly focused on urban footprint only, while the other characteristics of 3D building structure remain to be explored in more detail. Yet, building height and volume could have considerable impacts, for which this dataset provides a continental-scale source for further analysis.

Mapping 3D building structure at large spatial scales could further benefit from the accelerated developments of Artificial Intelligence (AI), which increasingly serves as a powerful tool for addressing complex problems (LeCun et al., 2015; Reichstein et al., 2019). However, one of the most essential and challenging parts of AI is that it needs to be trained through large amounts of precisely labelled reference data.

Currently, there are several databases available for universal objects such as the well-known ImageNet (Deng et al., 2009). Increasingly, there are some urban thematic benchmark databases such as DeepGlobe (Demir et al., 2018), BigEarthNet (Sumbul et al., 2019), and SEN12MS (Schmitt et al., 2019). Yet, these datasets focus mostly on the identification of objects, whereas they do not provide sufficient information on building height and volume. Therefore, we additionally developed a large amount of new reference data for this study specifically. In parallel, computer vision research has made great progress in detecting changes based on digital imagery (Kuehne et al., 2011; Soomro et al., 2012). These developments could greatly benefit urban scientists in characterizing changes in building structure based on time-series satellite data. Yet, as several of the data that feed into our analyses are only available for recent years, notably Sentinel-1 SAR data, change analysis of 3D building structure remains challenging.

4.2. Urban land use intensity and other applications of 3D building structure

Urban expansion plays an increasingly important role in the global competition for land, and impacts of urban expansion have been widely reported in scientific literature. For example, urban expansion on a global scale leads to the displacement of cropland and subsequent losses in natural areas (van Vliet, 2019). Consequently, increasing urban land use intensity could be a way to reduce urban expansion and thus alleviate the global competition for land. Population density has been used frequently for analyzing urban land use intensity. However, population

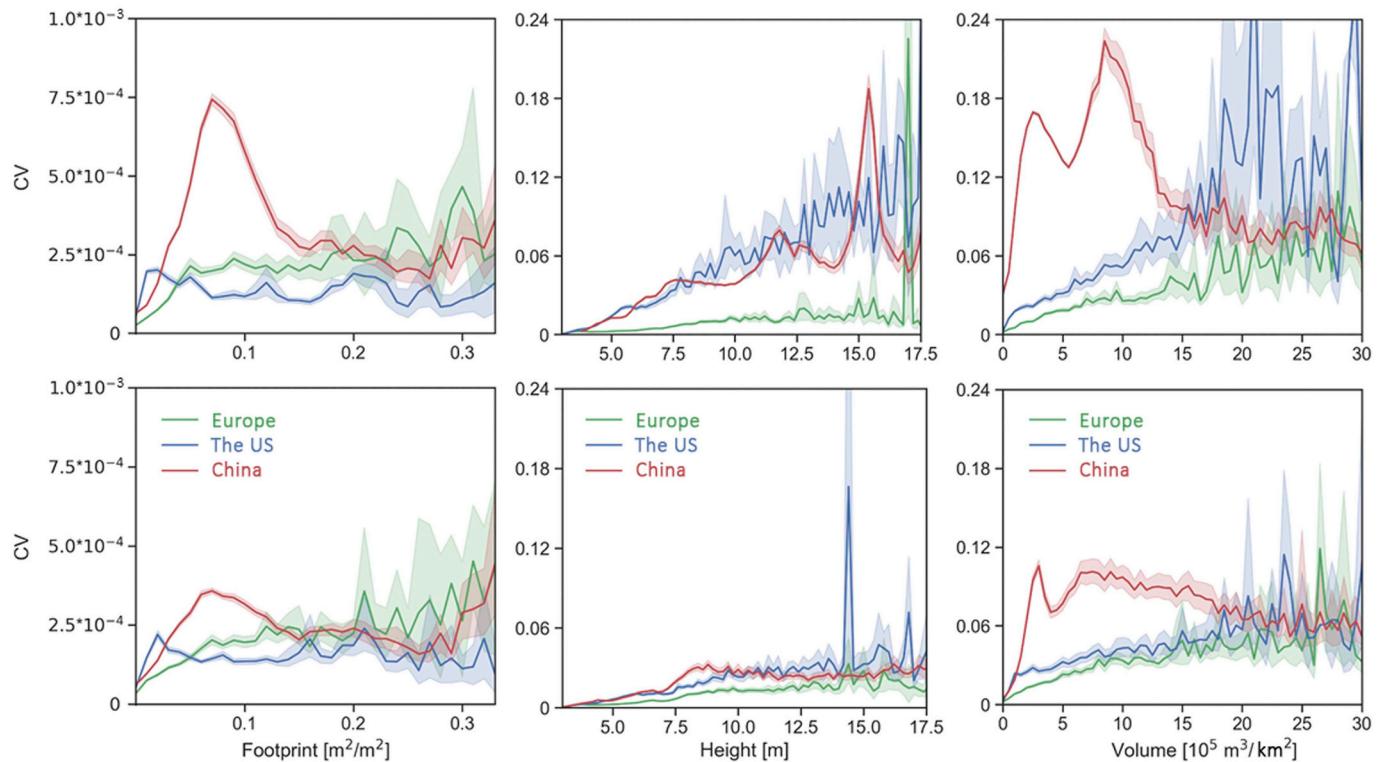


Fig. 9. Uncertainty, as expressed in the CV values of three predicted properties as a function of their values for these corresponding building properties. Upper row: model ran separately in each of the three regions; bottom row: model ran for all three regions combined. All these figures are plotted based on 100,000 randomly selected points within each region. The shaded areas represent the interquartile ranges (25%–75%) and the lines the ensemble-median values.

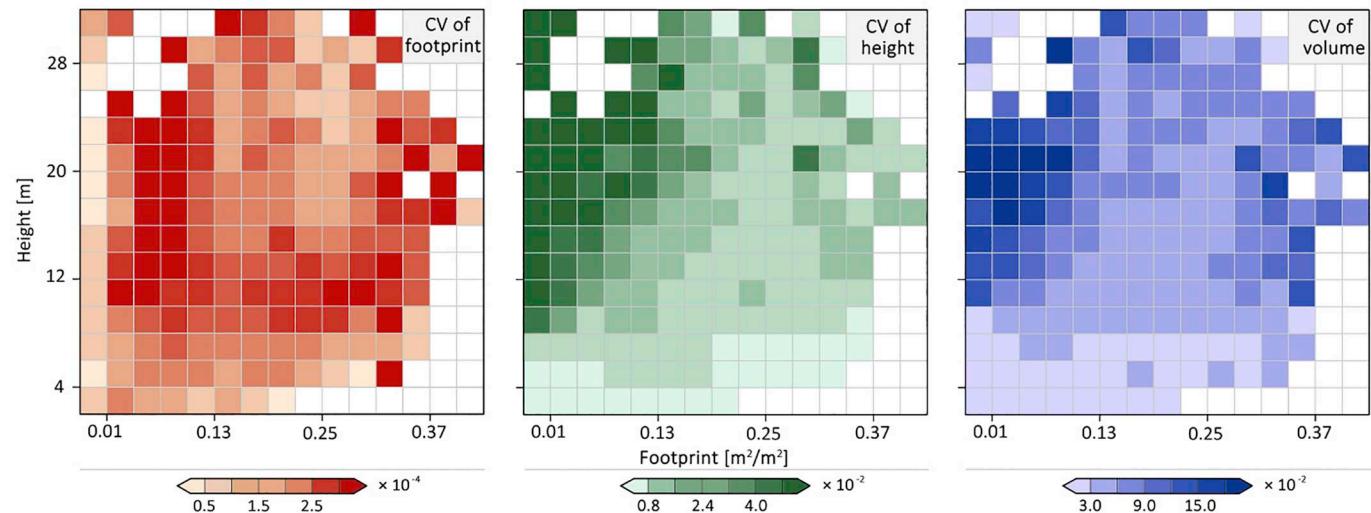


Fig. 10. Coefficient of variation (CV) plotted as a function of the combination of building footprint and building height, for the combined reference data set of all three study regions, based on results in a sample of 300,000 pixels in the three regions combined.

density maps are mostly produced by using a downscaling approach, based on a combination of census data and spatial data, such as nighttime light and built-up area (Florczyk et al., 2019; Wang et al., 2018). Therefore, while population data is typically rather accurate at the census level, they remain more uncertain at the local/pixel level. Moreover, population density reflects residential activities only, while other urban activities remain unaddressed (Dovey and Pafka, 2013).

Building characteristics as presented in this paper offer an alternative to population density data for characterizing urban land use intensity, as they present the footprint, height, and volume of buildings in a pixel. The same information is also underlying Local Climate Zones

(LCZs), a standardized classification of urban land use types presented by Stewart and Oke (2012) for urban climate research. The main difference is that our results are provided on a continuous scale, while LCZs include a limited range of discrete classes. Similar to previous studies of 3D urban structure, large scale analyses based on this framework are constrained to the scarcity of 3D building information. Previously, such information has already been presented for selected global megacities (e.g. Bagan and Yamagata (2012), Mertes et al. (2015), and Taubenböck et al. (2012)). These data provide only information of a limited area, while a large proportion of the built-up area is located outside these megacities (Li et al., 2019). The approach

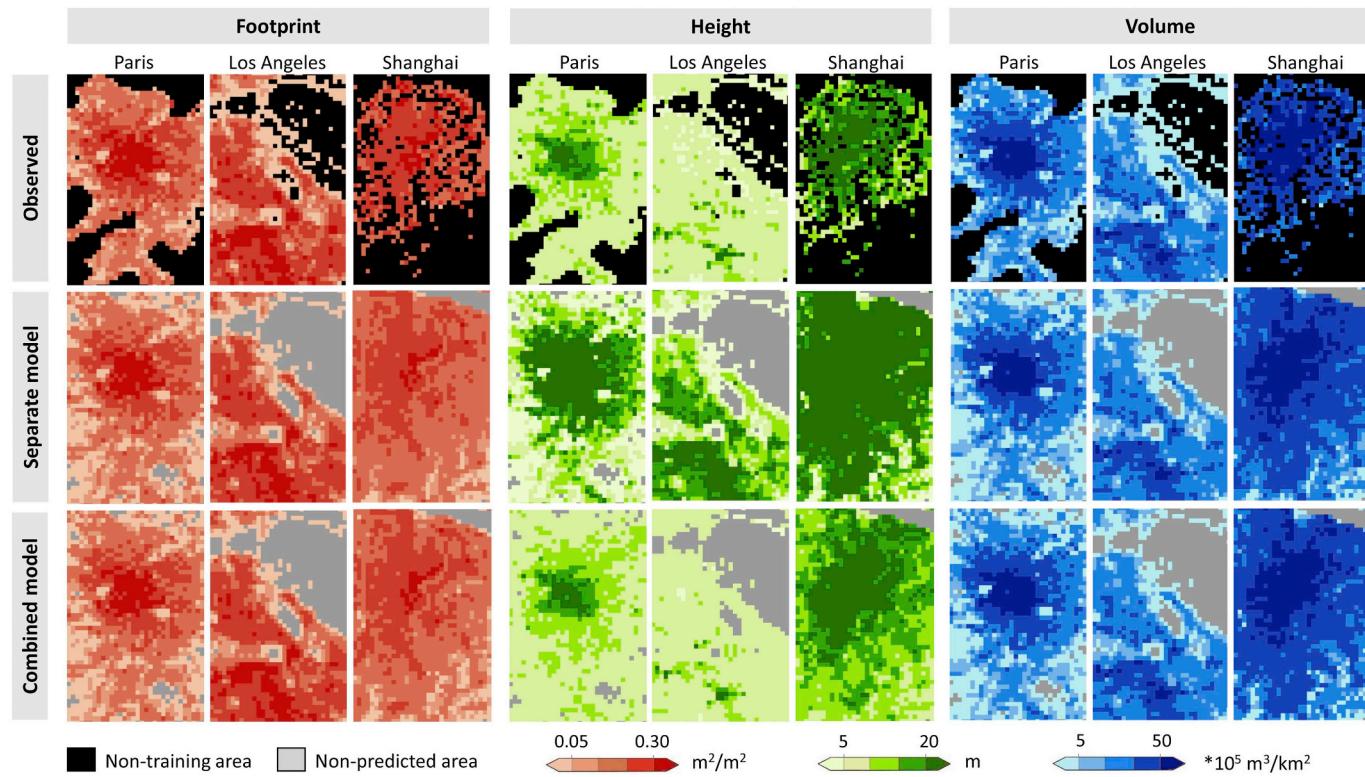


Fig. 11. Comparison of observed and predicted results for building structure in Paris, Los Angeles, and Shanghai. Each map is 30 × 50 km in size.

presented in this paper therefore complements population density as a measure for urban land use intensity on a continuous scale, and covers all types of human settlements regardless of their size. Conversely, a few studies have investigated population distribution based on building volume at a local scale (e.g. Dong et al. (2010), Tomás et al. (2015), and Zhao et al. (2017)). Hence the continental-scale building structure data produced in this study could also move the estimation of population distribution forward.

The comparison between building footprint and height shows that they are only partly correlated (0.55 for all three case regions combined). In other words, there is a considerable amount of variation in building height within locations with a comparable building footprint, thus justifying the mapping of these properties separately. This also implies that the analysis of 2D urban density as a proxy for urban intensity hides a significant part of the variation in actual building structure. Local patterns in the relations between building footprint and building height also differ across the three regions. The particular phenomenon in China that buildings along main roads tend to be relatively high, suggests that local conditions largely affect building structure. However, evidence for these differences as well as explanations for their causes is still sparse in the literature. For example, this particular phenomenon in China could be attributed to the mobility requirements of population (Wang et al., 2016), which facilitates the development of retail and service industries, resulting in higher buildings for mixed uses along main roads. Yet, this push-pull theory behind the spatial differences in 3D building structure is rather anecdotal.

The generated datasets on building footprint, height, and volume provide several opportunities for further analysis of urban structure and its impacts. First, this data can facilitate the classification of different settlement types (e.g. suburbs, slums, and business districts) based on a priori knowledge of these settlement types (Taubenböck et al., 2018), to further investigate social or environmental impacts of urban areas. Second, information on the urban vertical dimension is of much value for the field of disaster risk science as well. When knowing both the footprint and height of a building, one can much better specify the

potential vulnerability of a building to, for instance, floods, storms, and earthquakes (e.g. Du et al. (2018), Koks and Haer (2020) and Paprotny et al. (2020)). At the same time, knowing height of buildings is an essential metric for urban heat stress modelling and its potential socio-economic consequences as well (Lemonsu et al., 2015). Another application area is the impact of urban form on environmental conditions (Seto and Shepherd, 2009). To what extent urban climate is affected by building form and their mechanisms remains unclear, as conclusions vary across cases (Manoli et al., 2019; Yue et al., 2019; Zhou et al., 2017). However, most of these studies still focus on urban configuration in 2D dimensions such as city size and urban centricity.

Our study also reveals the potential to guide settlement development towards sustainable land use patterns for the benefit of human well-being. In the sustainability community, consensus has not been reached on whether urbanization is part of the problem or a solution to sustainability challenges (McFarlane, 2019; Seto et al., 2010). Either way, urban densification, both horizontally and vertically, is acknowledged as one of the tangible solutions to satisfy the increased urban land demand while conserving other land (Wang et al., 2019b). However, we also notice that local settlement trajectories should be guided in a large-scale context with broad considerations, including quality of life for inhabitants of human settlements, while these trade-offs and synergies remain largely unexplored.

5. Conclusion

This study presents the first continental-scale dataset on 3D building structure for Europe, the US and China. The presented data was generated using random forest (RF) models fed with optical remote sensing imagery, SAR imagery, remote sensing derived indices, and other spatial data. The RF models yield R^2 values of 0.90, 0.81, and 0.88 for building footprint, height, and volume, respectively, for all three continents combined. Our results show that building height is to a large extent independent from building footprint, emphasizing the importance of mapping these properties independently. The average

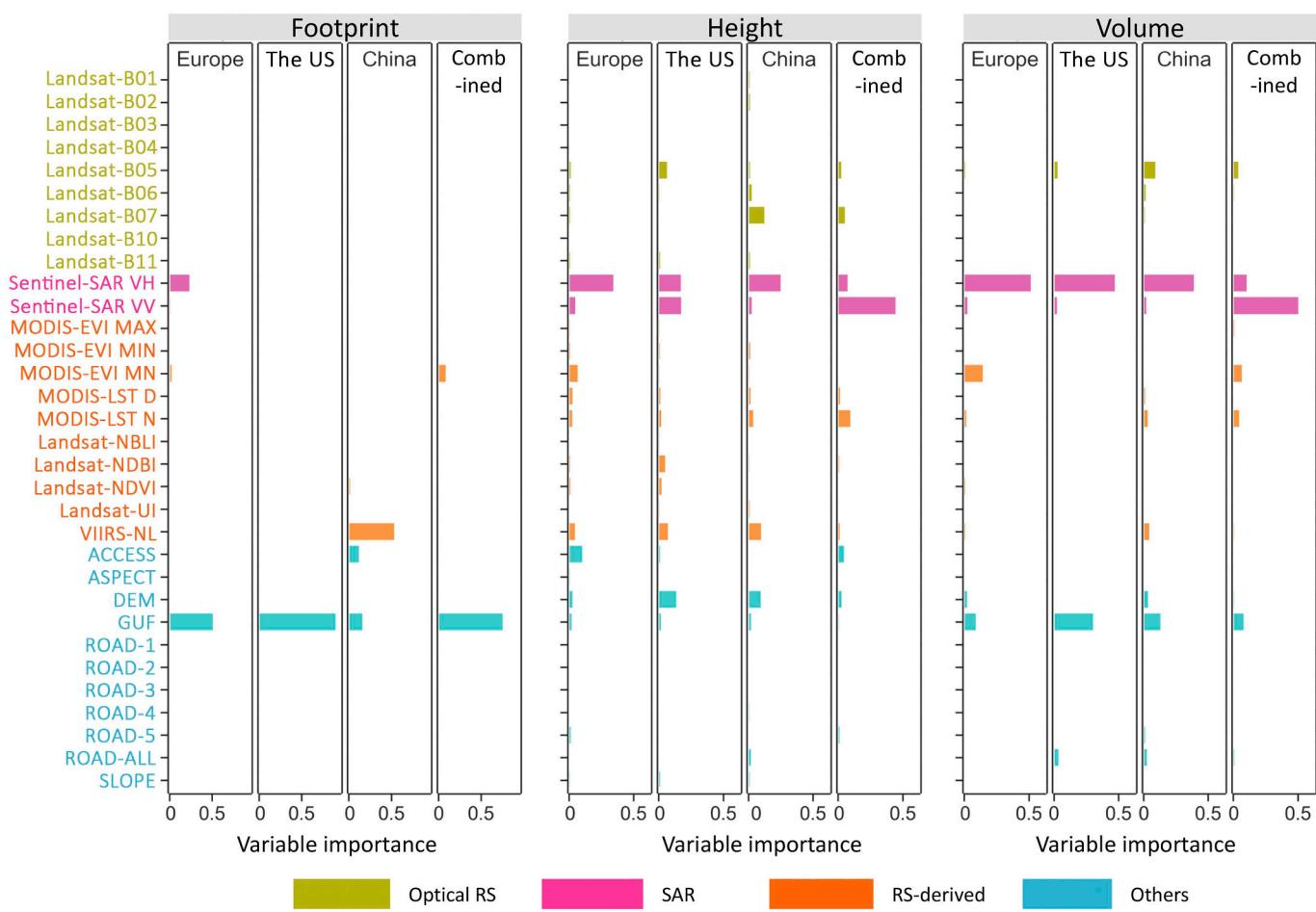


Fig. 12. Normalized variable importance based on the best-fitted run as identified by their R^2 . We run the model separately for Europe, the US, and China, as well as for all regions combined. For each property in one model, the summed variable importance equals 1.

building footprint per capita is only 29.2 m^2 in China, which is about one-third of that in the US (84.5 m^2), and about a half of that in Europe (54.9 m^2). Building volume per capita in China is 302.3 m^3 , which is 565.4 m^3 for the US, and 404.6 m^3 for Europe. The 3D building structure data produced in this study provide a nuanced representation of settlement patterns, which can be used for urban environmental analysis, spatial planning, and land use modelling that aim to guide the sustainable development of settlements. In itself, these data already reveal geographic peculiarities across different regions in the globe.

Data availability

The full list of satellite imageries downloaded from Google Maps and corresponding interpreted results are freely available (<https://doi.org/10.6084/m9.figshare.c.4672556>). All datasets used in our analysis, as well as the codes for model algorithm and statistical visualization are also available (<https://cscproject.github.io>).

CRediT authorship contribution statement

ML: Conceptualization, Data curation, Methodology, Visualization, Writing - original draft. EK: Methodology, Writing - review & editing. HT: Methodology, Writing - review & editing. JvV: Conceptualization, Methodology, Supervision, Writing - review & editing.

Declaration of interests

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.

Acknowledgements

ML would like to acknowledge funding from China Scholarship Council (Grant No 201706510011). JvV was further supported by the Netherlands Organization for Scientific Research NWO (Grant No VI.Vidi.198.008). All the authors thank the four anonymous reviewers for their insightful comments and critical reviews, which helped to improve our manuscript. This research contributes to the Global Land Programme (www.glp.earth).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2020.111859>.

References

- Angel, S., Parent, J., Civco, D.L., Blei, A., Potere, D., 2011. The dimensions of global urban expansion: estimates and projections for all countries, 2000–2050. *Prog. Plan.* 75, 53–107.
- van Asselen, S., Verburg, P.H., 2012. A land system representation for global assessments and land-use modeling. *Glob. Chang. Biol.* 18, 3125–3148.
- Badmos, O., Rienow, A., Callo-Concha, D., Greve, K., Jürgens, C., 2018. Urban development in West Africa—monitoring and intensity analysis of slum growth in Lagos: linking pattern and process. *Remote Sens.* 10, 1044.
- Bagan, H., Yamagata, Y., 2012. Landsat analysis of urban growth: how Tokyo became the world's largest megacity during the last 40 years. *Remote Sens. Environ.* 127, 210–222.

- Bagheri, H., Schmitt, M., d'Angelo, P., Zhu, X.X., 2018. A framework for SAR-optical stereogrammetry over urban areas. *ISPRS J. Photogramm. Remote Sens.* 146, 389–408.
- Barrington-Leigh, C., Millard-Ball, A., 2015. A century of sprawl in the United States. *Proc. Natl. Acad. Sci.* 112, 8244–8249.
- Berger, C., Rosentreter, J., Voltersen, M., Baumgart, C., Schmullius, C., Hese, S., 2017. Spatio-temporal analysis of the relationship between 2D/3D urban site characteristics and land surface temperature. *Remote Sens. Environ.* 193, 225–243.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Brunner, D., Lemoine, G., Bruzzone, L., Greidanus, H., 2010. Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique. *IEEE Trans. Geosci. Remote Sens.* 48, 1487–1504.
- Butler, J.J., Mills, S., Weiss, S., Liang, C., Xiong, X., Gu, X., 2013. VIIRS day/night band (DNB) stray light characterization and correction. In: *Earth Observing Systems XVIII*.
- Carlson, T., Sanchez-Azofeifa, G.A., 1999. Satellite remote sensing of land use changes in and around San Jose, Costa Rica. *Remote Sens. Environ.* 70, 247–256.
- Connors, J.P., Galletti, C.S., Chow, W.T.L., 2012. Landscape configuration and urban heat island effects: assessing the relationship between landscape characteristics and land surface temperature in Phoenix, Arizona. *Landscape Ecol.* 28, 271–283.
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raska, R., 2018. Deepglobe 2018: a challenge to parse the earth through satellite images. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, pp. 172–17209.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 248–255.
- Dong, P., Ramesh, S., Nepali, A., 2010. Evaluation of small-area population estimation using LiDAR, Landsat TM and parcel data. *Int. J. Remote Sens.* 31, 5571–5586.
- Dong, T., Jiao, L., Xu, G., Yang, L., Liu, J., 2019. Towards sustainability? Analyzing changing urban form patterns in the United States, Europe, and China. *Sci. Total Environ.* 671, 632–643.
- Dovey, K., Pafka, E., 2013. The urban density assemblage: modelling multiple measures. *Urban Des. Int.* 19, 66–76.
- Du, S., He, C., Huang, Q., Shi, P., 2018. How did the urban land in floodplains distribute and expand in China from 1992–2015? *Environ. Res. Lett.* 13, 034018.
- Ellis, E.C., Ramankutty, N., 2008. Putting people in the map: anthropogenic biomes of the world. *Front. Ecol. Environ.* 6, 439–447.
- Engelfriet, L., Koomen, E., 2017. The impact of urban form on commuting in large Chinese cities. *Transportation* 45, 1269–1295.
- ESA, 2019. The Sentinel-1 Toolbox. <https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1>.
- Esch, T., Marconcini, M., Felsbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., Dech, S., 2013. Urban footprint processor—fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. *IEEE Geosci. Remote Sens. Lett.* 10, 1617–1621.
- Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., Strano, E., 2017. Breaking new ground in mapping human settlements from space – the global urban footprint. *ISPRS J. Photogramm. Remote Sens.* 134, 30–42.
- Florczyk, A.J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., Sabo, F., Zanchetta, L., 2019. GHSL Data Package 2019. Luxembourg.
- Frolking, S., Milliman, T., Seto, K.C., Friedl, M.A., 2013. A global fingerprint of macro-scale changes in urban structure from 1999 to 2009. *Environ. Res. Lett.* 8, 024004.
- Geiß, C., Leichtle, T., Wurm, M., Pelizari, P.A., Standfuß, I., Zhu, X.X., So, E., Siedentop, S., Esch, T., Taubenböck, H., 2019. Large-area characterization of urban morphology—mapping of built-up height and density using TanDEM-X and Sentinel-2 data. *IEEE J-STARS* 12, 2912–2927.
- Glaeser, E., 2011. *Triumph of the City: How Urban Spaces Make us Human*. Pan Macmillan.
- Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., Zhou, Y., 2020. Annual maps of global artificial impervious area (GAIa) between 1985 and 2018. *Remote Sens. Environ.* 236, 111510.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27.
- Grace Wong, K.M., 2004. Vertical cities as a solution for land scarcity: the tallest public housing development in Singapore. *Urban Des. Int.* 9, 17–30.
- Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P.L., Frakias, M., Li, X., Seto, K.C., 2017. Global scenarios of urban density and its impacts on building energy use through 2050. *Proc. Natl. Acad. Sci.* 114, 8945–8950.
- Harun, S., Ogneva-Himmelberger, Y., 2013. Distribution of industrial farms in the United States and socioeconomic, health, and environmental characteristics of counties. *Geogr. J.* 2013.
- He, M., Zhu, Q., Du, Z., Hu, H., Ding, Y., Chen, M., 2016. A 3D shape descriptor based on contour clusters for damaged roof detection using airborne LiDAR point clouds. *Remote Sens.* 8.
- Hudeček, T., Hnilička, P., Dlouhý, M., Leňo Cutáková, L., Leňo, M., 2019. Urban structures, population density and municipal expenditures: an empirical study in the Czech Republic. *Urban Stud.* 56, 3450–3465.
- Kedron, P., Zhao, Y., Frazier, A.E., 2019. Three dimensional (3D) spatial metrics for objects. *Landscape Ecol.* 34, 2123–2132.
- Klotz, M., Kemper, T., Geiß, C., Esch, T., Taubenböck, H., 2016. How good is the map? A multi-scale cross-comparison framework for global settlement layers: evidence from Central Europe. *Remote Sens. Environ.* 178, 191–212.
- Koks, E.E., Haer, T., 2020. A high-resolution wind damage model for Europe. *Sci. Rep.* 10, 6866.
- Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., Serre, T., 2011. HMDB: a large video database for human motion recognition. In: *2011 International Conference on Computer Vision*. IEEE, pp. 2556–2563.
- Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., Levers, C., Lindner, M., Plutzar, C., Verkerk, P.J., van der Zanden, E.H., Reenberg, A., 2013. Challenges and opportunities in mapping land use intensity globally. *Curr. Opin. Environ. Sustain.* 5, 484–493.
- Kusno, A., 2019. Middling urbanism: the megacity and the kampung. *Urban Geogr.* 1–17.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Lee, T., Kim, T., 2013. Automatic building height extraction by volumetric shadow analysis of monoscopic imagery. *Int. J. Remote Sens.* 34, 5834–5850.
- Leichtle, T., Lakes, T., Zhu, X.X., Taubenböck, H., 2019. Has Dongying developed to a ghost city? – evidence from multi-temporal population estimation based on VHR remote sensing and census counts. *Comput. Environ. Urban. Syst.* 78, 101372.
- Lemonsu, A., Viguié, V., Daniel, M., Masson, V., 2015. Vulnerability to heat waves: impact of urban expansion scenarios on urban heat island and heat stress in Paris (France). *Urban Clim.* 14, 586–605.
- Li, M., van Vliet, J., Ke, X., Verburg, P.H., 2019. Mapping settlement systems in China and their change trajectories between 1990 and 2010. *Habitat Int.* 94, 102069.
- Li, X., Zhou, Y., Gong, P., Seto, K.C., Clinton, N., 2020. Developing a method to estimate building height from Sentinel-1 data. *Remote Sens. Environ.* 240, 111705.
- Liasis, G., Stavrou, S., 2016. Satellite images analysis for shadow detection and building height estimation. *ISPRS J. Photogramm. Remote Sens.* 119, 437–450.
- Lin, L., Homma, R., Iki, K., 2018. Preferences for a lake landscape: effects of building height and lake width. *Environ. Impact Assess. Rev.* 70, 22–33.
- Liu, Y., Fang, F., Li, Y., 2014. Key issues of land use in China and implications for policy making. *Land Use Policy* 40, 6–12.
- Ma, T., Zhou, C., Pei, T., Haynie, S., Fan, J., 2014. Responses of Suomi-NPP VIIRS-derived nighttime lights to socioeconomic activity in China's cities. *Rem. Sens. Lett.* 5, 165–174.
- Mahtta, R., Mahendra, A., Seto, K.C., 2019. Building up or spreading out? Typologies of urban growth across 478 cities of 1 million+. *Environ. Res. Lett.* 14, 124077.
- Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M.E., García-Santos, G., Fernandes, R., Berger, M., 2012. Sentinels for science: potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* 120, 91–101.
- Manoli, G., Fatichi, S., Schlapfer, M., Yu, K., Crowther, T.W., Meili, N., Burlando, P., Katul, G.G., Bou-Zeid, E., 2019. Magnitude of urban heat islands largely explained by climate and population. *Nature* 573, 55–60.
- Mathews, A.J., Frazier, A.E., Nghiem, S.V., Neumann, G., Zhao, Y., 2019. Satellite scatterometer estimation of urban built-up volume: validation with airborne lidar data. *Int. J. Appl. Earth Obs. Geoinf.* 77, 100–107.
- McFarlane, C., 2019. The urbanization of the sanitation crisis: placing waste in the city. *Dev. Chang.* 50, 1239–1262.
- Meijer, J., Huijbregts, M.A., Schotten, K., Schipper, A., 2018. Global patterns of current and future road infrastructure. *Environ. Res. Lett.* 13, 064006.
- Mertes, C.M., Schneider, A., Sulla-Menashe, D., Tatem, A.J., Tan, B., 2015. Detecting change in urban areas at continental scales with MODIS data. *Remote Sens. Environ.* 158, 331–347.
- Messerli, P., Heinemann, A., Epprecht, M., 2009. Finding homogeneity in heterogeneity—a new approach to quantifying landscape mosaics developed for the Lao PDR. *Hum. Ecol. Res.* 37, 291–304.
- Miles, R., Coutts, C., Mohamadi, A., 2012. Neighborhood urban form, social environment, and depression. *J. Urban Health* 89, 1–18.
- Mohammed, N.Z., Ghazi, A., Mustafa, H.E., 2013. Positional accuracy testing of Google earth. *Int. J. Multidiscipl. Sci. Eng.* 4, 6–9.
- Mushore, T.D., Odindi, J., Dube, T., Mutanga, O., 2017. Prediction of future urban surface temperatures using medium resolution satellite data in Harare metropolitan city, Zimbabwe. *Build. Environ.* 122, 397–410.
- Paprotny, D., Kreibich, H., Morales-Nápoles, O., Terefenko, P., Schröter, K., 2020. Estimating exposure of residential assets to natural hazards in Europe using open data. *Nat. Hazards Earth Syst. Sci.* 20, 323–343.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* 187, 156–168.
- Pulighe, G., Baiocchi, V., Lupia, F., 2015. Horizontal accuracy assessment of very high resolution Google Earth images in the city of Rome, Italy. *Int. J. Digital Earth* 9, 342–362.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven earth system science. *Nature* 566, 195–204.
- Schmitt, M., Hughes, L.H., Qiu, C., Zhu, X.X., 2019. SEN12MS—a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *arXiv preprint arXiv:1906.07789*.
- Schneider, A., Friedl, M.A., Potere, D., 2009. A new map of global urban extent from MODIS satellite data. *Environ. Res. Lett.* 4, 044003.
- Seto, K.C., Shepherd, J.M., 2009. Global urban land-use trends and climate impacts. *Curr. Opin. Environ. Sustain.* 1, 89–95.
- Seto, K.C., Sánchez-Rodríguez, R., Frakias, M., 2010. The new geography of contemporary urbanization and the environment. *Annu. Rev. Environ. Resour.* 35, 167–194.
- Soergel, U., Michaelsen, E., Thiele, A., Cadario, E., Thoennissen, U., 2009. Stereo analysis of high-resolution SAR images for building height estimation in cases of orthogonal

- aspect directions. *ISPRS J. Photogramm. Remote Sens.* 64, 490–500.
- Soomro, K., Zamir, A.R., Shah, M., 2012. UCF101: a dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402.
- Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* 93, 1879–1900.
- Straka, M., Sodoudi, S., 2019. Evaluating climate change adaptation strategies and scenarios of enhanced vertical and horizontal compactness at urban scale (a case study for Berlin). *Landsc. Urban Plan.* 183, 68–78.
- Sumbul, G., Charfuelan, M., Demir, B., Markl, V., 2019. BigEarthNet: a large-scale benchmark archive for remote sensing image understanding. arXiv preprint arXiv:1902.06148.
- Susaki, J., Kajimoto, M., Kishimoto, M., 2014. Urban density mapping of global megacities from polarimetric SAR images. *Remote Sens. Environ.* 155, 334–348.
- Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* 43, 1947–1958.
- Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring urbanization in mega cities from space. *Remote Sens. Environ.* 117, 162–176.
- Taubenböck, H., Kraft, N.J., Wurm, M., 2018. The morphology of the Arrival City - a global categorization based on literature surveys and remotely sensed data. *Appl. Geogr.* 92, 150–167.
- Tomás, L., Fonseca, L., Almeida, C., Leonardi, F., Pereira, M., 2015. Urban population estimation based on residential buildings volume using IKONOS-2 images and lidar data. *Int. J. Remote Sens.* 37, 1–28.
- Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E., Papale, D., 2015. Uncertainty analysis of gross primary production upscaling using random forests, remote sensing and eddy covariance data. *Remote Sens. Environ.* 168, 360–373.
- van Vliet, J., 2019. Direct and indirect loss of natural area from urban expansion. *Nat. Sustain.* 2, 755–763.
- van Vliet, J., Verburg, P.H., Grădinaru, S.R., Hersperger, A.M., 2019. Beyond the urban-rural dichotomy: towards a more nuanced analysis of changes in built-up land. *Comput. Environ. Urban. Syst.* 74, 41–49.
- van der Wal, D., Herman, P.M.J., Wielemaker-van den Dool, A., 2005. Characterisation of surface roughness and sediment texture of intertidal flats using ERS SAR imagery. *Remote Sens. Environ.* 98, 96–109.
- Wan, Z., Hook, S., Hulley, G., 2015. MOD11A2 MODIS/Terra land surface temperature/emissivity 8-day L3 global 1km SIN grid V006. In: NASA EOSDIS Land Processes DAAC. 10.
- Wang, Y., Liu, Y., Li, Y., Li, T., 2016. The spatio-temporal patterns of urban–rural development transformation in China since 1990. *Habitat Int.* 53, 178–187.
- Wang, L., Wang, S., Zhou, Y., Liu, W., Hou, Y., Zhu, J., Wang, F., 2018. Mapping population density in China between 1990 and 2010 using remote sensing. *Remote Sens. Environ.* 210, 269–281.
- Wang, J., Kuffer, M., Roy, D., Pfeffer, K., 2019a. Deprivation pockets through the lens of convolutional neural networks. *Remote Sens. Environ.* 234, 111448.
- Wang, Y., van Vliet, J., Pu, L., Verburg, P.H., 2019b. Modeling different urban change trajectories and their trade-offs with food production in Jiangsu Province, China. *Comput. Environ. Urban. Syst.* 77, 101355.
- Weiss, D.J., Nelson, A., Gibson, H.S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T.C.D., Howes, R.E., Tusting, L.S., Kang, S.Y., Cameron, E., Bisanzio, D., Battle, K.E., Bhatt, S., Gething, P.W., 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* 553, 333–336.
- Weissgerber, F., Colin-Koeniguer, E., Nicolas, J.-M., Trouvé, N., 2017. 3D monitoring of buildings using TerraSAR-X InSAR, DInSAR and PolSAR capacities. *Remote Sens.* 9, 1–24.
- Wellmann, T., Haase, D., Knapp, S., Salbach, C., Selsam, P., Lausch, A., 2018. Urban land use intensity assessment: the potential of spatio-temporal spectral traits with remote sensing. *Ecol. Indic.* 85, 190–203.
- Wentz, E.A., York, A.M., Alberti, M., Conrow, L., Fischer, H., Inostroza, L., Jantz, C., Pickett, S.T.A., Seto, K.C., Taubenböck, H., 2018. Six fundamental aspects for conceptualizing multidimensional urban form: a spatial mapping perspective. *Landsc. Urban Plan.* 179, 55–62.
- Xia, C., Yeh, A.G.-O., Zhang, A., 2020. Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: a case study of five Chinese megacities. *Landsc. Urban Plan.* 193, 103669.
- Yuan, F., Bauer, M.E., 2007. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. *Remote Sens. Environ.* 106, 375–386.
- Yue, W., Qiu, S., Xu, H., Xu, L., Zhang, L., 2019. Polycentric urban development and urban thermal environment: a case of Hangzhou, China. *Landsc. Urban Plan.* 189, 58–70.
- Zhang, W., Huang, B., 2015. Land use optimization for a rapidly urbanizing city with regard to local climate change: Shenzhen as a case study. *J. Urban Plan. Develop.* 141, 05014007.
- Zhang, W., Li, W., Zhang, C., Hanink, D.M., Liu, Y., Zhai, R., 2018. Analyzing horizontal and vertical urban expansions in three East Asian megacities with the SS-coMCRF model. *Landsc. Urban Plan.* 177, 114–127.
- Zhao, Y., Ovando-Montejo, G.A., Frazier, A.E., Mathews, A.J., Flynn, K.C., Ellis, E.A., 2017. Estimating world and home population using lidar-derived building volumes. *Int. J. Remote Sens.* 38, 1180–1196.
- Zhao, W., Duan, S.-B., Li, A., Yin, G., 2019. A practical method for reducing terrain effect on land surface temperature using random forest regression. *Remote Sens. Environ.* 221, 635–649.
- Zhou, C., Wang, Z., Chen, Q., Jiang, Y., Pei, J., 2014. Design optimization and field demonstration of natural ventilation for high-rise residential buildings. *Energy Build.* 82, 457–465.
- Zhou, B., Rybski, D., Kropp, J.P., 2017. The role of city size and urban form in the surface urban heat island. *Sci. Rep.* 7, 1–9.
- Zhu, Z., Zhou, Y., Seto, K.C., Stokes, E.C., Deng, C., Pickett, S.T.A., Taubenböck, H., 2019. Understanding an urbanizing planet: strategic directions for remote sensing. *Remote Sens. Environ.* 228, 164–182.