



Developing an annual building volume dataset at 1-km resolution from 2001 to 2019 in China

Wenting Yan, Jianping Wu, Chaoqun Zhang, Xiuzhi Chen, Jiashun Ren, Zhenzhen Xiao, Ziyin Liao, Raffaele Laforteza, Xueyan Li & Yongxian Su

To cite this article: Wenting Yan, Jianping Wu, Chaoqun Zhang, Xiuzhi Chen, Jiashun Ren, Zhenzhen Xiao, Ziyin Liao, Raffaele Laforteza, Xueyan Li & Yongxian Su (2024) Developing an annual building volume dataset at 1-km resolution from 2001 to 2019 in China, International Journal of Digital Earth, 17:1, 2330690, DOI: [10.1080/17538947.2024.2330690](https://doi.org/10.1080/17538947.2024.2330690)

To link to this article: <https://doi.org/10.1080/17538947.2024.2330690>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 25 Mar 2024.



Submit your article to this journal [↗](#)



Article views: 205



View related articles [↗](#)



View Crossmark data [↗](#)



Developing an annual building volume dataset at 1-km resolution from 2001 to 2019 in China

Wenting Yan^{a,b*}, Jianping Wu^{b*}, Chaoqun Zhang^{a,b}, Xiuzhi Chen^a, Jiashun Ren^{a,b}, Zhenzhen Xiao^b, Ziyin Liao^b, Raffaele Laforteza^c, Xueyan Li^b and Yongxian Su^b

^aGuangdong Province Data Center of Terrestrial and Marine Ecosystems Carbon Cycle, School of Atmospheric Sciences, Sun Yat-sen University (Zhuhai), Zhuhai, People's Republic of China; ^bGuangdong Provincial Key Lab of Remote Sensing and Geographical Information System, Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou Institute of Geography, Guangdong Academy of Sciences, Guangzhou, People's Republic of China; ^cDepartment of Agricultural and Environmental Sciences, University of Bari 'A. Moro', Bari, Italy

ABSTRACT

Urban vertical features are crucial for understanding urban morphology. However, long-term information on three-dimensional buildings, which are important fundamental data for studying on the historical urbanization processes, remains scarce in China. In this study, we proposed a Random Forest model to generate an annual 1-km resolution building volume dataset covering mainland China from 2001 to 2019, by integrating the nighttime light data, population demographics, electricity consumption records, carbon dioxide emissions data, and various optical and statistical datasets. This new building volume data are highly consistent with that derived from Baidu Maps on 1-km scale, with Pearson's correlation coefficient (R) of 0.847, root mean square error (RMSE) of $9.17 \times 10^5 \text{ m}^3/\text{km}^2$ and mean absolute error (MAE) of $5.86 \times 10^5 \text{ m}^3/\text{km}^2$. Notably, cross-validation indicate that the blooming problem was greatly improved when compared with previous model-based building three-dimensional data. The proposed method holds significant advantages, benefiting from low-cost implementation based on free open-source data and providing extendable algorithm to estimate the 3D shape of cities in the future. The time-series historical building volume data offer comprehensive insights into the historical development of urban structures, and provide valuable fundamental data for future urban planning, urban climate models and land use projections.

ARTICLE HISTORY



Received 8 December 2023
Accepted 10 March 2024

KEYWORDS

Urbanization; building volume; time-series information; China; machine learning; urban planning

1. Introduction

Urbanization has undergone a rapid ascent over the past century, with more than 50% of the world's population now residing in urban areas (He et al. 2023a). The 2018 Revision of World Urbanization Prospects suggest this upward trend will persist (Daes 2019), with an estimated increase to 68% by 2050 (United Nations 2017). Thus, the structure of cities (i.e. urban form) has been greatly changed, with rapid expansion in their horizontal boundary (urban area, Li

CONTACT Yongxian Su  suyongxian@gdas.ac.cn  Guangzhou Institute of Geography, Guangdong Academy of Sciences, No. 100 Xianliezhong Road, Yuexiu District, Guangzhou, Guangdong, People's Republic of China

*These authors contributed equally to this work.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

et al. 2020b) and increases in vertical boundary (building height, He et al. 2023b). These changes in urban three dimensional (3D) features may have significant impacts on socio-economic conditions, residents' living standards and public health, and energy consumption (Perini and Magliocco 2014; Borck 2016; Li and Zhou 2017); and may also triggered a cascade of adverse ecological effects (Alahmadi, Atkinson, and Martin 2013; Perini and Magliocco 2014). Therefore, it is important to track the long-term historical changes in global urban's 3D features.

Over the past decades, the horizontal urban boundaries has been well detected by determining the extent of impervious surface using multiple state-of-art satellite remote sensing technologies (Zhao et al. 2015; Leyk et al. 2019a; Heris et al. 2020; Gong et al. 2020) and statistical datasets (Frantz et al. 2021), based on machine learning methodologies (He et al. 2023b). However, comprehensively assessing or quantifying the urban form and morphology changes requires the assessment of vertical (3D) features within urban built-up areas (Gong et al. 2011; Li and Zhou 2017). The vertical features significantly influence urban environments such as solar radiation, wind pathways, pollution dispersion, and the urban heat island (UHI) effects (Clinton and Gong 2013; Rodríguez, Dupont-Courtade, and Oueslati 2016; Wang et al. 2018). Ignoring the urban height estimation, it's difficult to reflect the actual building density, land use intensity of urban areas and the heterogeneity of urban internal structure (Li et al. 2020a). Therefore, novel approaches are needed that surpass the detection of urban cover and can furnish consistent information regarding 3D traits that predict characteristics of urban areas at large scale.

Currently, extensive efforts have been dedicated to explore the 3D urban features at local and regional scales (Li et al. 2020c), using airborne light detection and ranging (LIDAR) (Yu et al. 2010; Bonczak and Kontokosta 2019) and other high-resolution optical remote sensing images (Thiele et al. 2007; Hao, Zhang, and Cao 2016). Given the outstanding capability of capturing 3D spatial information by LIDAR, some studies have directly employed it to model building height and morphology at both local and regional scales (Thiele et al. 2007; Yu et al. 2010; Hao, Zhang, and Cao 2016). However, obtaining such data is typically expensive and difficult, especially for large-scale and continuous estimation over large areas.

With the publication of open, free and globally available data, such as the Sentinel-1 Synthetic Aperture Radar (SAR) data and the Advanced Land Observing Satellite (ALOS) global digital surface model (DSM), increasing studies used a series of multivariate data to map large-scale vertical features based on machine learning methods (Li et al. 2020a; Frantz et al. 2021). Li et al. (2020a) applied a Random Forest (RF) models using Landsat, Moderate-resolution Imaging Spectroradiometer (MODIS), Sentinel-1, and other data collected from various sources to fill a gap in continental-scale and global-scale 3D building structures with a resolution of 1 km for 2015. Huang et al. (2022) devised a direct method using the Advanced Land Observing Satellite (ALOS) World 3D-30 m (AW3D30) Digital Surface Model (DSM) to estimate building height in China for 2010. Wu et al. (2023) generated the first Chinese building height map at 10-m resolution (CNBH-10 m) by integrating multiple-source data (radar, optical, and night light images) through a RF model for 2020. Despite these advancements, radar data sources such as Sentinel-1, commonly used for such analyses, mainly focus on specific time points after 2014, lacking the comprehensive long-term information that is crucial for understanding the continuous urban development process. Recently, He et al. (2023b) enriched time-series information on urban vertical features with a global 3D expansion dataset (1990–2010). They generated a building height dataset using the ALOS AW3D30 data version 2, global DSM data at 30-m resolution in 2010. Subsequently, annual global construction land data (1990–2010) were derived based on various sources, forming an original time-series 3D expansion dataset from 1990 to 2010. However, the dataset's endpoint is 2010 due to the limited availability of ALOS AW3D30 data.

Beyond these raster datasets, many technology companies' map products, such as Google Earth 3D, Baidu Map, and OpenStreetMap, generate rich building vector datasets worldwide with high spatial resolution (Daniel et al. 2019). However, obtaining the time-series data for national or global investigations is typically challenging and costly (He et al. 2023b), as the services they provide are

commercial and not entirely open (Yu and Gong 2012). Additionally, in certain cities, many important buildings are incomplete or inaccurate on the maps (Haklay and Weber 2008). What is more important, these datasets provide only limited temporal coverage and lack of the time-series data, particularly the historical data (He et al. 2023b). In summary, the majority of existing 3D city datasets either have restricted coverage or are not freely available. Thus, it is necessary to develop a time-series urban 3D feature dataset with complete coverage for all the cities.

To fill the gap of lacking of time-series historical 3D urban feature data, we aim to develop a long-term (from 2001 to 2009) building volume (including both horizontal and vertical building features) datasets covering all the cities in mainland China (excluding Hong Kong, Macao, and Taiwan) based on free open-source data at low cost quickly and automatically. The RF model has been demonstrated great potential in generating a spatial and temporal continuous gridded building volume dataset (Tramontana et al. 2015; Zhang et al. 2023). Thus, by combining the building volume data of 75 cities from Baidu Map in 2019 with 17 environmental covariates data from various sources spanning the year 2001 and 2019, we employed RF machine-learning model to derive a national map of building volume at a 1-km resolution covering the period from 2001 to 2019 in mainland China. The reference building volume data obtained from Baidu Map were spanned various city sizes (i.e. Small-sized, Medium-sized city, Large-sized city, Mega city and Super city), ensuring a diverse representation.

2. Data and methods

The workflow for estimating building volume in mainland China employing the RF model is illustrated in Figure 1. In the first, we collected the vector building data of 75 cities in 2019 from Baidu Map for training model and other 17 environmental covariates data from 2001 to 2019 for estimation of building volume. These environmental covariates data included economic-social, land cover-related and topography variables are sourced from satellite-based observations and statistical data (Table 1). All these data underwent preprocessing via the Google Earth Engine (GEE) platform and ArcGIS 10.8 (Figure 2), including unified resampling into 1 km-resolution annual data. Then, we created a spatially predictive RF model of building volume using the Baidu Map's data as dependent variable and environmental covariates in corresponding year as independent variables. Based on this RF model, we generated an annual average building volume dataset covering mainland China from 2001 to 2019. The third stage involved evaluating the accuracy of our estimated time-series data. Finally, we conducted a comprehensive analysis of the spatial and temporal characteristics derived from the estimated building volume dataset in mainland China.

2.1. Study area

Situated in eastern Asia along the western Pacific coast, China stands as the world's largest developing nation, hosting a significant population that accounts for about 20% of the global populace (Cao et al. 2023; Guo et al. 2023). Following the inception of free-market policies in 1978, China swiftly underwent urbanization (Yang et al. 2019; Cai, Liu, and Cao 2020), with approximately two-thirds of its population now residing in cities (China Statistics Press 2021). Notably, urban development in China exhibits considerable imbalances, which can be divided into five distinct sizes based on population size: Small-sized, Medium-sized, Large-sized, Mega, and Super cities (Song, Wang, and Qi 2015; Shi et al. 2021). In this work, the analysis focuses on cities in Mainland China, excluding Hong Kong, Macao, and Taiwan due to unavailability of statistical data.

2.2. Data collection and processing

This study used 3D vector building data from Baidu Map and other 17 environmental covariates data from multiple sources to create the national 1-km resolution 3D urban expansion dataset

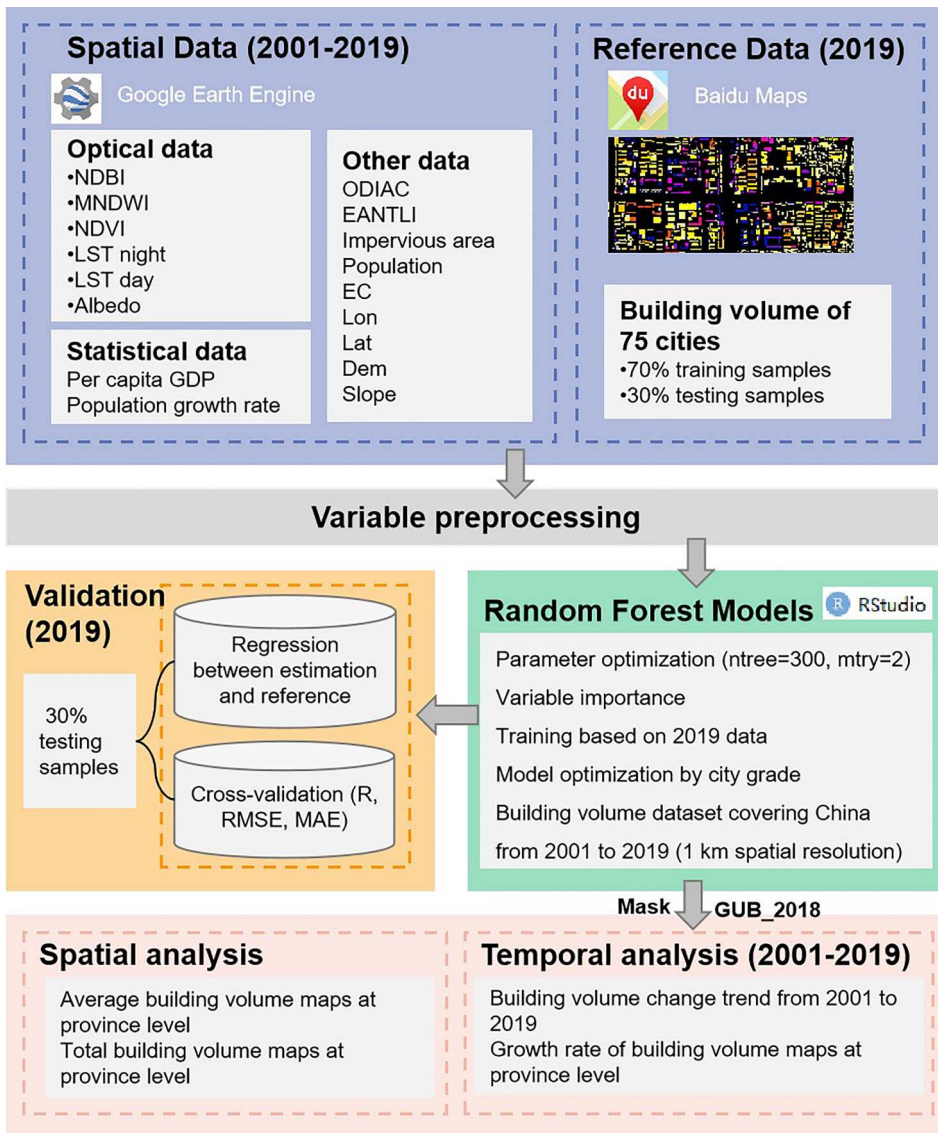


Figure 1. Methodological flowchart of building volume estimation from 2001 to 2019.

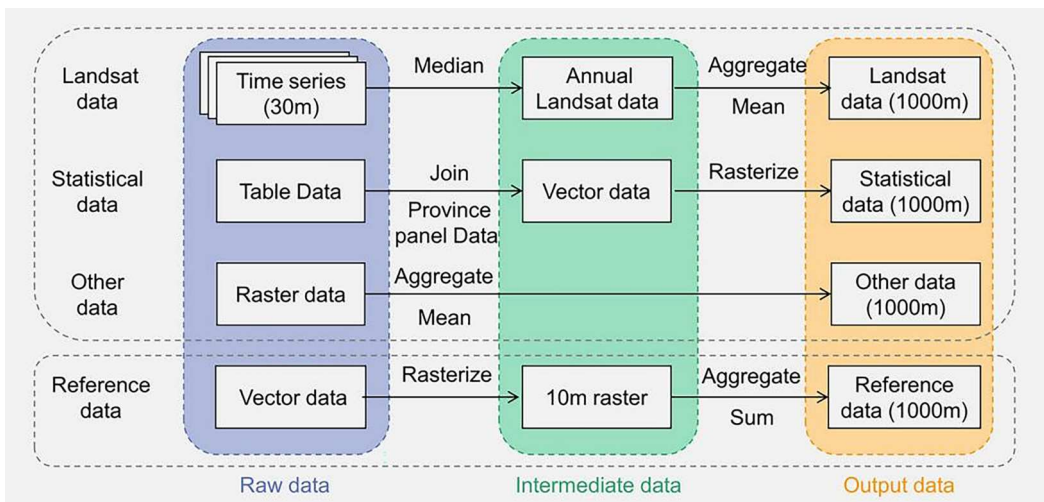
from 2001 to 2019. The detailed description of all data is shown in Table 1. Before building model, all these data were resampled to 1-km annual resolution as shown in Figure 2.

2.2.1. Building volume data used for train model

We downloaded the building data of 75 cities for the year in 2019 from Baidu Maps (<http://map.baidu.com>), one of the most widely used digital map platforms in China that offer satellite imagery, street maps, a route planner, and so on (Xue and Li 2020). In this study, the selected 75 cities include 2 Medium-sized cities, 24 Large-sized cities, 37 Mega cities, and 12 Super cities (Classification criteria from The State Council of the People's Republic of China 2014) (Figure 3(a)). The raw building data with vector format contain the information relating to the footprint and number of floors. We assumed a 3-m floor height to calculate the height of each building (Leichtle et al. 2019). Vector building data were converted into raster data with a 10-m spatial resolutions and calculated in a 1-

Table 1. Datasets used in this work.

Data	Type	Period	Spatial resolution	Source	Abbreviation
Reference building volume data	vector	2019	Vector	http://www.map.baidu.com	/
Nighttime light data	raster	2000–2020	1 km	https://doi.org/10.3974/geodb.2022.06.01.V1	EANTLI
Carbon dioxide emission data	raster	2000–2021	1 km	http://db.cger.nies.go.jp/dataset/ODIAC/	ODIAC
Electricity consumption data	raster	1992–2019	1 km	https://doi.org/10.6084/m9.figshare.19517272.v1	EC
Artificial impervious area data	raster	1990–2019	30 m	https://doi.org/10.5281/zenodo.4417810	AIA
Population data	raster	2000–2022	1 km	https://landscan.ornl.gov	Pop
Gross Domestic Product per capita data	table	2001–2022	Provincial	National Bureau of Statistics of China	GDP
Population growth rate data	table	2001–2022	Provincial	National Bureau of Statistics of China	PGR
Normalized Difference Built-up Index	raster	1984–2023	30 m	Landsat5; Landsat7; Landsat8	NDBI
Modified Normalized Difference Water Index	raster	1984–2023	30 m	Landsat5; Landsat7; Landsat8	MNDWI
Normalized Difference Vegetation Index	raster	1998–2019	1 km	https://www.resdc.cn/data.aspx?DATAID=257	NDVI
Albedo	raster	2000–2022	500 m	https://doi.org/10.5067/MODIS/MCD43A3.061	Albedo
LST Day	raster	2000–2022	1 km	https://doi.org/10.5067/MODIS/MOD11A2.006	LSTD
LST Night	raster	2000–2022	1 km	https://doi.org/10.5067/MODIS/MOD11A2.006	LSTN
Digital Elevation Model	raster	2000	90 m	http://www.earthenv.org/DEM.html	DEM
Slope	raster	2000	90 m	obtain based on DEM data	Slope
Latitude	raster	/	1 km	/	Lat
Longitude	raster	/	1 km	/	Lon
Global Urban Boundary data	vector	1990, 1995, 2000, 2005, 2010, 2015, 2018	/	http://data.ess.tsinghua.edu.cn	GUB

**Figure 2.** The flowchart of input dataset processing.

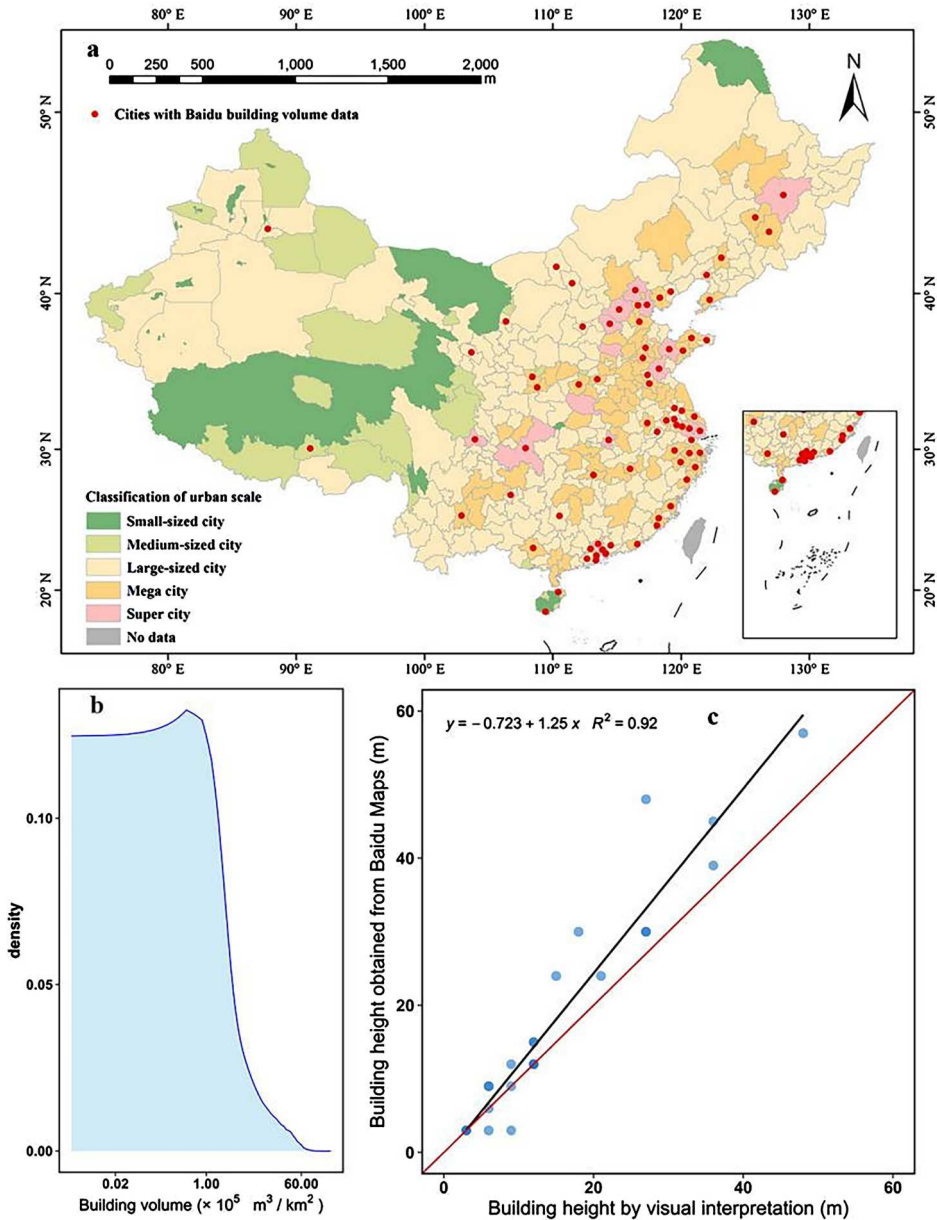


Figure 3. Building volume data in 2019 from Baidu Map for model development. (a) Geographical distribution of the selected 75 cities. Background colors indicating different city sizes based on population criteria (Small-sized city: population < 500,000; Medium-sized city: $500,000 \leq$ population < 1,000,000; Large-sized city: $1,000,000 \leq$ population < 5,000,000; Mega city: $5,000,000 \leq$ population < 10,000,000; Super city: population \geq 10,000,000). Red points denote the locations of the selected 75 cities. (b) Kernel density plot of building volume in each 1-km grid cell for the selected 75 cities. The x-axis scaled logarithmically. (c) Reliability analysis of building height by visual interpretation compared to reference data obtained from Baidu Maps.

km grid cell based on the method provided by Frantz et al. (2021) (Equation 1). Finally, 80,615 1-km grid cell samples were used for building model. The result showed that almost 80% of city had building volume more than $4394.4 \times 10^5 \text{ m}^3$ and almost 70% of the building volume in each 1-km grid cell for these 75 cities were large than $1 \times 10^5 \text{ m}^3/\text{km}^2$ (Figure 3(b)). To assess the reliability of reference samples, we used thirty buildings in Guangzhou to access the accuracy base on visual

interpretation approach, and these buildings has not changed since 2019. In Figure 3(c), the comparison shows very high reliability.

$$BV_{1\text{ km}} = \sum_f^n H_f A_f \quad (1)$$

where, $BV_{1\text{ km}}$ represents the total building volume of all buildings in a 1-km pixel; H_f and A_f denote the height and area of each reference building, respectively.

2.2.2. Socio-economic data

Close to 70% of global energy consumption is related to cities, and more than one-third of energy consumption stems from residential and commercial buildings (Gurney et al. 2009; Seto et al. 2014; Zhong et al. 2021). In this work, carbon dioxide (CO₂) emissions were taken from The Open-Data Inventory for Anthropogenic Carbon dioxide (ODIAC) (Oda, Maksyutov, and Andres 2018), which is a global monthly fossil fuel CO₂ emissions dataset with a 1-km horizontal resolution, from 2000 to 2021. It was generated through global power plant emissions estimates and satellite-observed nighttime light data (Oda, Maksyutov, and Andres 2018). We calculated yearly CO₂ emissions data based on each monthly data.

Electricity consumption is an indispensable commodity in the daily life of modern residents and provides useful information for understanding their economic status and quality of life (Shi et al. 2020; Cui et al. 2021). We used global yearly 1 km × 1 km gridded electricity consumption from 1992 to 2019, which was generated based on calibrated nighttime light data (Chen et al. 2022).

Population is an essential indicator for social and economic development and influences built-up planning and development (Fang et al. 2014; Calka, Nowak Da Costa, and Bielecka 2017; Leyk et al. 2019b). In this study, population size data were derived from the Oak Ridge National Laboratory (ORNL) (<https://landscan.ornl.gov>). The Landscan population dataset, first developed in 1998 and then annually from 2000 to 2019, consists of global gridded data with 1-km spatial resolution (ORNL 2019). It is based on statistical data from the Geographic Studies Branch, US Bureau of Census and geospatial input data, which contains land cover, roads, slope, urban areas, village locations, and high-resolution satellite imagery analysis (ORNL 2019).

Per capita gross domestic product (GDP) and population growth rate were obtained from provincial-level statistical data released by the National Bureau of Statistics of China. The processing of statistical data includes two steps. The initial data are in table format; thus, we first combined the information in the table with the provincial-level vector data in China. Secondly, the vector data were converted into raster data with a 1-km spatial resolution.

2.2.3. Nighttime light data

Nighttime light data reflects the brightness of the surface night light and represents the urbanization process and intensity of human activities (Levin et al. 2020; Sutton et al. 2001; Chen et al. 2023). Herein, we utilized nighttime light data corrected based on enhanced vegetation index (EANTLI), which provide yearly nighttime lighting information at a 1-km resolution in China from 2000 to 2020 (Zhong et al. 2022). This dataset was derived from a combination of data sources, including the Operational Linescan System (DMSP/OLS) from the Meteorological Satellite Program and Visible Infrared Imaging Radiometer (NPP/VIIRS) from the National Polar-orbiting Partnership.

2.2.4. Artificial impervious area

Built-up density is one of the most relevant measures in urban research (e.g. Taubenböck et al. 2016). In this work, built-up density denotes the share of a 1-km pixel that is occupied by artificial impervious surface that was extracted from the annual China land cover dataset (CLCD) (Yang and Huang 2021). The CLCD is multi-temporal, with 30-m resolution spanning from 1990 to 2019, and contains dynamic land cover change information. The calculated built-up density is expressed in

the following Equation 2:

$$BD_{1\text{ km}} = \frac{\sum_i^n 30\text{ m} \times 30\text{ m}}{1000\text{ m} \times 1000\text{ m}} \quad (2)$$

where $\sum_i^n 30\text{ m} \times 30\text{ m}$ denotes the total areas of artificial impervious surface in a 1-km pixel.

2.2.5. Optical data

Remotely sensed data consist of land surface temperature (LST), albedo, and relevant indices derived from Landsat. LST directly reflects surface features and thermal conditions (Du et al. 2016; Zhang, Murray, and Turner Li 2017; Gao et al. 2019). LST in urban built-up areas is affected by many factors, which include not only meteorological conditions and location but also land cover and urban form (Du et al. 2016; Zhang, Murray, and Turner Li 2017; Gao et al. 2019). LST data were obtained from the MOD11A2 Version 6 product, which provides an average 8-day period with a 1-km spatial resolution, where daytime and nighttime data are independently stored. The raw LST data extracted from a 1-year period were calculated as the yearly average LST data using the GEE platform.

The albedo with a range from 0 to 1 is the ratio of the reflected to the incident solar radiation over a horizontal plane, and represents the reflecting power of a surface (Rutherford et al. 2017). At the urban scale, albedo represents the ability of the urban surface features to reflect radiation back to the sky (Qin 2015; Salvati et al. 2022). Urban form and material reflectance are considered as influencing factors for albedo (Salvati et al. 2022). We used black sky albedo to denote real albedo, which is available for an 8-day period with a 500-m spatial resolution from MOD43A3 products. The raw (1-year) albedo data was computed as the yearly average albedo data. Subsequently, the albedo data were spatially aggregated to 1-km resolution using a mean function in the GEE platform.

The Normalized Difference Vegetation Index (NDVI), Modified Normalized Difference Water Index (MNDWI) and Normalized Difference Built-up Index (NDBI) were also considered as independent variables in our study. NDVI could provide additional features to effectively avoid over-estimation of built-up areas in urban cores (Zheng et al. 2023). Annual NDVI data were obtained using a 1-km spatial resolution from the Resource and Environment Science and Data Center of the Chinese Academy of Sciences (Xu 2018). MNDWI and NDBI were derived from Landsat images, which provide a 16-day period with a 30-m spatial resolution (Mushore et al. 2017). Firstly, we calculated the median of the raw Landsat data within one year, then aggregated the data into 990-m data using an average function and resampled to a 1-km resolution. Later, we computed MNDWI and NDBI based on Equations 3 and 4. All of the above processing was performed using the GEE platform.

$$\text{MNDWI} = \frac{\text{GREEN} - \text{SWIR1}}{\text{GREEN} + \text{SWIR1}} \quad (3)$$

$$\text{NDBI} = \frac{\text{SWIR1} - \text{NIR}}{\text{SWIR1} + \text{NIR}} \quad (4)$$

where GREEN, SWIR1, and NIR represent green, blue, near infrared, and shortwave infrared.

2.2.6. Geospatial data

Considering the complexity of the Chinese terrain, geospatial data are used as key variables for the estimation of building volume (Huang et al. 2022). The geospatial data include location information (longitude and latitude) and topographic information, Digital Elevation Model (DEM and slope). The DEM data were obtained from EarthEnv-DEM90 data, which is a global and free access product with 90-m spatial resolution (Robinson, Regetz, and Guralnick 2014). The slope data were

computed based on the DEM data. The data were aggregated into 990-m data using an average function and resampled to 1-km resolution. All of the processing is based on ArcGIS 10.8.

2.2.7. Urban boundary data

Similar to Zhou et al. (2022), global urban boundaries (GUB) in 2018 were used to mask the final map of building volume (Li et al. 2020b). The GUB dataset is the boundary of urban areas with vector format in 1990, 1995, 2000, 2005, 2010, 2015, and 2018. It is generated based on the global artificial impervious area (GAIA) data (Gong et al. 2020). Herein, the vector data were converted into raster data with 1-km spatial resolution.

2.2.8. Datasets for potential uncertainty analysis

To assess potential uncertainties of independent variables, we substituted four alternative datasets for carbon dioxide emissions, nighttime light data, impervious area, and population, which are crucial factors in our model for estimating building volume. The carbon dioxide emissions data was replaced with Particulate Matter 10 μm (PM10) data, a daily (1-km resolution) high-quality PM10 dataset in China spanning from 2015 to 2019 (Wei et al. 2021). Nighttime light data was replaced with a dataset obtained from Li et al. (2020d), offering a global dataset with integrated and consistent time-series (1992–2018) (Zhao et al. 2022). Impervious area was substituted with terrestrial Human Footprint data, reflecting the annual dynamics of the global human footprint from 2000 to 2018 (Mu et al. 2022). Population data was replaced with data obtained from Worldpop, an open dataset for spatial demography (Tatem 2017).

2.3. Random forest algorithms for developing the machine learning model

The RF regression model is a combined algorithm and generally demonstrates high accuracy (Breiman. 2001; Tramontana et al. 2015; Zhang et al. 2023). Currently, the RF regression model has received more attention for estimating vertical attributes (Li et al. 2020a; Potapov et al. 2021). We first split the total number of 1-km grid cells into a training set and a test set using an 70/30 random split, stratified by cities sizes. We used the training set to determine the best machine-learning algorithm and set of hyperparameters, and to train the final model. We used the test set to assess out-of-sample error. In our study, the ‘randomForest’ package in R language environment was used to set up the RF regression model and compute the relative importance of each independent variable (Breiman et al. 2018). To improve model accuracy, it was necessary to determine optimal values for two essential parameters: ntree and mtry (Ma et al. 2023). The values of ntree and mtry represent the number of decision trees and the number of variables included in each decision tree, respectively. Increasing the value of ntree can obtain better results, but the errors of the results stabilize when above a certain value (Guan et al. 2013). We found that the model error tended to be stable when the value of ntree was greater than 300. Subsequently, we attempted to apply each of the values of mtry from 2 to 17 to the model until we found the optimal value. Meanwhile, we also conducted a cross-validation to find the best combination of parameters. Therefore, after initial tuning experiments, we set the values of ntree, mtry, minimum node size, maximum depth and sample fraction to 300, 2, 5, 30 and 1, respectively (Figure 4).

2.4. Metrics for accuracy assessment

The performance of the building volume model was assessed based on three indicators – Pearson’s correlation coefficient (R), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). To test the stability of the model, the bootstrap sampling method was applied for cross-validation. The reference samples were randomly divided into two subsets based on the bootstrap sampling method. Seventy percent of the reference samples were used to train the building volume model.

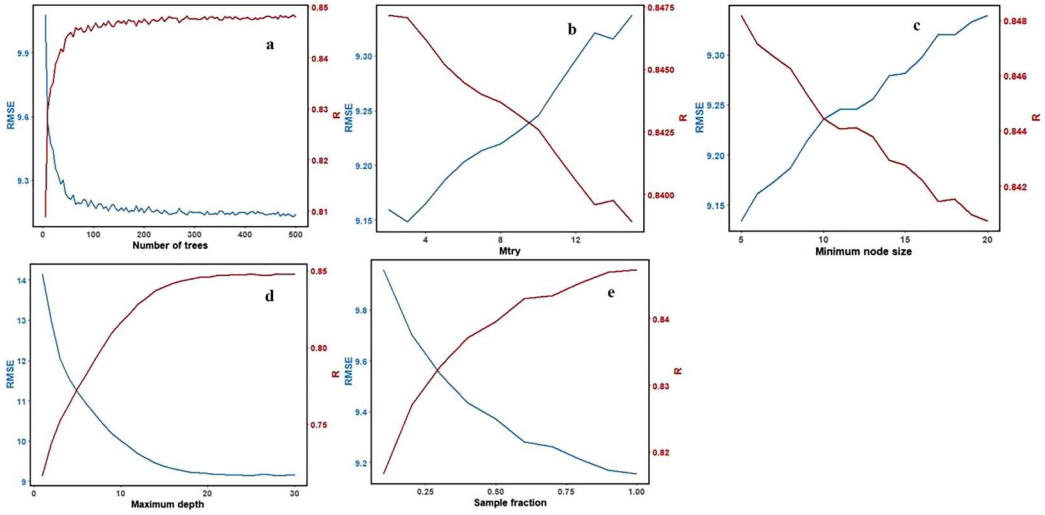


Figure 4. Precision of different parameters configurations during the modeling process. Red and blue lines represent the Pearson's correlation coefficient (R , red line) and the root mean square error (RMSE, blue line), respectively. The parameters include the number of decision trees (a), the number of variables included in each decision tree (b), minimum node size (c), maximum depth of decision trees (d), and sample fraction (e).

Subsequently, we computed the R , RMSE and MAE using the remaining samples (30%) (Eq. 5, Eq. 6 and Eq. 7). The process was repeated randomly 200 times.

In addition, both two one-year 3D building data provided by Wu et al. (2023) and Li et al. (2020a), respectively, and one time-series data produced by He et al. (2023b) were used to further assess model performance. For comparison with a referenced dataset at the same spatial resolution, we calculated the total building volume of annual reference data in each 1-km pixel based on ArcGIS 10.8. The GUB data from 2018 were used to mask the reference building volume maps. The final reference data were compared with our results from 2001 to 2010, which comprise of the two datasets from the overlaid period. In the Equations, n denotes the number of validation samples, $BV_{est,i}$ and $BV_{ref,i}$ represent the estimated building volume and reference building volume, respectively.

$$R = \sqrt{1 - \frac{(n-1) \sum_{i=1}^n (BV_{est,i} - BV_{ref,i})^2}{(n-2) \sum_{i=1}^n (BV_{est,i} - BV_{ref,i})^2}} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (BV_{est,i} - BV_{ref,i})^2}{n}} \quad (6)$$

$$MAE = \frac{\sum_{i=1}^n |BV_{est,i} - BV_{ref,i}|}{n} \quad (7)$$

3. Results and discussion

3.1. Model development, validations and uncertainty assessments

3.1.1. Relative importance of independent variables

The relative importance of each independent variable is shown in Figure 5(a). The independent variables with the most significant importance (top five) are ODIAC (13.1%), EANTLI (13.0%), impervious area (11.3%), population (10.9%), and electricity consumption (9.2%). The relative

importance of the remaining independent variables was less than 9%. In summary, socio-economic variables exhibited the highest importance for estimating building volume.

The regressions between the top five variables with high relative importance and reference building volume are shown in Figure 5(b-f). Based on the results of the regression, these variables expressed a highly positive correlation with reference building volume according to the coefficient of determination (R^2) values. The relationship between impervious area, electricity consumption and reference building volume conformed to a power function, and the remaining independent variables all presented a positive linear correlation with reference building volume (Figure 5(b-f)).

3.1.2. Cross-validation of model and model adjustment

The random cross-validation results of building volume are shown in Figure 6. We found that the estimated building volume showed a strong statistical correlation with the building volume from Baidu Map. The mean values of R , RMSE, and MAE were 0.847, $9.17 \times 10^5 \text{ m}^3/\text{km}^2$, and $5.86 \times 10^5 \text{ m}^3/\text{km}^2$, respectively (Figure 6(a-c)). However, the result also showed that the model performed a bit underestimate and different between four-scale cities. The mean R values in the Medium-sized city was 0.80, while that for Large-sized city, Mega city and Super city ranged from 0.844–0.852. In addition, the mean values of RMSE and MAE were higher in the larger scale city than in the smaller scale city. The lowest mean values of RMSE and MAE were $5.06 \times 10^5 \text{ m}^3/\text{km}^2$ and $3.83 \times 10^5 \text{ m}^3/\text{km}^2$, respectively, in the Medium-sized city, whereas the highest mean values of RMSE and MAE were $9.62 \times 10^5 \text{ m}^3/\text{km}^2$ and $6.35 \times 10^5 \text{ m}^3/\text{km}^2$, respectively, in the Super city (Figure 6(d-g)).

In order to further improve accuracy, we adjusted the building volume for four levels of cities separately according to the relationship of the scatter plot between model-based building volume and baidu-map-based building volume for four-scales cities. The initial estimated building volumes were compared with the adjusted building volumes, demonstrating that the regression line of adjusted building volumes was closer to the 1:1 line than the regression line of the initial estimated

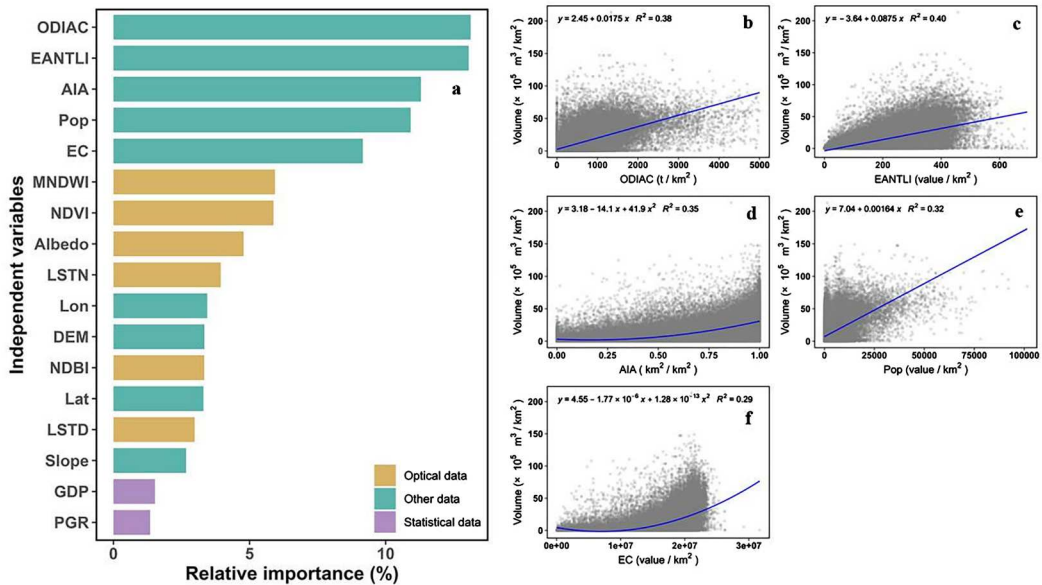


Figure 5. The relationship between independent and dependent variables in this RF model. (a) Average relative importance of each independent variable in influencing the building volume data. The variables used to estimate building volume are classified into optical data (yellow), statistical data (purple) and other data (green). (b–f) Scatter diagrams between building volume per square kilometer and CO₂ emissions data (ODLAC, b), nighttime light data (EANTLI, c), impervious area (AIA, d), population count (Pop, e), and electricity consumption (EC, f).

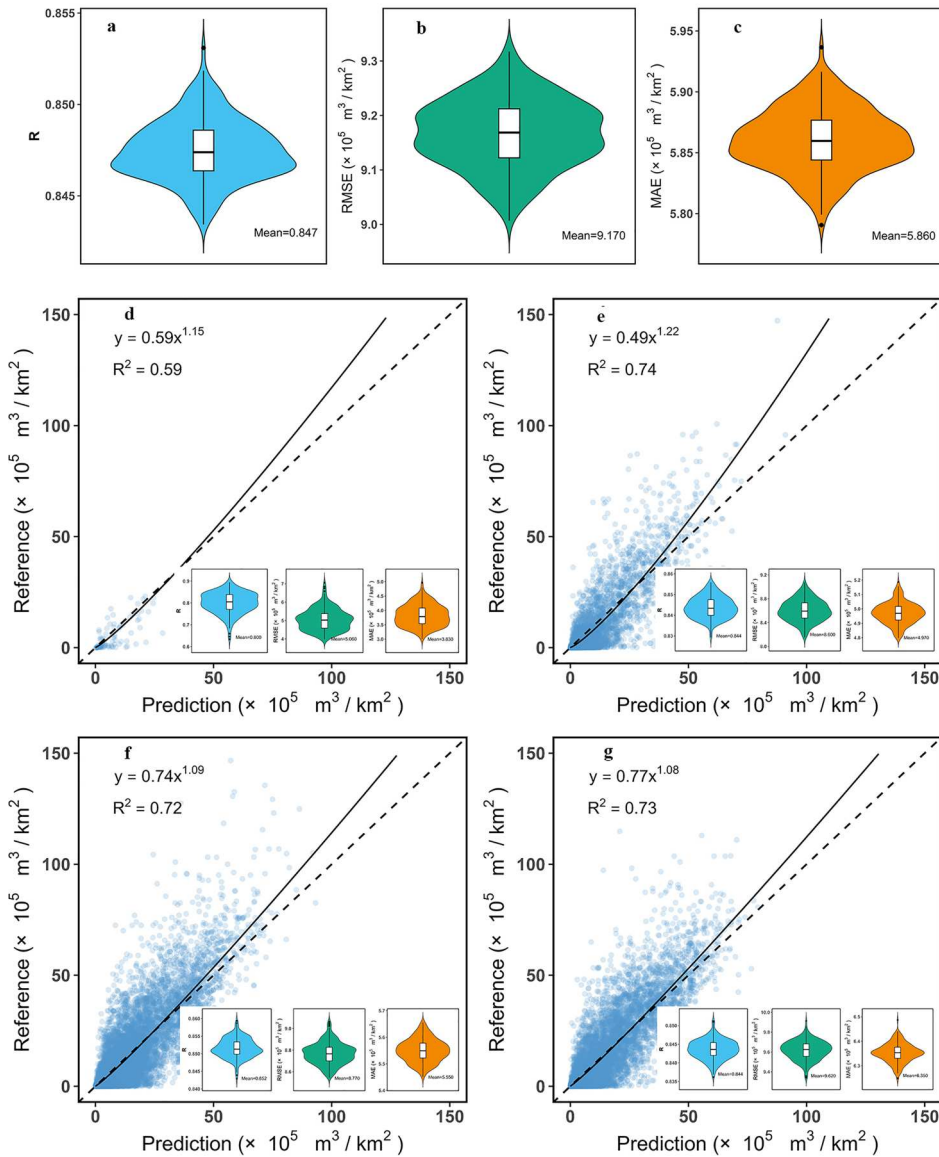


Figure 6. Random cross-validations of the estimated building volume using the Random Forest machine learning model. (a–c) The distribution of cross-validation results for building volume per square kilometer. Violin plots depict three performance metrics: Pearson's correlation coefficient (R) (a, blue), Root mean square error (RMSE) (b, green), and Mean absolute error (MAE) (c, orange) during cross-validations utilizing the bootstrap sampling method. Boxplots within each violin show the median, quartiles, and range of estimation accuracy, with black dots indicating outliers. (d–g) Relationship between building volume obtained from Baidu Map and estimated building volume in Medium-sized city (d), Large-sized city (e), Mega city (f), and Super city (g).

building volumes (Figure 7(a and b)). After adjusted, the mean values of RMSE and MAE for 75 cities declined from $8.65 \times 10^5 \text{ m}^3 / \text{km}^2$ to $8.55 \times 10^5 \text{ m}^3 / \text{km}^2$, $5.82 \times 10^5 \text{ m}^3 / \text{km}^2$ to $5.57 \times 10^5 \text{ m}^3 / \text{km}^2$, respectively. Specially, Xi'an ($1.1 \times 10^5 \text{ m}^3 / \text{km}^2$) and Lanzhou ($2.2 \times 10^5 \text{ m}^3 / \text{km}^2$) exhibit considerable of difference values between initial RMSE and adjusted RMSE. However, building volume was still evidently underestimated in Lanzhou (mean difference = -3.32) and overestimated in Guilin (mean difference = 2.56) (Figure 7(c) and Figure 8).

3.1.3. Potential uncertainty analysis

We estimated the building volume data using other machine learning algorithms such as multiple linear regression (MLR), support vector machines (SVM), artificial neural networks (ANN), decision trees (DT) models, and convolutional neural networks (CNNs) (Figure 9(a–e)). The results showed a strong linear correlation between the building volume data from Baidu Map and estimated building volume data from MLR ($R^2 = 0.55$), SVM ($R^2 = 0.67$), ANN ($R^2 = 0.62$), DT models ($R^2 = 0.61$), CNNs ($R^2 = 0.46$) and indicating the robustness of our model. RF might lead to overfitting due to the spatial autocorrelation problems (Ploton et al. 2020). In addition, we explored the inter-annual changes of built-up areas in China based on proportion of impervious areas within 1 km grid for 2018 and 2019, to examine the impacts of inner-annual dynamics on the estimation accuracy. The results showed that only 13.1% 1-km grid cells appeared slightly building volume change (difference of proportion between 2018 and 2019 > 0.5%), with average magnitude less than $0.0047 \text{ km}^2/\text{km}^2$ (Figure 9(f and g)). This indicated that the inter-annual variation of the built-up area is not obvious in most of pixels, suggesting a less significant impact from inner-annual changes of urban built-up areas.

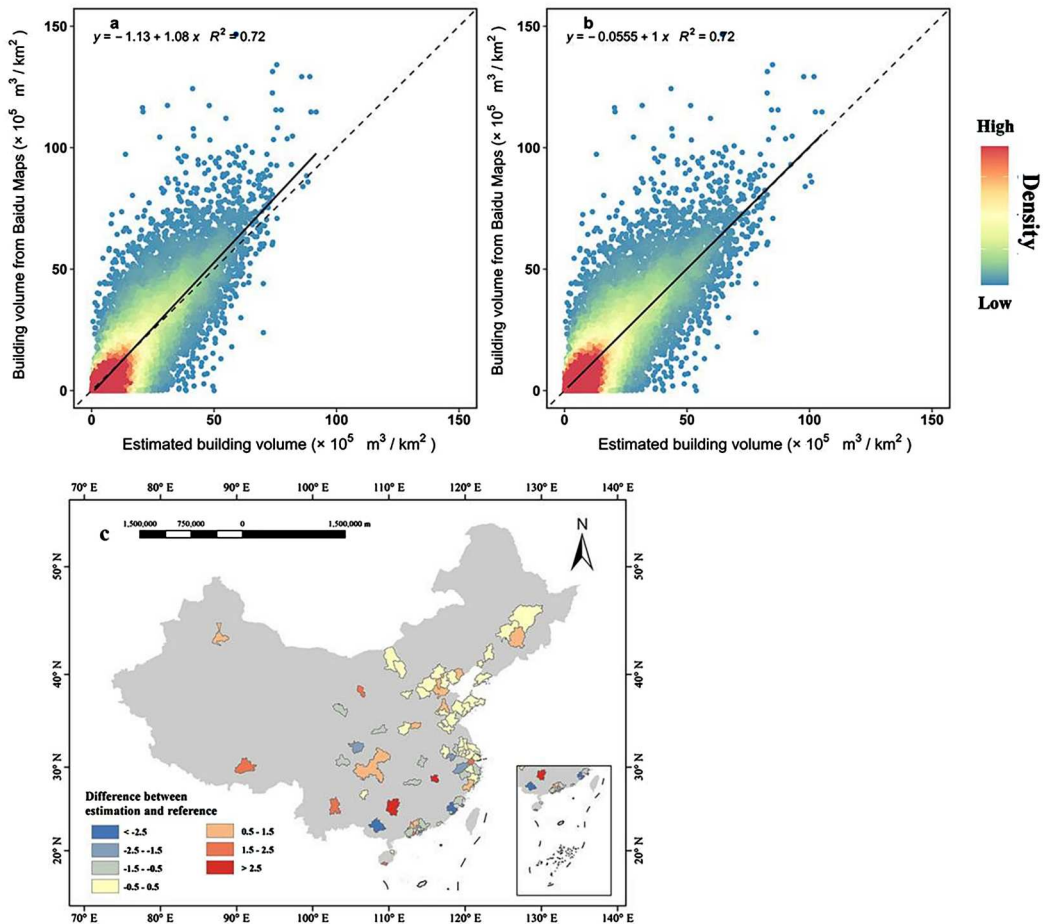


Figure 7. Validation of building volume per square kilometer on a left-out 30% sample of the training dataset. (a–b) Validation of the initial building volume (a) and the adjusted building volume (b) with the building volume obtained from Baidu Map. Solid and dotted lines represent the regressed line and the 1:1 line, respectively. (c) Inter-city comparison of mean difference between adjusted building volume and building volume obtained from Baidu Map across 75 cities.

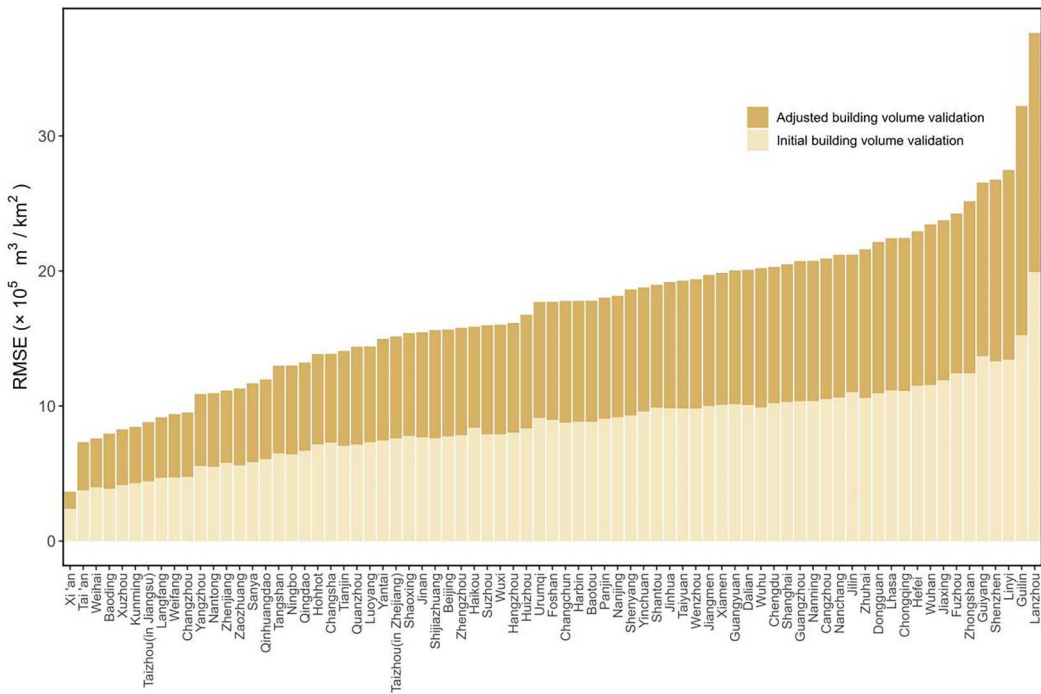


Figure 8. Accurate assessment of initial and adjusted building volume for 75 cities based on root mean square error (RMSE).

Meanwhile, we used four other datasets on carbon dioxide emissions, nighttime light data, impervious area and population, which are four most important factors in our model to estimate building volume (Figure 5(a)), to test potential uncertainties from independent variables in our model (Figure 9(h–k)). The result based on these four new datasets showed marginally small differences compared with those derived from original data, confirming the robustness of the model for estimating building volume.

Meyer et al. (2018) have revealed that it is an effective way to use a non-random validation method to check the potential overfitting problem that might existed in the RF estimations. Following Meyer’s method, we selected the data of 30, 35, 40, 45, 50, 55 and 60 cities to train the models. Then, we used the data from remaining cities for cross-validation (Figure 10(a–g)). This non-random validation could avoid the spatial autocorrelation problem. Results showed that the accuracy increases with the numbers of cities that used for training the models, the process was repeated randomly 10 times (Figure 10(h)). Thus, the estimation models, using 70% cities for projection, showed a high estimation accuracy (mean $R = 0.80$; mean $RMSE = 11.03 \times 10^5 \text{ m}^3/\text{km}^2$).

3.2. Comparison with existing 3D building products

To further assess the accuracy of estimated building volume in this study, a comparison was made with the 1-km building volume data of Li et al. (2020a), a study that simulated 3D building structure using RF models across China in 2015 (Figure 11(a and b)). The models they utilized contains both a combined and separate models. Our results reveal a strong linear relationship with those obtained from the combined ($R^2 = 0.58$) and separate ($R^2 = 0.68$) models. In comparison with the Baidu’s building volume, the mean R^2 value (0.72) of our model aligned closely with both the combined ($R^2 = 0.72$) and separate ($R^2 = 0.73$) models from Li et al. (2020a). What’s more, our model

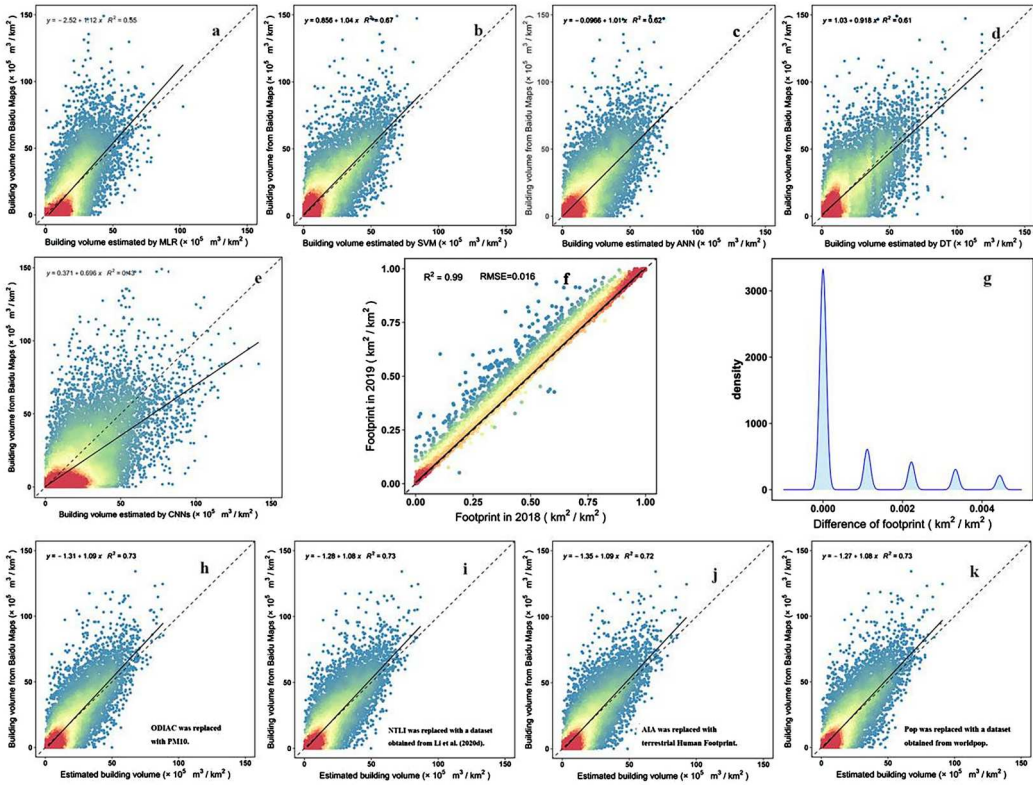


Figure 9. Uncertainty analysis of the estimated building volume data derived from random forest machine learning model. (a–e) Regression analysis between building volume obtained from Baidu Map and building volume predicted based on multiple linear regression model (MLR) (a), support vector machines model (SVM) (b), artificial neural networks model (ANN) (c), decision trees model (DT) (d) and convolutional neural networks (CNNs) (e). (f) Regression analysis between proportion of impervious area within 1-km pixel in 2018 and 2019. (f) Kernel density plot of difference between proportion between impervious areas proportion within 1-km pixel in 2019 and 2018. (h–k) Validation analysis between building volume obtained from Baidu Map and estimated building volume derived from our model trained by four replaced factors, i.e. the carbon dioxide emissions was replaced with PM10 data obtained from Wei et al. (2021) (h), nighttime light data was replaced with dataset obtained from Li et al. (2020d) (i), impervious area was replaced with terrestrial Human Footprint data obtained from Mu et al. (2022) (j), and population data was replaced with data obtained from Worldpop (k). Solid and dotted lines represent the regressed line and the 1:1 line, respectively.

showcased lower mean RMSE and MAE values ($RMSE = 9.17 \times 10^5 \text{ m}^3/\text{km}^2$; $MAE = 5.86 \times 10^5 \text{ m}^3/\text{km}^2$) in contrast to the combined ($RMSE = 11.07 \times 10^5 \text{ m}^3/\text{km}^2$; $MAE = 8.17 \times 10^5 \text{ m}^3/\text{km}^2$) and separate ($RMSE = 10.74 \times 10^5 \text{ m}^3/\text{km}^2$; $MAE = 8.09 \times 10^5 \text{ m}^3/\text{km}^2$) models (Li et al. 2020a). Specifically, the separate model tended to exhibit a more pronounced overestimation compared to the combined model. Notably, this overestimation was more evident in building volume with high values than that with low values (Figure 11(a)). Similarly, Huang et al. (2022) noted the separate model's overestimation in city centers when comparing with reference building height data. Specially, Figure 11(c) illustrates the spatial distribution across Beijing, Shanghai, and Guangzhou, comparing our building volume maps in 2015 with those derived by Li et al. (2020a) using both the combined and separate models. The spatial patterns in both sets of maps consistently revealed a decline in building volume from the urban center towards the suburbs. However, both the combined and separate models exhibited an overestimation trend in the suburbs (Figure 11(c)). Our model notably demonstrated more accurate estimations of suburb building volume than the model by Li et al. (2020a). This advantage might be attributed to our reference samples encompassing diverse city categories, including Medium-sized, Large-sized, Mega and Super cities, in contrast

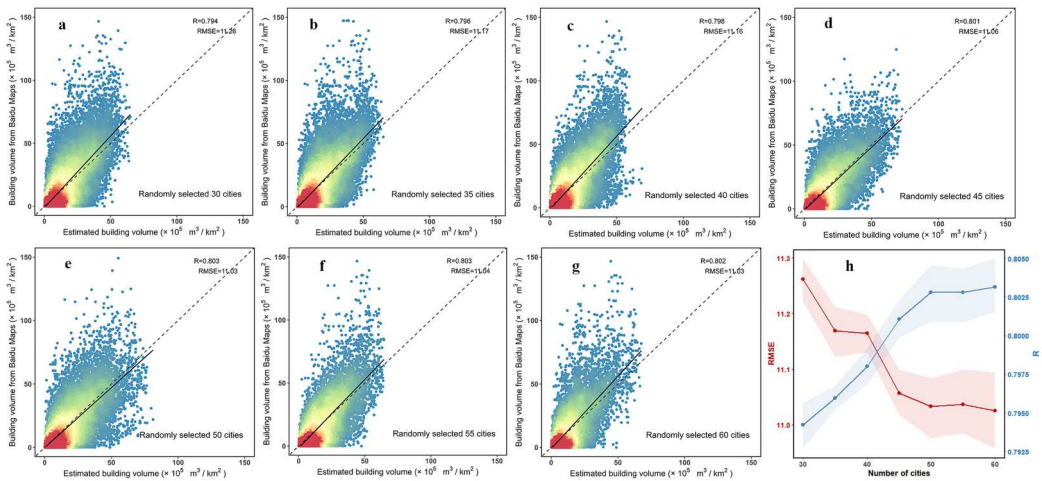


Figure 10. Assessments of the building volume estimations using non-random cross-validations. Data of the randomly selected 30 (a), 35 (b), 40 (c), 45 (d), 50 (e), 55 (f), and 60 (g) cities were used to train the RF model and the data of remaining 45 (a), 40 (b), 35 (c), 30 (d), 25 (e), 20 (f), and 15 (g) cities were used for cross-validations. (h) The relationship between the number of input city for training and the Pearson's correlation coefficient (R, blue line) and Root mean square error (RMSE, red line) of the estimation models.

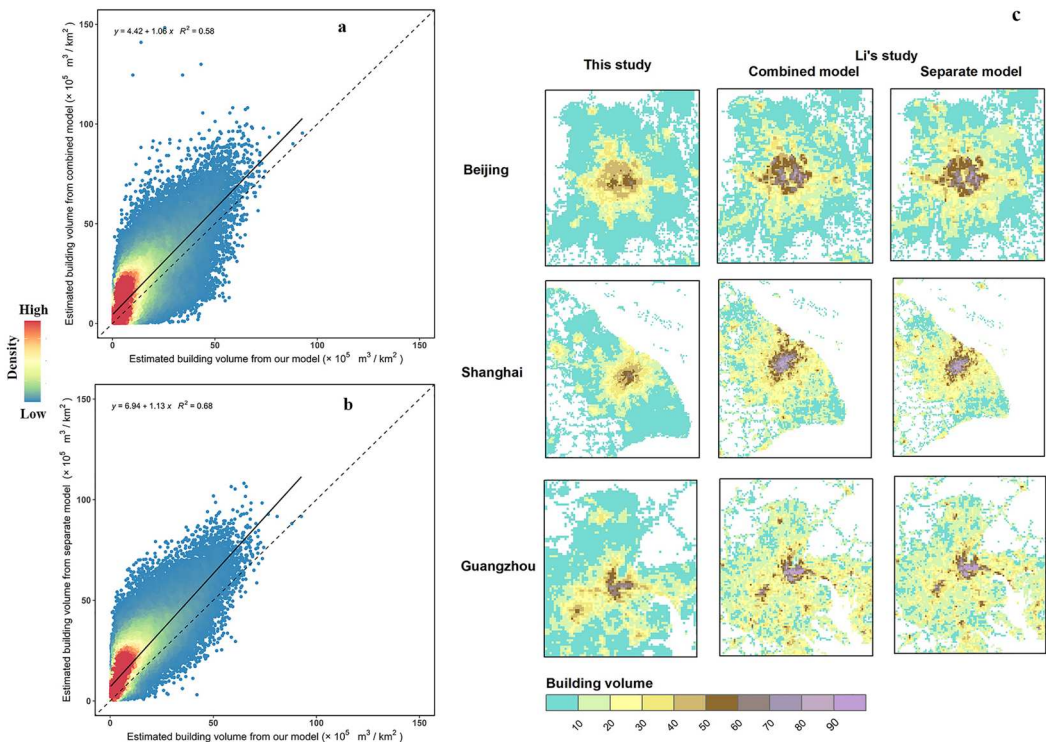


Figure 11. Cross-validations of estimated results from our model and other existing models. (a–b) Regression analysis of estimated building volume obtained from our random forest (RF) model with that from combined model (a) and separate model (b) of Li et al. (2020a). (c) Spatial diagram of estimated building volume in 2015 based on our model and the models of Li's study (2020a) in Beijing, Shanghai, and Guangzhou.

to the research by Li et al. (2020a), which considered only large cities, leading to overestimation of building volume. Moreover, our maps exhibited minimal pixels with building volumes exceeding $60 \times 10^5 \text{ m}^3/\text{km}^2$, hinting at a saturation issue observed in a few prior studies (Cao and Huang 2021; Frantz et al. 2021; Ma et al. 2023). In addition, various studies aiming to map building height at a larger scale have conducted regional analyses of total building volume at city level, which serve as a validation of our results. For instance, Cai et al. (2023) calculated the total building volume of Beijing ($6.29 \times 10^9 \text{ m}^3$) and Shanghai ($7.1 \times 10^9 \text{ m}^3$). Interestingly, their findings exhibited considerable consistency with our results (Beijing: $8.01 \times 10^9 \text{ m}^3$; Shanghai: $7.05 \times 10^9 \text{ m}^3$). This coherence reinforces the reliability of our estimations.

Meanwhile, we further compared our results with two other existing building heights at high spatial resolutions, which were converted into building volume data at 1-km resolution. One is a 30-meters spatiotemporal 3D urban expansion dataset from 1990 to 2010 provided by He et al.(2023b), which is generated using World Settlement Footprint 2015 data, GAIA data, and ALOS AW3D30 data. Another is building height estimate at 10-m resolution (CNBH-10 m) in 2020 using multi-source earth observations and machine learning (Wu et al. 2023). The maps from He et al. (2023b) at 30-m spatial resolution and Wu et al. (2023) at 10-m spatial resolution presented stronger blooming problem in comparison with Google Earth images (Figure 12(c-e)), resulting in considerable biases in the building volume estimation. By comparison, we found that the estimated building volume data of our study were lower than that from Wu et al. (2023)

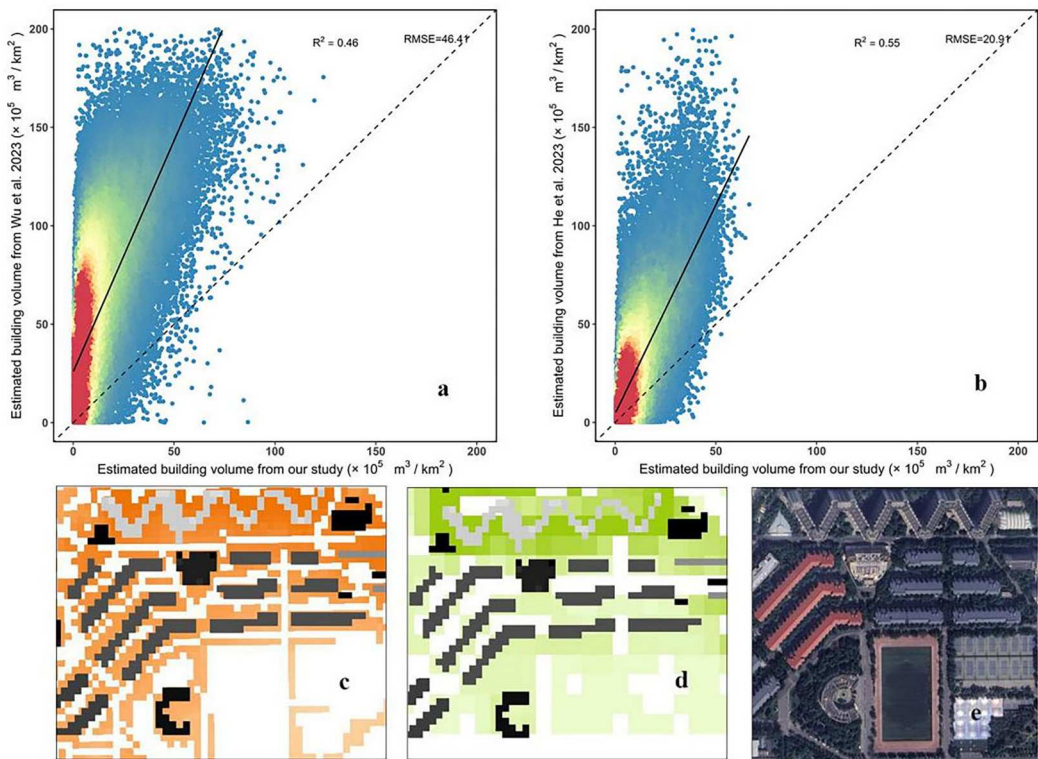


Figure 12. Comparison between estimated building volume obtained from our model and other existing products. (a–b) Regression analysis of estimated building volume obtained from our random forest model with that from Wu et al. (2023) in 2019 (a) and He et al. (2023b) in 2010 (b). (c–d) Spatial patterns of several building features in Beijing. The grid cells with orange color, green color, and gray color represent the spatial patterns of building based on Wu et al.'s product (10-m resolution), He et al.'s product (30-m resolution), and Baidu Map (10-m resolution), respectively. (e) Building patterns in Google Earth images. These buildings have remained consistent since 2010.

and He et al. (2023b) (Figure 12(a and b)), indicating that the blooming problem existed in both Wu and He's datasets were greatly improved in our data.

3.3. Spatial pattern of building volume of mainland China in 2019

The 1-km resolution building volume map covering China in 2019 is illustrated in Figure 13(a). The building volumes of each 1-km pixel of the map ranged from 0.09 to $118.15 \times 10^5 \text{ m}^3/\text{km}^2$, with a total building volumes of $1991497.2 \times 10^5 \text{ m}^3$ and building area of $2.68 \times 10^5 \text{ km}^2$. Figure 13(b and c) showcases the mean building volumes density and total building volumes covering mainland China at the provincial level. For the distribution of mean building volume, the provinces with the highest values are Chongqing ($15.18 \times 10^5 \text{ m}^3/\text{km}^2$), Sichuan province ($13.13 \times 10^5 \text{ m}^3/\text{km}^2$) and Shanghai ($12.66 \times 10^5 \text{ m}^3/\text{km}^2$). Contrarily, the provinces with the lowest mean building volume are Hebei ($3.84 \times 10^5 \text{ m}^3/\text{km}^2$), Shandong ($4.42 \times 10^5 \text{ m}^3/\text{km}^2$) and Henan province ($5.54 \times 10^5 \text{ m}^3/\text{km}^2$). The distribution of total building volume is obviously different from that of the mean building volume. Based on Figure 13(c), the provinces with the greater total building volume are primarily concentrated in the southeast coastal regions and Guangdong province. Specifically, the top five provinces with the highest total building volume are Guangdong ($194074.13 \times 10^5 \text{ m}^3$), Jiangsu ($169112.41 \times 10^5 \text{ m}^3$), Shandong ($134570.87 \times 10^5 \text{ m}^3$), Zhejiang ($118498.67 \times 10^5 \text{ m}^3$), and Liaoning province ($108163.45 \times 10^5 \text{ m}^3$).

We further analyzed the average and associated coefficients of variation (CV) of the estimated building volume for the five city categories, as shown in Figure 13(e and d). The CV value is the standard deviation divided by the mean and represents the degree of dispersion of the data near the mean value (Cottis 2021). The building volume in larger cities tended to exhibit higher CV and average values. It is evident that these cities often showcase greater building volumes characterized by high capacity and complex structural forms.

3.4. Dynamic characterization of building volume density from 2001 to 2019

The dynamics and trends in estimated building volume from 2001 to 2019 are displayed in Figure 14. The temporal changes of mean building volume density and across the five city categories in China from 1985 to 2019 are depicted in Figure 14(a and b), respectively. According to Figure 14(a), the mean building volume presents a clear upward trend from 2001 to 2019. In terms of

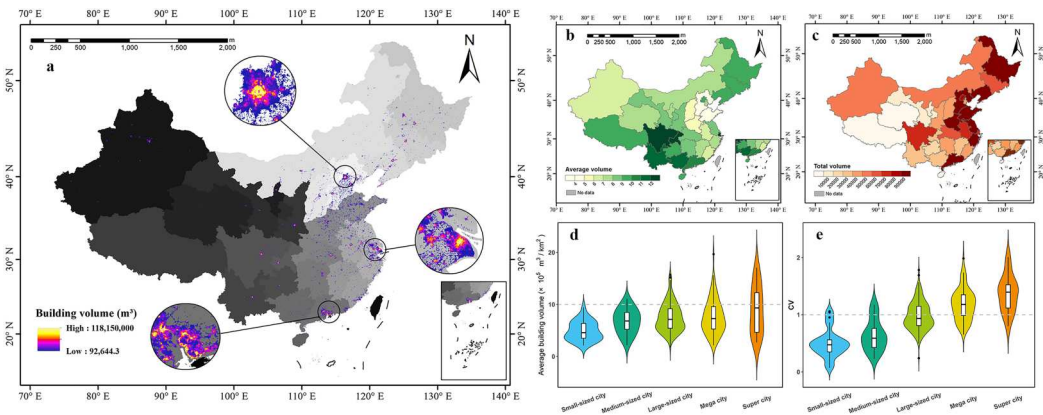


Figure 13. Spatial pattern of building volume of mainland China in 2019. (a) The 1-km resolution building volume map of mainland China. The different background gray categories represent different provinces in China. (b) Mean building volume density of provincial level in 2019. (c) Total building volume of provincial level in 2019. (d–e) Average building volume density (d) and Coefficients of variation (CV) of building volume density (e) for five city categories in 2019.

change in magnitude, the mean building volume first rapidly increased from 2001 to 2012, then the upward trend slightly fluctuated from 2013 to 2019. The temporal changes of building volume for the five city categories were similar to those in whole mainland China. The difference in time-series building volume for the five city categories was not evident, and the annual mean building volume in the super cities was higher than for the other cities (Figure 14(b)). Figure 14(c) showed that the provinces with the highest growth rate of mean building volume are mostly located in southern China. Specifically, the top three provinces with the highest growth rate of mean building volume are Chongqing (395.30%), Sichuan (296.26%) and Guangxi province (240.86%). In contrast, the provinces with the lowest growth rate of mean building volume are Shanxi (30.81%), Hebei (33.10%) and Beijing (42.70%) at the provincial level.

We further compared the dynamic characterization of building volume density from 2001 to 2019 between our study and He et al. (2023b). Figure 14(d) illustrates that although the building volume density estimated by He et al. (2023b) is much larger than that from our study, a consistent and clear upward trend from 2001 to 2010 in both our study and those of He et al. (2023b).

4. Advantages and limitations

Although previous studies have made significant contributions to research on urban height and three-dimensional structure estimation, there are still issues with either insufficient space coverage or insufficient time coverage (He et al. 2023b). In this study, we employed the Random Forest model to estimate China's annual building volume from 2001 to 2019 at a 1-km resolution, incorporating spatially and temporally continuous independent variables. Our estimations demonstrate reliability when compared to reference data from Baidu Maps and other existing products.

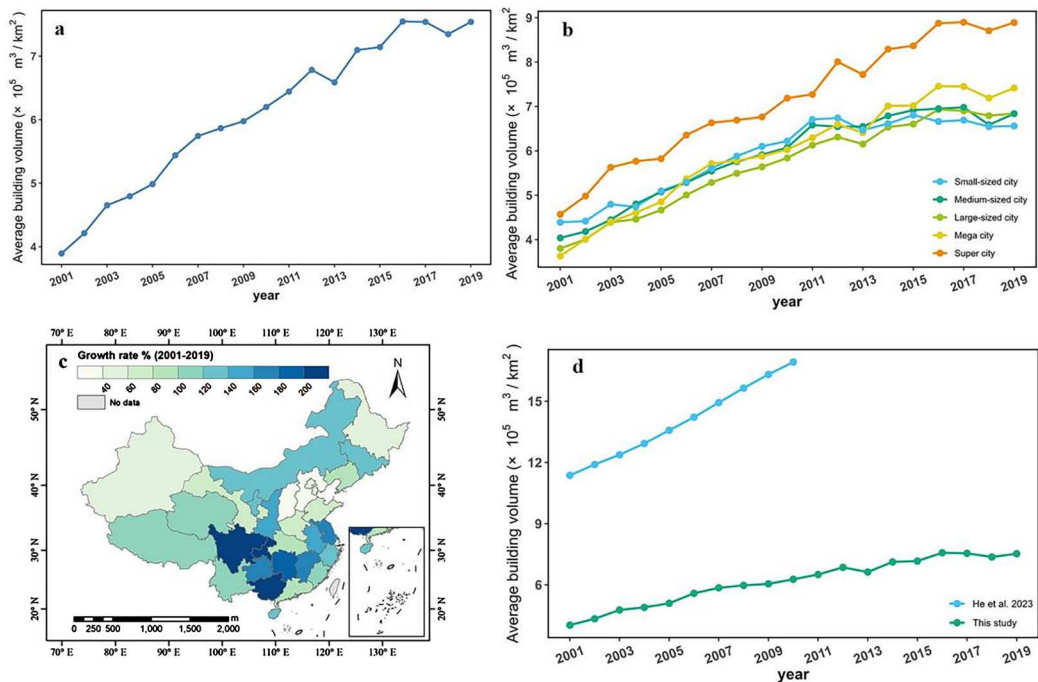


Figure 14. Dynamics of the estimated building volume density from 2001 to 2019. (a-b) Time-series building volume density from 2001 to 2019 for the whole mainland China (a) and five city categories (b). (c) Growth rate of building volume density from 2001 to 2019 at the provincial level. (d) Time-series building volume derived from our study and He et al.'s study (2023b) from 2001 to 2019.

In general, our study not only extends the spatiotemporal perspective in building volume in China but also enhances estimation accuracy. From the temporal perspective, we first generated the 1-km resolution time-series change data of urban 3D expansion from 2001 to 2019, covering the whole mainland China. Our estimated historical building volume offers comprehensive insights into the dynamic development of urban structures, proving invaluable for future urban planning, construction endeavors, and policy-making. Meanwhile, the results of this study can be utilized for further analysis of urban climate mechanisms, urban climate models, and land use projections. From the accuracy perspective, our method demonstrated higher consistency in the estimation of building volumes at a 1-km resolution when compared with previous urban building data, particularly aligning closely with the 3D building data provided by Baidu Maps. Notably, given that most of these high-precision 3D urban datasets, such as Baidu Maps, are not freely available due to their commercial properties, our method holds significant advantages, benefiting from low-cost implementation based on free open-source data and providing extendable algorithm to estimate the 3D shape of cities in the future.

However, there are a few limitations in the estimation of building volume and these need to be addressed in future studies. Although the applied RF model can well mitigate overfitting issues compared to other models, it can still occur and exhibit some noise. Future efforts to improve the accuracy and precision of estimation can be pursued through advanced methods. In addition, RF models are based on decision trees which are influenced by the input data. Biases from input data may have an impact on the results. Although we try to limit these biases by using multiple sets of alternative data (Figure 9(h–k)), it should be noted that the uncertainties are still possible. Important factors like land use policies, historical urbanization patterns, or socio-economic disparities at a more localized level might also have an effect on the three-dimensional expansion of individual city, but have not involved in our study. To fill this gap, in future studies, we will use local conditions to improve the accuracy of building structure estimation in different urban settings, and we are also committed to experimenting with a broader array of advanced techniques to enhance the precision and applicability of our study results.

Acknowledgements

We thank Raffaele Laforteza for his constructive suggestions on this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the National Natural Science Foundation of China [grant numbers 31971458, 41971275], Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) [grant number 311021009], ‘GDAS’ Project of Science and Technology Development [grant numbers 2020GDASYL-20200102002, 2022GDASZH-2022010105, 2023GDASQNR-0217], Science and Technology Program of Guangzhou, China [grant number 2024A04J3347].

Data availability statement

Reference building data are available at <http://map.baidu.com>. Nighttime light data are available at <https://doi.org/10.3974/geodb.2022.06.01.V1>. CO₂ emissions data are available at <http://db.cger.nies.go.jp/dataset/ODIAC/>. Electricity consumption data are available at <https://doi.org/10.6084/m9.figshare.19517272.v1>. Artificial impervious area data are available at <https://doi.org/10.5281/zenodo.4417810>. Population data are available at <https://landscan.ornl.gov>. GDP per capita and Population growth rate data are available at National Bureau of Statistics of China. Normalized Difference Vegetation Index are available at <https://www.resdc.cn/data.aspx?DATAID=257>. Modified Normalized Difference Water Index and Normalized Difference Built-up Index are available at Landsat5, Landsat7, Landsat8.

Albedo data are available at <https://doi.org/10.5067/MODIS/MCD43A3.061>. Land surface temperature data are available at <https://doi.org/10.5067/MODIS/MOD11A2.006>. Digital Elevation Model data are available at <http://www.earthenv.org/DEM.html>. Global urban boundaries data are available at <http://data.ess.tsinghua.edu.cn>.

Code availability

The codes that were used in this work are available upon author request.

Author contributions

Y.S. designed the study, proposed the scientific hypothesis and revised the manuscript. W.Y. collected and analyzed the data, drew the figures and wrote the manuscript. J.W. and X.C. performed the analyses and revised the manuscript. All the authors reviewed and revised the paper.

References

- Alahmadi, M., P. Atkinson, and D. Martin. 2013. "Estimating the Spatial Distribution of the Population of Riyadh, Saudi Arabia Using Remotely Sensed Built Land Cover and Height Data." *Computers, Environment and Urban Systems* 41: 167–176. <https://doi.org/10.1016/j.compenvurbsys.2013.06.002>.
- Bonczak, B., and C. E. Kontokosta. 2019. "Large-Scale Parameterization of 3D Building Morphology in Complex Urban Landscapes Using Aerial LiDAR and City Administrative Data." *Computers, Environment and Urban Systems* 73: 126–142. <https://doi.org/10.1016/j.compenvurbsys.2018.09.004>.
- Borck, R. 2016. "Will Skyscrapers Save the Planet? Building Height Limits and Urban Greenhouse Gas Emissions." *Regional Science and Urban Economics* 58: 13–25. <https://doi.org/10.1016/j.regsciurbeco.2016.01.004>.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., A. Cutler, A. Liaw, and M. Wiener. 2018. Package 'randomForest': Breiman and Cutler's Random Forests for Classification and Regression.
- Cai, Z., Q. Liu, and S. Cao. 2020. "Real Estate Supports Rapid Development of China's Urbanization." *Land use Policy* 95: 104582. <https://doi.org/10.1016/j.landusepol.2020.104582>.
- Cai, B., Z. Shao, X. Huang, X. Zhou, and S. Fang. 2023. "Deep Learning-Based Building Height Mapping Using Sentinel-1 and Sentinel-2 Data." *International Journal of Applied Earth Observation and Geoinformation* 122: 103399. <https://doi.org/10.1016/j.jag.2023.103399>.
- Calka, B., J. Nowak Da Costa, and E. Bielecka. 2017. "Fine Scale Population Density Data and Its Application in Risk Assessment." *Geomatics, Natural Hazards and Risk* 8 (2): 1440–1455. <https://doi.org/10.1080/19475705.2017.1345792>.
- Cao, Q., A. Hou, X. Li, and C. Jiang. 2023. "Spatial Impact of Government Venture Capital on Urbanization and Its Path—Evidence from the Yangtze River Delta Urban Agglomeration in China." *Land* 12 (1): 53. <https://doi.org/10.3390/land12010053>.
- Cao, Y., and X. Huang. 2021. "A Deep Learning Method for Building Height Estimation Using High-Resolution Multi-View Imagery Over Urban Areas: A Case Study of 42 Chinese Cities." *Remote Sensing of Environment* 264: 112590. <https://doi.org/10.1016/j.rse.2021.112590>.
- Chen, J., M. Gao, S. Cheng, W. Hou, M. Song, X. Liu, and Y. Liu. 2022. "Global 1 km×1 km Gridded Revised Real Gross Domestic Product and Electricity Consumption During 1992–2019 Based on Calibrated Nighttime Light Data." *Scientific Data* 9 (1): 202. <https://doi.org/10.1038/s41597-022-01322-5>.
- Chen, Z., S. Yu, X. You, C. Yang, C. Wang, J. Lin, W. Wu, and B. Yu. 2023. "New Nighttime Light Landscape Metrics for Analyzing Urban-Rural Differentiation in Economic Development at Township: A Case Study of Fujian Province, China." *Applied Geography* 150: 102841. <https://doi.org/10.1016/j.apgeog.2022.102841>.
- China Statistics Press. 2021. "China's 2020 1% Population Sample Survey Data." In, edited by Population Statistics Department of National Bureau of Statistics.
- Clinton, N., and P. Gong. 2013. "MODIS Detected Surface Urban Heat Islands and Sinks: Global Locations and Controls." *Remote Sensing of Environment* 134: 294–304. <https://doi.org/10.1016/j.rse.2013.03.008>.
- Cottis, R. A. 2021. "Electrochemical Noise for Corrosion Monitoring." *Techniques for Corrosion Monitoring*, 99–122. <https://doi.org/10.1016/B978-0-08-103003-5.00005-9>.
- Cui, W., J. Li, W. Xu, and B. Güneralp. 2021. "Industrial Electricity Consumption and Economic Growth: A Spatio-Temporal Analysis Across Prefecture-Level Cities in China from 1999 to 2014." *Energy* 222: 119932. <https://doi.org/10.1016/j.energy.2021.119932>.
- Daes. 2019. *World Urbanization Prospects: The 2018 Revision*. United Nations.

- Daniel, M., A. Lemonsu, M. Déqué, S. Somot, A. Alias, and V. Masson. 2019. “Benefits of Explicit Urban Parameterization in Regional Climate Modeling to Study Climate and City Interactions.” *Climate Dynamics* 52 (5-6): 2745–2764. <https://doi.org/10.1007/s00382-018-4289-x>.
- Du, S., Z. Pingong, Y. C. Wang, and L. Guo. 2016. “Quantifying the Multilevel Effects of Landscape Composition and Configuration on Land Surface Temperature.” *Remote Sensing of Environment* 178: 84–92. <https://doi.org/10.1016/j.rse.2016.02.063>.
- Fang, J., S. Sun, P. Shi, and J. A. Wang. 2014. “Assessment and Mapping of Potential Storm Surge Impacts on Global Population and Economy.” *International Journal of Disaster Risk Science* 5 (4): 323–331. <https://doi.org/10.1007/s13753-014-0035-0>.
- Frantz, D., F. Schug, A. Okujeni, C. Navacchi, W. Wagner, S. van der Linden, and P. Hostert. 2021. “National-Scale Mapping of Building Height Using Sentinel-1 and Sentinel-2 Time Series.” *Remote Sensing of Environment* 252: 112128. <https://doi.org/10.1016/j.rse.2020.112128>.
- Gao, M., F. Chen, H. Shen, M. Barlage, H. Li, Z. Tan, and L. Zhang. 2019. “Efficacy of Possible Strategies to Mitigate the Urban Heat Island Based on Urbanized High-Resolution Land Data Assimilation System (U-HRLDAS).” *Journal of the Meteorological Society of Japan. Ser. II* 97 (6): 1075–1097. <https://doi.org/10.2151/jmsj.2019-060>.
- Gong, P., Z. Li, H. Huang, G. Sun, and L. Wang. 2011. “ICESat GLAS Data for Urban Environment Monitoring.” *IEEE Transactions on Geoscience and Remote Sensing* 49 (3): 1158–1172. <https://doi.org/10.1109/TGRS.2010.2070514>.
- Gong, P., X. Li, J. Wang, Y. Bai, B. Chen, T. Hu, X. Liu, et al. 2020. “Annual Maps of Global Artificial Impervious Area (GAIA) Between 1985 and 2018.” *Remote Sensing of Environment* 236: 111510. <https://doi.org/10.1016/j.rse.2019.111510>.
- Guan, H., J. Li, M. Chapman, F. Deng, Z. Ji, and X. Yang. 2013. “Integration of Orthoimagery and Lidar Data for Object-Based Urban Thematic Mapping Using Random Forests.” *International Journal of Remote Sensing* 34 (14): 5166–5186. <https://doi.org/10.1080/01431161.2013.788261>.
- Guo, W., J. Liu, X. Zhao, W. Hou, Y. Zhao, Y. Li, W. Sun, and D. Fan. 2023. “Spatiotemporal Dynamics of Population Density in China Using Nighttime Light and Geographic Weighted Regression Method.” *International Journal of Digital Earth* 16 (1): 2704–2723. <https://doi.org/10.1080/17538947.2023.2233493>.
- Gurney, K. R., D. L. Mendoza, Y. Zhou, M. L. Fischer, C. C. Miller, S. Geethakumar, and S. de la Rue du Can. 2009. “High Resolution Fossil Fuel Combustion CO₂ Emission Fluxes for the United States.” *Environmental Science & Technology* 43 (14): 5535–5541. <https://doi.org/10.1021/es900806c>.
- Haklay, M., and P. Weber. 2008. “Openstreetmap: User-Generated Street Maps.” *IEEE Pervasive Computing* 7 (4): 12–18. <https://doi.org/10.1109/MPRV.2008.80>.
- Hao, L., Y. Zhang, and Z. Cao. 2016. “Building Extraction from Stereo Aerial Images Based on Multi-Layer Line Grouping with Height Constraint.” In *2016 IEEE International Geoscience and Remote Sensing Symposium*, 445–448. <https://doi.org/10.1109/IGARSS.2016.7729110>.
- He, T., Y. Lu, W. Yue, W. Xiao, X. Shen, and Z. Shan. 2023a. “A New Approach to Peri-Urban Area Land Use Efficiency Identification Using Multi-Source Datasets: A Case Study in 36 Chinese Metropolitan Areas.” *Applied Geography* 150: 102826. <https://doi.org/10.1016/j.apgeog.2022.102826>.
- He, T., K. Wang, W. Pingao, S. Xu, M. Li, R. Yang, and W. Yue. 2023b. “Global 30 Meters Spatiotemporal 3D Urban Expansion Dataset from 1990 to 2010.” *Scientific Data* 10 (1): 321. <https://doi.org/10.1038/s41597-023-02240-w>.
- Heris, M. P., N. L. Foks, K. J. Bagstad, A. Troy, and Z. H. Ancona. 2020. “A Rasterized Building Footprint Dataset for the United States.” *Scientific Data* 7 (1): 207. <https://doi.org/10.1038/s41597-020-0542-3>.
- Huang, H., P. Chen, X. Xu, C. Liu, J. Wang, C. Liu, N. Clinton, and P. Gong. 2022. “Estimating Building Height in China from ALOS AW3D30.” *ISPRS Journal of Photogrammetry and Remote Sensing* 185: 146–157. <https://doi.org/10.1016/j.isprsjprs.2022.01.022>.
- Leichtle, T., T. Lakes, X. Zhu, and H. Taubenboeck. 2019. “Has Dongying Developed to a Ghost City?-Evidence from Multi-Temporal Population Estimation Based on VHR Remote Sensing and Census Counts.” *Computers, Environment and Urban Systems* 78: 101372. <https://doi.org/10.1016/j.compenvurbsys.2019.101372>.
- Levin, N., C. C. Kyba, Q. Zhang, A. S. de Miguel, M. O. Román, X. Li, B. A. Portnov, et al. 2020. “Remote Sensing of Night Lights: A Review and an Outlook for the Future.” *Remote Sensing of Environment* 237: 111443. <https://doi.org/10.1016/j.rse.2019.111443>.
- Leyk, S., D. Balk, B. Jones, M. R. Montgomery, and H. Engin. 2019a. “The Heterogeneity and Change in the Urban Structure of Metropolitan Areas in the United States, 1990–2010.” *Scientific Data* 6 (1): 321. <https://doi.org/10.1038/s41597-019-0329-6>.
- Leyk, S., A. E. Gaughan, S. B. Adamo, A. de Sherbinin, D. Balk, S. Freire, A. Rose, et al. 2019b. “The Spatial Allocation of Population: A Review of Large-Scale Gridded Population Data Products and Their Fitness for Use.” *Earth System Science Data* 11 (3): 1385–1409. <https://doi.org/10.5194/essd-11-1385-2019>.
- Li, X., P. Gong, Y. Zhou, J. Wang, Y. Bai, B. Chen, T. Hu, et al. 2020b. “Mapping Global Urban Boundaries from the Global Artificial Impervious Area (GAIA) Data.” *Environmental Research Letters* 15 (9): 094044. <https://doi.org/10.1088/1748-9326/ab9be3>.

- Li, M., E. Koks, H. Taubenböck, and J. van Vliet. 2020a. "Continental-Scale Mapping and Analysis of 3D Building Structure." *Remote Sensing of Environment* 245: 111859. <https://doi.org/10.1016/j.rse.2020.111859>.
- Li, X., and Y. Zhou. 2017. "Urban Mapping Using DMSP/OLS Stable Night-Time Light: A Review." *International Journal of Remote Sensing* 38 (21): 6030–6046. <https://doi.org/10.1080/01431161.2016.1274451>.
- Li, X., Y. Zhou, P. Gong, K. C. Seto, and N. Clinton. 2020c. "Developing a Method to Estimate Building Height from Sentinel-1 Data." *Remote Sensing of Environment* 240: 111705. <https://doi.org/10.1016/j.rse.2020.111705>.
- Li, X. C., Y. Y. Zhou, M. Zhao, and X. Zhao. 2020d. "A Harmonized Global Nighttime Light Dataset 1992–2018." *Scientific Data* 7: 168. <https://doi.org/10.1038/s41597-020-0510-y>.
- Ma, X., G. Zheng, X. Chi, L. Yang, Q. Geng, J. Li, and Y. Qiao. 2023. "Mapping Fine-Scale Building Heights in Urban Agglomeration with Spaceborne Lidar." *Remote Sensing of Environment* 285: 113392. <https://doi.org/10.1016/j.rse.2022.113392>.
- Meyer, H., C. Reudenbach, T. Hengl, M. Katurji, and T. Nauss. 2018. "Improving Performance of Spatio-Temporal Machine Learning Models Using Forward Feature Selection and Target-Oriented Validation." *Environmental Modelling & Software* 101: 1–9. <https://doi.org/10.1016/j.envsoft.2017.12.001>.
- Mu, H. W., X. C. Li, Y. N. Wen, J. X. Huang, P. J. Du, W. Su, S. X. Miao, and M. Q. Geng. 2022. "A Global Record of Annual Terrestrial Human Footprint Dataset from 2000 to 2018." *Scientific Data* 9: 176. <https://doi.org/10.1038/s41597-022-01284-8>.
- Mushore, T. D., J. Odindi, T. Dube, and O. Mutanga. 2017. "Prediction of Future Urban Surface Temperatures Using Medium Resolution Satellite Data in Harare Metropolitan City, Zimbabwe." *Building and Environment* 122: 397–410. <https://doi.org/10.1016/j.buildenv.2017.06.033>.
- Oda, T., S. Maksyutov, and R. J. Andres. 2018. "The Open-Source Data Inventory for Anthropogenic CO₂, Version 2016 (ODIAC2016): A Global Monthly Fossil Fuel CO₂ Gridded Emissions Data Product for Tracer Transport Simulations and Surface Flux Inversions." *Earth System Science Data* 10 (1): 87–107. <https://doi.org/10.5194/essd-10-87-2018>.
- ORNL. 2019. Documentation. <https://landscan.ornl.gov/documentation>.
- Perini, K., and A. Magliocco. 2014. "Effects of Vegetation, Urban Density, Building Height, and Atmospheric Conditions on Local Temperatures and Thermal Comfort." *Urban Forestry & Urban Greening* 13 (3): 495–506. <https://doi.org/10.1016/j.ufug.2014.03.003>.
- Ploton, P., F. Mortier, M. Réjou-Méchain, N. Barbier, N. Picard, V. Rossi, C. Dormann, et al. 2020. "Spatial validation reveals poor predictive performance of large-scale ecological mapping models." *Nature communications* 11 (1): 4540. <https://doi.org/10.1038/s41467-020-18321-y>.
- Potapov, P., X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, et al. 2021. "Mapping Global Forest Canopy Height Through Integration of GEDI and Landsat Data." *Remote Sensing of Environment* 253: 112165. <https://doi.org/10.1016/j.rse.2020.112165>.
- Qin, Y. 2015. "Urban Canyon Albedo and Its Implication on the Use of Reflective Cool Pavements." *Energy and Buildings* 96: 86–94. <https://doi.org/10.1016/j.enbuild.2015.03.005>.
- Robinson, N., J. Regetz, and R. P. Guralnick. 2014. "EarthEnv-DEM90: A Nearly-Global, Void-Free, Multi-Scale Smoothed, 90 m Digital Elevation Model from Fused ASTER and SRTM Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 57–67. <https://doi.org/10.1016/j.isprsjprs.2013.11.002>.
- Rodríguez, M. C., L. Dupont-Courtade, and W. Oueslati. 2016. "Air Pollution and Urban Structure Linkages: Evidence from European Cities." *Renewable and Sustainable Energy Reviews* 53: 1–9. <https://doi.org/10.1016/j.rser.2015.07.190>.
- Rutherford, W., T. Painter, S. Ferrenberg, J. Belnap, G. Okin, C. Flagg, and S. Reed. 2017. "Albedo Feedbacks to Future Climate via Climate Change Impacts on Dryland Biocrusts." *Science Report* 7: 44188. <https://doi.org/10.1038/srep44188>.
- Salvati, A., M. Kolokotroni, A. Kotopouleas, R. Watkins, R. Giridharan, and M. Nikolopoulou. 2022. "Impact of Reflective Materials on Urban Canyon Albedo, Outdoor and Indoor Microclimates." *Building and Environment* 207: 108459. <https://doi.org/10.1016/j.buildenv.2021.108459>.
- Seto, K. C., S. Dhakal, A. Bigio, H. Blanco, G. Carlo Delgado, D. Dewar, L. Huang, et al. 2014. "Human Settlements, Infrastructure, and Spatial Planning."
- Shi, Y., X. Ren, K. Guo, Y. Zhou, and J. Wang. 2020. "Research on the Economic Development Pattern of Chinese Counties Based on Electricity Consumption." *Energy Policy* 147: 111881. <https://doi.org/10.1016/j.enpol.2020.111881>.
- Shi, K., J. Shen, Y. Wu, S. Liu, and L. Li. 2021. "Carbon Dioxide (CO₂) Emissions from the Service Industry, Traffic, and Secondary Industry as Revealed by the Remotely Sensed Nighttime Light Data." *International Journal of Digital Earth* 14 (11): 1514–1527. <https://doi.org/10.1080/17538947.2021.1946605>.
- Song, Q., Y. Wang, and Y. Qi. 2015. "Study on Present Status of Carbon Emissions in China's Low-Carbon Pilot Cities." *China Population Resources and Environment* 25 (1): 78–82.
- The State Council of the People's Republic of China. 2014. "GB/T 000014349/2014-00135." Circular of the State Council on Adjusting the Criteria for the Division of City Scale.

- Sutton, P., D. Roberts, C. Elvidge, and K. Baugh. 2001. "Census from Heaven: An Estimate of the Global Human Population Using Night-Time Satellite Imagery." *International Journal of Remote Sensing* 22 (16): 3061–3076. <https://doi.org/10.1080/01431160010007015>.
- Tatem, A. 2017. "WorldPop, Open Data for Spatial Demography." *Scientific Data* 4: 170004. <https://doi.org/10.1038/sdata.2017.4>.
- Taubenböck, H., I. Standfuß, M. Klotz, and M. Wurm. 2016. "The Physical Density of the City—Deconstruction of the Delusive Density Measure with Evidence from Two European Megacities." *ISPRS International Journal of Geo-Information* 5 (11): 206. <https://doi.org/10.3390/ijgi5110206>.
- Thiele, A., E. Cadario, K. Schulz, U. Thonnessen, and U. Soergel. 2007. "Building Recognition from Multi-Aspect High-Resolution InSAR Data in Urban Areas." *IEEE Transactions on Geoscience and Remote Sensing* 45 (11): 3583–3593. <https://doi.org/10.1109/TGRS.2007.898440>.
- Tramontana, G., K. Ichii, G. Camps-Valls, E. Tomelleri, and D. Papale. 2015. "Uncertainty Analysis of Gross Primary Production Upscaling Using Random Forests, Remote Sensing and Eddy Covariance Data." *Remote Sensing of Environment* 168: 360–373. <https://doi.org/10.1016/j.rse.2015.07.015>.
- United Nations. 2017. "World Urbanization Prospects: The 2018 Revision." <https://population.un.org/wpp/>.
- Wang, W., Y. Xu, E. Ng, and S. Raasch. 2018. "Evaluation of Satellite-Derived Building Height Extraction by CFD Simulations: A Case Study of Neighborhood-Scale Ventilation in Hong Kong." *Landscape and Urban Planning* 170: 90–102. <https://doi.org/10.1016/j.landurbplan.2017.11.008>.
- Wei, J., Z. Li, W. Xue, L. Sun, T. Fan, L. Liu, T. Su, and M. Cribb. 2021. "The ChinaHighPM10 Dataset: Generation, Validation, and Spatiotemporal Variations from 2013 to 2019 Across China." *Environment International* 146: 106290. <https://doi.org/10.1016/j.envint.2020.106290>.
- Wu, W. B., J. Ma, E. Banzhaf, M. E. Meadows, Z. W. Yu, F. X. Guo, D. Sengupta, X. X. Cai, and B. Zhao. 2023. "A First Chinese Building Height Estimate at 10 m Resolution (CNBH-10 m) Using Multi-Source Earth Observations and Machine Learning." *Remote Sensing of Environment* 291: 113578. <https://doi.org/10.1016/j.rse.2023.113578>.
- Xu, X. 2018. *China Annual Vegetation Index (NDVI) Spatial Distribution Dataset*. Beijing: Resource and Environment Science and Data Center.
- Xue, Y., and C. Li. 2020. "Extracting Chinese Geographic Data from Baidu Map API." *The Stata Journal* 20 (4): 805–811. <https://doi.org/10.1177/1536867X20976313>.
- Yang, J., and X. Huang. 2021. "The 30 m Annual Land Cover Dataset and Its Dynamics in China from 1990 to 2019." *Earth System Science Data* 13 (8): 3907–3925. <https://doi.org/10.5194/essd-13-3907-2021>.
- Yang, Y., J. Liu, Y. Lin, and Q. Li. 2019. "The Impact of Urbanization on China's Residential Energy Consumption." *Structural Change and Economic Dynamics* 49: 170–182. <https://doi.org/10.1016/j.strueco.2018.09.002>.
- Yu, L., and P. Gong. 2012. "Google Earth as a Virtual Globe Tool for Earth Science Applications at the Global Scale: Progress and Perspectives." *International Journal of Remote Sensing* 33 (12): 3966–3986. <https://doi.org/10.1080/01431161.2011.636081>.
- Yu, B., H. Liu, J. Wu, Y. Hu, and L. Zhang. 2010. "Automated Derivation of Urban Building Density Information Using Airborne LiDAR Data and Object-Based Method." *Landscape and Urban Planning* 98 (3-4): 210–219. <https://doi.org/10.1016/j.landurbplan.2010.08.004>.
- Zhang, Y., A. T. Murray, and B. L. Turner II. 2017. "Optimizing Green Space Locations to Reduce Daytime and Nighttime Urban Heat Island Effects in Phoenix, Arizona." *Landscape and Urban Planning* 165: 162–171. <https://doi.org/10.1016/j.landurbplan.2017.04.009>.
- Zhang, H., J. Peng, R. Wang, M. Zhang, C. Gao, and Y. Yu. 2023. "Use of Random Forest Based on the Effects of Urban Governance Elements to Forecast CO₂ Emissions in Chinese Cities." *Heliyon*. <https://doi.org/10.1016/j.heliyon.2023.e16693>.
- Zhao, C., X. Cao, X. Chen, and X. Cui. 2022. "A Consistent and Corrected Nighttime Light Dataset (CCNL 1992–2013) from DMSP-OLS Data." *Scientific Data* 9 (1): 424. <https://doi.org/10.1038/s41597-022-01540-x>.
- Zhao, S., D. Zhou, C. Zhu, W. Qu, J. Zhao, Y. Sun, D. Huang, W. Wu, and S. Liu. 2015. "Rates and Patterns of Urban Expansion in China's 32 Major Cities Over the Past Three Decades." *Landscape Ecology* 30 (8): 1541–1559. <https://doi.org/10.1007/s10980-015-0211-7>.
- Zheng, Q., K. C. Seto, Y. Zhou, S. You, and Q. Weng. 2023. "Nighttime Light Remote Sensing for Urban Applications: Progress, Challenges, and Prospects." *ISPRS Journal of Photogrammetry and Remote Sensing* 202: 125–141. <https://doi.org/10.1016/j.isprsjprs.2023.05.028>.
- Zhong, X., M. Hu, S. Deetman, B. Steubing, H. X. Lin, G. A. Hernandez, C. Harpprecht, C. Zhang, A. Tukker, and P. Behrens. 2021. "Global Greenhouse Gas Emissions from Residential and Commercial Building Materials and Mitigation Strategies to 2060." *Nature Communications* 12 (1): 6126. <https://doi.org/10.1038/s41467-021-26212-z>.
- Zhong, X., Q. Yan, F. Li, M. Li, and G. Li. 2022. "Long Time Series Nighttime Light Dataset of China (2000–2020)." *Digital Journal of Global Change Data Repository* 6 (3): 416–424. <https://doi.org/10.3974/geodb.2022.06.01.V1>.
- Zhou, Y., X. Li, W. Chen, L. Meng, Q. Wu, P. Gong, and K. C. Seto. 2022. "Satellite Mapping of Urban Built-Up Heights Reveals Extreme Infrastructure Gaps and Inequalities in the Global South." *Proceedings of the National Academy of Sciences* 119 (46): e2214813119. <https://doi.org/10.1073/pnas.2214813119>.