A Simple and Reliable Method for Estimating Building-Scale Height Based on Multisource Datasets

Yangzi Che, Xuecao Li, Qian Shi, *Senior Member, IEEE*, and Xiaoping Liu, *Member, IEEE*

***Abstract*—Building height dataset is crucial in urban studies, holding significant importance in understanding the interactions of human activities and the built-up environment. However, high-resolution three-dimensional building datasets covering large areas are limited. A rapid and accurate method for revealing fine-scale urban morphology is required. In this study, we developed a method for estimating building heights at the building scale. First, we integrated multisource datasets (i.e., synthetic aperture radar, optical, terrain, social-economic, and vector-based datasets) and built the machine learning model for building height estimation. Second, we applied the model to 11 cities in the U.S. and assessed its performance. Our results were consistent with the reference data, indicating that the effectiveness of our method is applicable [i.e., the R^2 was 0.82, and the root mean square error (RMSE) was 3.39 m]. The evaluated results in various cities, across different height intervals, and within distinct regions also show the good agreement with reference heights according to the correlation (R^2: 0.51–0.86, RMSE: 2.57–5.97 m in cities) and similar height distribution. Moreover, our results also showed the superiority by comparing with other height datasets at different scales. Finally, we mapped the building-scale height to characterize the urban morphology. These results demonstrate our proposed method's usable accuracy and the vast application potential in estimating building heights. Our proposed method's refined building height information can significantly help socioeconomic and climatological urban studies.**

***Index Terms*—Building height, high-resolution, multisource datasets, urban morphology, urbanization.**

I. Introduction

THREE-DIMENSIONAL (3-D) morphology of buildings significantly influenced socio-economic indicators during the process of urbanization. A comprehensive and quantitative understanding of buildings is instrumental in achieving the United Nations' 2023 Sustainable Development Agenda, particularly the objectives related to "Sustainable Cities and Communities." During the accelerated urbanization in the past two centuries, the vertical form of buildings has been rapidly changed, reshaping the 3-D structure of cities [1], [2], [3], [4], [5]. These changes are closely linked to issues such as urban energy usage [6], [7], human activity intensity [8], [9], and climate change [10], [11]. As the vertical dimension of building structures, the acquisition of building height data plays a crucial role in assessing the environmental impact of human activities, facilitating urban planning and management, and realizing sustainable development goals [12], [13].

Recently, remote sensing and geographic information systems (GIS) have been widely used in building height estimation studies. Three distinct types of data are used individually for high-resolution building height estimation, including:

1) high-resolution optical images;
2) airborne light detection and ranging (LiDAR);
3) interferometric synthetic aperture radar (InSAR).

High-resolution optical imageries provide spatial information of land surface. Through geometric analysis of the interactions among the sun, buildings, and shadows in high-resolution imagery, we can obtain information related to building heights [13], [14], [15], [16]. These high-resolution images provide detailed spatial information, enabling estimation of building heights at fine scales (i.e., $<$10 m). However, high-resolution optical images are limited in estimating building heights in areas with dense buildings [e.g., central business district (CBD)] where shadows cover each other [17]. Alternatively, LiDAR shows the potential to remit this issue. LiDAR can capture 3-D information of objects through point clouds. By analyzing the points related to building structures, it is possible to obtain building heights at individual building scale since the point cloud directly provides the coordinates of building roofs [18], [19], [20]. Although this type of data is efficient, it is costly, time-consuming, and labor-intensive in processing. Thus, the coverage of LiDAR is still limited, remaining it difficult to estimate the building height on a big scale. By contrast, InSAR can obtain 3-D information in urban and rural areas [21], [22], [23]. InSAR measures changes of the elevation on the ground surface by comparing the phase of two or more synthetic aperture radar images collected at different times. [24], [25], [26]. Li et al. [27] proposed an indicator of backscatter intensity and estimated building height in seven cities using the Sentinel-1 ground range detected data. The results showed that the InSAR-recorded backscatter values are strongly related to building height. However, this InSAR index performs well in estimating building heights above 500 m, with a decrease in accuracy at the scale of individual buildings [28]. In addition, the radiation of InSAR includes all objects

Manuscript received 4 January 2024; revised 24 February 2024; accepted 20 March 2024. Date of publication 12 April 2024; date of current version 19 April 2024. This work was supported in part by the National Science Fund for Distinguished Young Scholars under Grant 42225107, in part by the National Natural Science Foundation of China under Grant 42171409, in part by the National Natural Science Foundation of China under Grant 42171410, and in part by the National Key Research & Development Program of China under Grant 2019YFA0607203. *(Corresponding author: Xiaoping Liu.)*

Yangzi Che, Qian Shi, and Xiaoping Liu are with the Guangdong Province Key Laboratory for Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: cheyz@mail2.sysu.edu.cn; shixi5@mail.sysu.edu.cn; liuxp3@mail.sysu.edu.cn).

Xuecao Li is with the College of Land Science and Technology, China Agricultural University, Beijing 100083, China (e-mail: xuecaoli@cau.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3386124

indiscriminately, resulting in irrelevant information (e.g., trees) contained in the received data, which brings about a negative impact on building height estimation. In addition, some studies combined these datasets together and got better estimated heights using machine learning algorithm, indicating that the synthetic use of multisources datasets can be taken into account to avoid the uncertainty [29], [30], [31], [32]. For example, Frantz et al. [33] synthesized Sentinel-1 and multispectral Sentinel-2 time series and estimated the building height using random forest algorithm, increasing the accuracy compared to single-dataset model. Additionally, this method is suitable for estimating building heights at a finer scale, providing insights into building-scale height estimation.

However, most of the derived building height datasets are currently limited in terms of spatial resolution, especially lacking methods for estimating heights of individual buildings at a regional or national scale. The United States Geological Survey (USGS) released a national-scale building height map that was collected using radar [34]. However, the dataset was at a block scale with category information (ranging from "low" to "very high"), making it challenging to reveal the upward structure of buildings in fine-grained urban studies. Li et al. [27] produced a 1 km building height map in Europe, the United States (U.S.), and China using SAR (i.e., Sentinal-1) and other optical remote sensing images (i.e., Landsat and MODIS). However, the 1 km building height dataset is too coarse for urban system studies. Similarly, Li et al. [28] estimated the building height in major U.S. cities at a 500 m scale with the help of Sentinel-1 Ground Range Detected data. However, these 500-m grid-scale maps are still too coarse for urban studies due to the complex landscape within the city. In addition to these global building height results, some products have been developed to characterize building heights at a more fine-grained resolution, although these studies were mostly performed at small-regional scale (e.g., city scale) [13], [35], [36], [37], [38]. Thus, it is challenging to acquire data on a national or global scale, owing to the considerable time required for both pre-processing and model training. In addition, some studies have used machine learning techniques to estimate building heights at high resolution [33], [39], which rely heavily on the detailed vector shape of buildings.

To fill the research gap mentioned above, we propose a method for mapping building-scale height based on multisource datasets and evaluated the accuracy of our model. This study focusses on the following objects.

1) Propose a simple and reliable building-scale height estimation method based on machine learning and estimate the building height of 11 cities in the U.S.
2) Integrate multisource features (SAR, optical, socio-economical, terrain, and vector-based) and incorporate statistical values of remote-sensing images to enhance accuracy of building height estimation model.
3) Demonstrate the competitive applicability and effectiveness of the model to cover large region by validation in various cities, across different height intervals, within distinct regions, and intercomparison with other existing datasets.

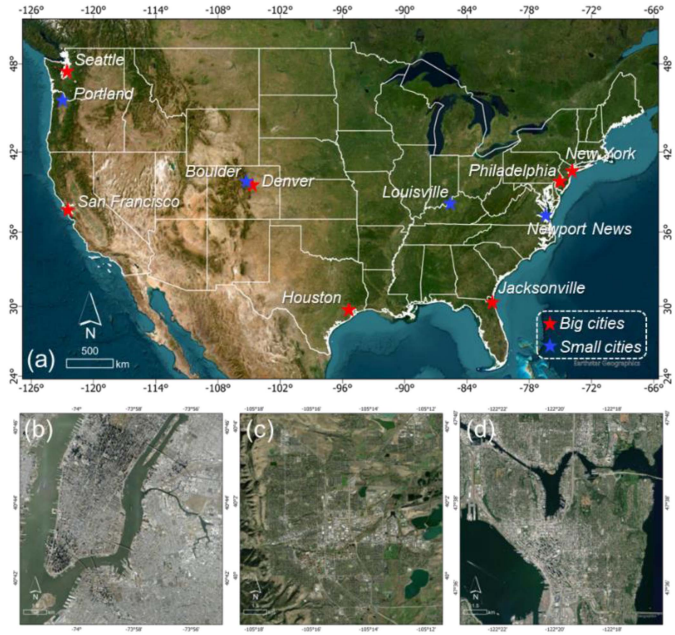

Fig. 1. Locations of 11 U.S. cities (a) and the enlarged view of (b) Seattle, (c) Boulder, and (d) New York.

II. Materials

A. Study Area

In total, 11 cities in the U.S. were selected to apply our model, including seven big and four small cities. These cities include some of the most densely populated megacities, such as New York, Los Angeles, and Houston, as well as cities with small population size, such as Portland and Newport News. Considering the heterogeneity of buildings, we chose cities widely located in all regions of the U.S. (see Fig. 1), enabling us to validate the feasibility of the height estimation model for the conterminous U.S. With a diverse morphology (e.g., building height, area, and the spatial patterns) of big cities (e.g., New York and Seattle) and small cities, we can comprehensively validate the model's robustness and effectiveness under complex urban landscapes.

B. Datasets

1) Multisource Datasets for Model Training: We utilized a diverse range of data sources to training building height estimation model in our study, including SAR, optical, terrain, social-economical, and vector data.

We extract synthetic aperture radar (SAR) features from Sentinel-1 and phased array type L-band synthetic aperture radar (PALSAR) images. Sentinel-1 is equipped with a C-band SAR, designed to provide high-quality radar imaging data for monitoring Earth's surface changes. This satellite provides two types of images [i.e., ground range detected (GRD) and single look complex (SLC)] at high resolution (i.e., 10 m) [40]. We used the backscatter coefficient with dual polarization (V.V. and V.H.) from GRD images to capture the structural and electromagnetic properties of urban elements [41]. We also used PALSAR as the auxiliary SAR features. PARSAR employs radar signals in the

TABLE I
FEATURES WITH MULTIPLE SOURCES USED IN OUR STUDY

Category	Datasets	Feature name	Resolution	Provider	Link
Optical data	Sentinel-2 (band2, band3, band4, band8)	B2, B3, B4, B8	10m	European Union/ESA/Copernicus	https://earth.esa.int/
	Nighttime light	light rad	463.83m	Earth Observation Group, Payne Institute for Public Policy, Colorado School of Mines	https://payneinstitute.mines.edu/
Synthetic aperture radar data	Sentinel-1 (VV, VH)	VV, VH	10m	European Union/ESA/Copernicus	https://earth.esa.int/
	PALSAR (HH, HV)	HH, HV	25m	JAXA EORC	https://www.eorc.jaxa.jp/ALOS/
Terrain data	DEM	dem, dsm, ndsm*	30m	NASA/USGS/JPL-Caltech	https://cmr.earthdata.nasa.gov/
	DSM		30m	JAXA Earth Observation Research Center	https://www.eorc.jaxa.jp/ALOS/
Population data	WorldPop	population	92.77m	WorldPop	https://www.worldpop.org/
Vector-derived data	Area of single building	Area	/	Microsoft	https://wiki.openstreetmap.org/wiki/Microsoft_Building_Footprint_Data
	Perimeter of single building	Length	/		

*ndsm is calculated by DSM-DEM

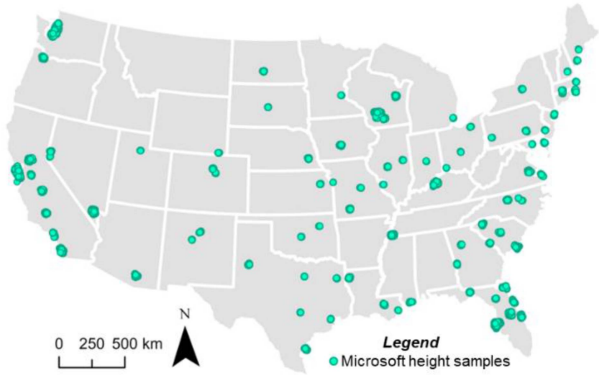

Fig. 2. Microsoft building height samples in the conterminous U.S.

TABLE II
BUILDING HEIGHT DATA FORM GIS GOVERNMENT PORTAL

City	Datasets	Link
Boulder	Buildings in Boulder, Wyoming	https://hub.arcgis.com
Pasadena	Buildings in Pasadena, California	
Portland	Buildings in Portland, Oregon	
Jefferson	Buildings in Jefferson County, Jefferson	
Newport News	Buildings in Newport News, Virginia	
Sauk	Buildings in Sauk, Wisconsin	

L-band frequency range to capture high-resolution images (i.e., 25 m) of the Earth's surface.

Sentinel-2 provides the spectral features as the inputs of the model. Sentinel-2 is a pivotal Earth observation program under the ESA's Copernicus initiative. It equipped with the multi spectral instrument (MSI), which can capture data across 13 spectral bands, including visible light (e.g., blue, green, and red), near-infrared, red-edge, and shortwave infrared. We selected Bands 2, 3, 4, and 8 at 10-m resolution. These bands provide valuable information related to impervious surface characteristics, as they capture data in the visible and near-infrared spectra [42].

We utilized the NASA topography mission elevation data and the advanced land observing satellite digital surface model (DSM) to enrich our comprehension of terrain elevation and surface conditions.

In addition to these primary satellite datasets, we incorporated various auxiliary datasets into our modeling approach, including WorldPop Global Project Population Data and Visible Infrared Imaging Radiometer Suite day/night datasets. These social-economical datasets are associated with building morphology.

We also calculated two key geometric indicators to capture geometric features that have proven effective in building height estimation [39]. The area and premiers of buildings were computed as predictor variables.

The utilization of these diverse datasets allowed us to construct a comprehensive model to estimate building heights. Further details on the features sourced can be found in Table I.

2) Reference Height Data and Building Footprints: We collected the building footprint datasets of 11 cities from Microsoft Building Footprints. The US Microsoft Building Footprint dataset is an open dataset provided by Microsoft in 2018, aiming to offer geographic information about building outlines and spatial locations across the United States [43]. The deep neural network model and polygonization methods were employed for extracting building boundaries. The footprints achieved an accuracy of 98.5% precision and 92.4% recall, providing precise building boundaries for height estimation.

The reference building height were collected from Microsoft and GIS government portals of cities. While releasing building footprints, Microsoft also provided partial building height data in 44 states (see Fig. 2). However, these height datasets primarily focus on urban centers with high-rise buildings and do not encompass entire cities and rural areas. For example, 2% of buildings in New York State were released with height information. To balance the sample representation of smaller cities, we also collected height dataset from through GIS government portals in some cities (see Table II).

III. METHODOLOGY

This study proposed a multisource feature fusion method to estimated building height of 11 US cities at fine-scale (i.e., building scale) and assessed the accuracy and applicability of the model through various validation methods (see Fig. 3). First, we extracted the statistical features of multi-source remote sensing datasets and calculated the geometry features of buildings. Second, we trained the model using random forest regression model. We evaluated the synthetic use of multisource data and the effectiveness of statistical values in building height model. Also, we evaluated the model accuracy in in various cities, across different height intervals, and within distinct regions. Third, we mapped the building-scale height in 11 cities and compared the accuracy of our estimated building height with other existing height products.

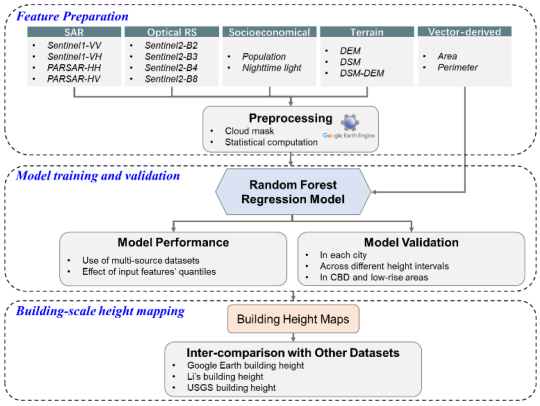

Fig. 3. Overall framework.

A. Feature Preparation and Preprocessing

We used SAR, optical, socio-economical, terrain, and vector-based features for model training. We filtered the high-quality images and calculated the statistical values of the images before training the model.

We constructed a multisource dataset to reflect building properties. Firstly, we utilized the VV and VH polarization bands of SAR (i.e., Sentinel-1), as the backscatter coefficient reflects the reflection and scattering information of buildings [44]. Second, we integrated visible light bands (blue, green, red) and near-infrared bands of Sentinel-2 as input features. Visible light provides information on building boundaries and textures, while the near-infrared band can reflect the thermal radiation capability of materials, which is related to building height [17]. In addition, we utilized terrain features to reflect surface undulations. Especially, difference of DSM and DEM can reflect the heights of surface objects [45]. Furthermore, we included building area and perimeter, as morphological features of buildings are related to height according to fractal theory [46]. In addition, we inducted population and nighttime light data as auxiliary data, as these datasets reflect land use intensity in relation to buildings [47]. Finally, 93 features were used in our height estimation model.

To ensure the quality of input features, we filtered out images in which the percentage of clouds was above 20% and calculated the mean value of all the images in 2020. All images were resampled to 10-m resolution for feature extraction.

To fully use the information of images, we calculated the statistic values (i.e., mean values, mean value, standard deviation, and quantiles (i.e., 5%, 25%, 50%, 75%, and 95%) of the raster pixels intersecting with building polygon vectors as the remote-sensing features. Subsequently, we integrated the statistical values and vector features of each building as the independent features to characterize different dimensions of building attributes, providing a more comprehensive information related to building heights.

B. Model Training and Validation

We used the random forest regression to develop the building height estimation model. We used stratified sampling to strategy to extract training and testing samples, covering all U.S. states. Within each state, we conducted stratified sampling at predefined intervals. The number of training samples at each interval was adjusted to correspond the height distribution in studies by Esch et al. [48] and Li et al. [27], ensuring that the sample height distribution closely resembled the nationwide distribution.

We conducted a sensitivity analysis of input sample sizes for the model. We employed a ten-fold cross-validation strategy and selected the most suitable model for height estimation in this study, considering both the model's efficiency and accuracy. We used ordinary least squares regression to evaluate model performance and quantified the height uncertainty using the coefficient of determination (R^2) (1) and the root mean square error (RMSE) (2). We also plotted the distribution of height values to assess the height precision in the comparison across different height intervals and within different regions (i.e., in CBDs and low-rise areas)

$$R^2 = 1 - \frac{(n-1)\sum_{i=1}^{n}(H_{e,i} - H_{r,i})^2}{(n-2)\sum_{i=1}^{n}(H_{e,i} - H_{r,i})^2} \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(H_{e,i} - H_{r,i})^2}{n}}. \tag{2}$$

C. Building-Scale Height Mapping

We mapped the building height of 11 U.S. cities based on the height estimation model. We compared our height datasets with those of other existing datasets. The building-scale height map was aggregated to different scales (i.e., 1 km, and block scales) (3) for comparison with Google Earth Pro building height, building height in Li et al. [27], and building height provided by USGS, respectively

$$\text{BH} = \frac{\sum_{i}^{n} H_i \times A_i}{\sum_{i}^{n} A_i} \tag{3}$$

where H_i and A_i represent the height and area of individual buildings, respectively, while $\sum_{i}^{n} A_i$ denotes the total area of all buildings within the statistical area (i.e., 1 km grid and block).

IV. Results

A. Synthetic Use of Multisource Data

A synergistic model with multisource data input, including radar, optical, terrain, socioeconomic, and vector features, is better than a single feature-based model for predicting building height. To quantitatively understand the role of different categories of remote sensing data in height model, we compared the accuracy of seven models trained with different categories of features. The results show that radar-only and optical-only models can estimate building height at R^2 of 0.56 and 0.63, respectively [see Fig. 4(a) and (b)], indicating that radar and optical features can reflect considerable vertical information [33], [48]. Terrain-only model can get good results, with an R^2 of 0.62, similar to the results estimated by the model using optical features only [see Fig. 4(c)]. The nDSM feature (calculated by DSM-DEM) characterizes the height of objects on the Earth's surface by calculating the differences between the

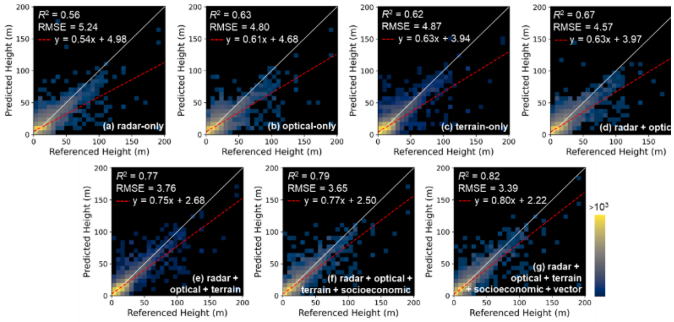

Fig. 4. Regression results of different input feature: (a) radar-only; (b) optical-only; (c) terrain-only; (d) radar and optical; (e) radar, optical and terrain; (f) radar, optical, terrain, and socioeconomic; (g) radar, optical, terrain, socioeconomic, and vectors. The red lines represent the ordinary least squares regression line in each city. The white lines are the one-to-one lines.

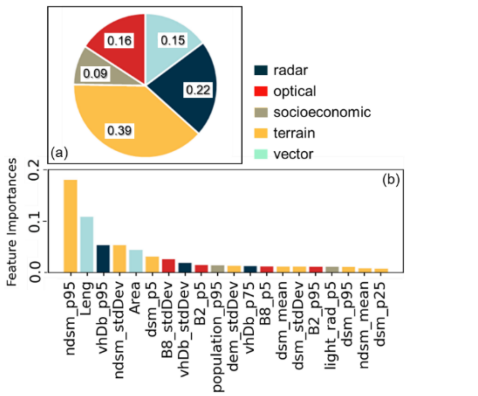

Fig. 5. Feature importance of the predictors used in the regression model. (a) Contributions from each category. (b) Ranking of variable importance.

DSM and DEM, providing significant building vertical information. However, it is worth noting that synthetic models can obtain more accurate estimated height results. The synergistic model using radar and optical datasets is more accurate than single-dataset models using radar-only or optical-only features [see Fig. 4(d)]. Furthermore, the model synthetically using radar, optical, and terrain datasets can estimate height with R^2 of 0.77 [see Fig. 4(e)]. Socioeconomic datasets, including population and nighttime-light features, slightly improved the accuracy of the model [see Fig. 4(f)]. In addition, geometric features can effectively increase the accuracy of the height estimation model [see Fig. 4(g)] by providing information related to building forms.

By further calculating the model's feature contributions, we quantified the impacts of different categories of features (see Fig. 5). We further calculated the Gini index of features to quantified contributions of different categories of datasets. The Gini index of each feature represents its ability to ensure the purity of the divided datasets in classification and regression trees, which is the base learner in the random forest model [49]. Terrain and radar features contributed significantly (61%) to the model's accuracy according to the calculated Gini feature importance. These two features provide vertical information about surface objects. In the terrain category, the difference

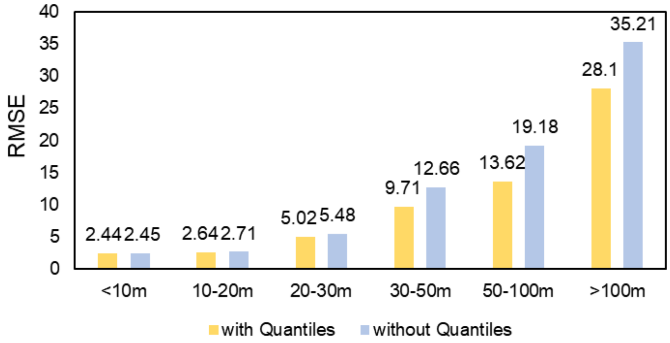

Fig. 6. RMSEs of models with and without quantiles as input features.

between the DSM and DEM distinctly depicted the height of objects relative to the land surface, contributing the most to the model, followed by the SAR images. SAR images provide interferometric stripes containing the topographic information of target regions, reflecting the height of buildings [50]. The high resolution of the Sentinel-1 datasets (i.e., 10 m) was also effective for building scale estimations. Vector-derived features are also critical variables for estimating building height. Notably, the two vector-derived features (i.e., area and perimeter) are both in the top five in the feature importance ranking. The geometric attributes of buildings tend to maintain specific ratios to ensure functionality (e.g., to acquire natural light-penetrating buildings) [51], indicating the relationship between other geometric features and building heights. In addition to these categories, optical and socioeconomic features provide information for estimating the building height. Optical images contain information regarding ground objects, whereas socioeconomic features reflect the built-up environment and anthropogenic activities that are crucial for building height estimation.

B. Effectiveness of Statistical Values in Building Height Models

The difference of RMSE between models with and without quantiles indicated that using quantiles as input features improved the accuracy of model, especially for high-rise buildings (see Fig. 6). We respectively calculated the RMSEs for models using solely the mean values and other statistical values (i.e., standard deviation, 5%, 25%, 50%, 75%, and 95% quantiles) of the raster within building boundaries as the input features. These calculations of RMSEs were performed on the test dataset within various height intervals (i.e., $<$10 m, 10–20 m, 20–30 m, 30–50 m, 50–100 m, $>$100 m). The results indicate that quantiles exhibit a relatively minor influence on accuracy enhancement when estimating low-rise buildings. For instance, when estimating heights less than 10 m, the inclusion of quantiles only results in an improvement of RMSE by approximately 0.01 m. However, as estimating higher buildings, the impact of quantiles on accuracy becomes increasingly significant. The utilization of quantiles can lead to a reduction in RMSE by approximately 3 m (comparable to the height of a single-story building) when estimating buildings with height range of 30–50 m. Furthermore, these quantiles enhanced accuracy by approximately 6

TABLE III
BUILDING HEIGHT VALIDATION ON 10% LEFT SAMPLES

City	R^2	RMSE(m)
New York	0.86	4.13
Philadelphia	0.85	5.97
San Francisco	0.83	3.65
Houston	0.62	4.40
Seattle	0.79	3.31
Denver	0.77	2.85
Jacksonville	0.51	3.11
Portland	0.62	2.57
Louisville	0.61	3.11
Boulder	0.61	3.16
Newport News	0.56	4.37
All 11 cities	**0.88**	**4.32**

and 7 m in estimating buildings with height of 50–100 m and >100 m, respectively. The comparative results demonstrate the utility of the quantiles as input features in height estimation model.

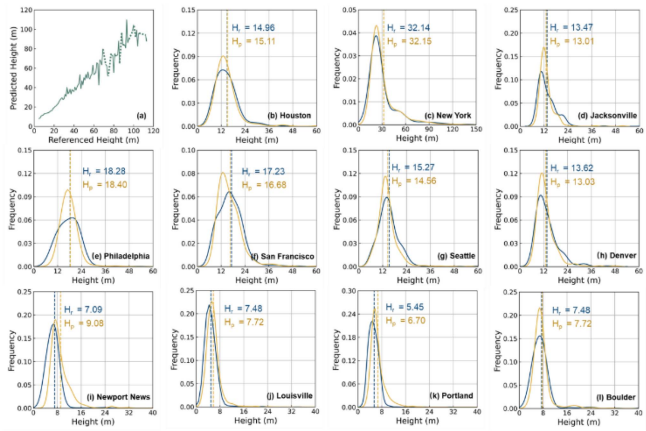

Fig. 7. Height frequency of estimated building heights and reference building heights. (a) Average estimated heights in reference height intervals (in 1-m increments). Dash lines are drawn where there is no reference value. (b–l) Height distributions in cities. The blue lines represent the frequency of reference heights. The yellow lines represent the estimated heights. H_r is the mean height of reference buildings. H_p is the mean value of the predicted height.

C. Validation of the Estimated Building Height

To evaluate the model's accuracy and its robustness across different urban scenarios and spatial contexts, we evaluated the model accuracy in various cities, across different height intervals, and within distinct regions (i.e., CBDs and low-rise building dominated areas). Details of the model spatial transferability can be found in Section V-B.

1) Performance in Cities: Overall, the model can reasonably estimate building heights in both big and small cities, demonstrating the model application in cities with different sizes. The R^2 between estimated and referred building heights is 0.88, and the RMSE is 4.32 m (i.e., around the height of one story of 3 m) (see Table III) in total 11 cities. In big cities, the estimated heights agree with the reference heights. The R^2 of big cities is close to 1 (e.g., 0.83 in San Francisco, 0.85 in New York, and 0.79 in Seattle), with RMSEs of approximately 3 m. However, although R^2 in New York reached 0.85, the RMSE was relatively large (9.97 m), which was caused by the underestimation of skyscrapers in the Manhattan region (i.e., the central area of New York City). In small cities, the reference heights were consistent with the predicted heights. The R^2 in small cities reached around 0.6, and the height uncertainty (i.e., RMSE) ranged from 2.57 to 4.37 m, which is usually lower than that for one story.

2) Performance on Different Height Intervals: The frequency distributions of the estimated and reference heights were in good agreement in all height intervals in all 11 cities, suggesting good model performance in estimating both high-rise and low-rise buildings (see Fig. 7). The line of the mean estimated heights in all height classes (with an interval of 1 m) of the reference heights was close to the one-to-one line [see Fig. 7(a)]. However, for high-rise buildings (above 30 m), the estimated heights fluctuated violently, indicating an increase in the uncertainty of predicting the height of high-rise buildings. The line has a small bump at height intervals of less than 10 m, indicating a slight overestimation of low-rise buildings.

The distribution of estimated heights and reference heights were well matched in each big and small cities [see Fig. 7(b)–(l)], suggesting the applicability in both high-rise and low-rise dominated cities. As for the big cities with more high-rise buildings, their mean building height ranged from 13 to 15 m. The mean difference between the two datasets for these cities was approximately 0.6 m. In particular, the building in New York City was notably higher than other cities, with a mean height of 32.14 m. The difference between the estimated and reference height in New York was 0.01 m, demonstrating that our model is applicable to such cities dominated by high-rise buildings. Besides, in small cities with the mean height about 5–8 m, the mean difference between estimated and the reference heights was about 1 m (e.g., the difference was only 0.24 m in Louisville.). The results showed that the model also achieves relatively accurate results in areas dominated by low-rise buildings.

3) Model Performance in CBD and Low-Rise Areas: Our dataset showed the consistency with the reference height in both CBDs and low-rise areas, confirming that our model can provided heights in dense high-rise building scenarios and sparse low-rise building scenarios. We plotted the distribution of buildings in these two regions separately, and compared them with reference height and results in Li et al. [27]. We distinguished the buildings in the 5 km × 5 km extent centered on the CBDs [28], and the remaining were regarded as located in the low-rise building dominated areas. The number of buildings at all height intervals (i.e., 5 m) of our estimated height datasets was close to that of the reference height datasets in the CBDs [see Fig. 8(a)]. The mean difference between our estimated heights and the reference heights was approximately 0.3 m in the CBDs, which outperforms the results of Li et al. [27] (approximately 8 m). For buildings in low-rise building-dominated areas, our estimated height performed better than the results of Li et al. [27] [see Fig. 8(b)]. The mean difference between our height datasets and the reference height was approximately 2 m in low-rise

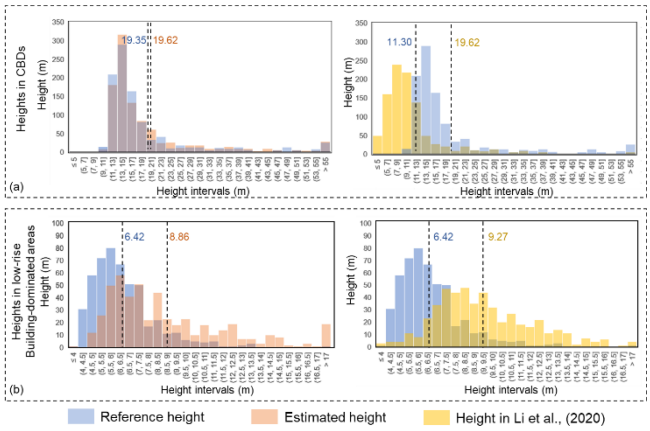

Fig. 8. Histograms of high-rise and low-rise buildings. (a) Height distribution in CBDs. (b) Height distribution in low-rise building-dominated areas.

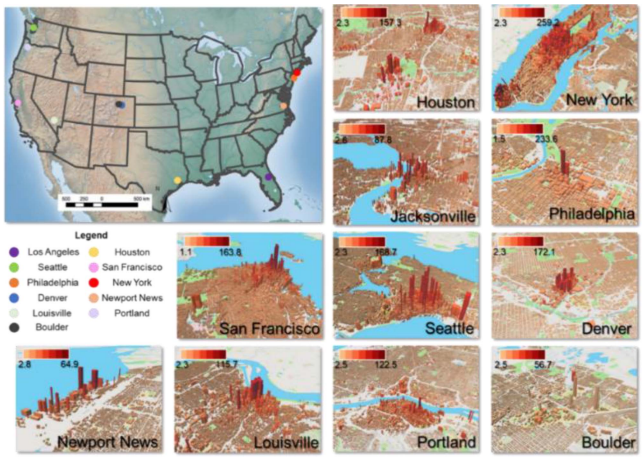

Fig. 9. Map of building height in example cities.

building-dominated areas, which is closer to the reference height datasets.

D. Three-Dimensional Morphology Within the Cities

The underlying urban surface showed a distinct roughness owing to variations in building heights according to our building-scale height results (see Fig. 9). The height of buildings in large cities and small cities show different height patterns. The average height of buildings in big cities was higher than that of buildings in small cities. For instance, in big cities, the average building height usually exceeds 10 m, whereas, in small cities, the average building height is generally less than 10 m. In addition, the height of the buildings decreased from the city center to the surrounding fringe areas within the cities. Generally, tall buildings are located in the core areas of cities (e.g., CBDs), whereas low-rise buildings are dominant on the fringes of cities.

V. Discussion

A. Cross-Validation With Other Existing Height Datasets

1) Validation at the Building Scale: The results for both big and small cities showed relatively high agreement regarding

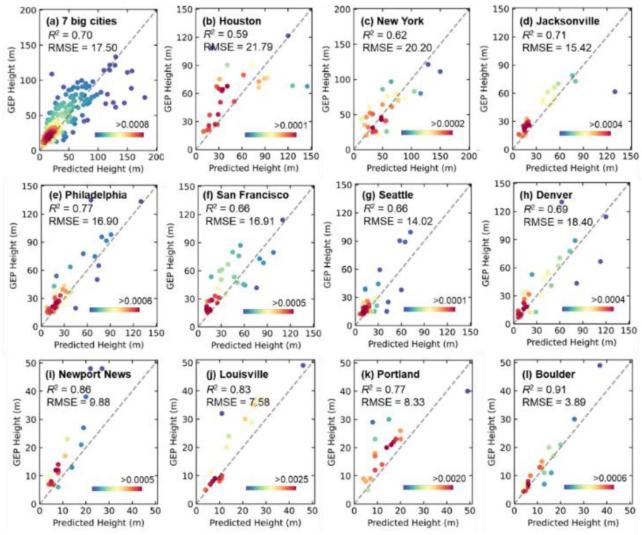

Fig. 10. Scatter plots between measured height on Google Earth and estimated heights in (a) seven big cities and individual (b–h) big and (i–l) small cities.

accuracy compared to the manually measured heights from Google Earth Pro. We measured the height of 50 buildings in each city with the help of 3D ruler tool in Google Earth Pro. The total R^2 was 0.70, and the total RMSE was 17.50 m, with the test samples randomly selected from the entire city [see Fig. 10(a)]. The model characterized the building heights in big cities with an R^2 of approximately 0.70 [see Fig. 10(b)–(h)]. The heights of buildings in big cities ranged from low to high, increasing the uncertainty of the estimated heights. We found that buildings below 50 m in height could be accurately estimated. The results may be underestimated or overestimated when building heights exceed 50 m. The model also accurately estimated building heights in small cities [see Fig. 10(i)–(l)]. The R^2 of small cities was approximately 0.85, and the RMSEs of small cities were approximately 6 m. In small cities, building heights were mostly concentrated at approximately 10 m (e.g., Newport News, Louisville, and Boulder), above which a slight overestimation was observed. Notably, buildings that could be measured on Google Earth were located mainly in the city center, where high-rise buildings with larger height values are dominant; therefore, the RMSEs were relatively larger, especially in big cities.

2) Validation at 1-km Resolution: The spatial patterns of our estimated and reference heights demonstrate the spatial consistency of the two datasets at a resolution of 1 km (see Fig. 11). The reference building height was collected and aggregated from Microsoft buildings with height in their attributes. The height result in Li et al. [27] was estimated at 1 km-scale using Earth observation datasets with random forest algorithm. Compared with the heights in Li et al. [27], our estimated results are similar to the reference height in terms of the spatial patterns of building heights in most representative cities. Our results can accurately depict high-rise buildings within cities [e.g., Manhattan in New York, see Fig. 11(a) and the CBD in Houston, see Fig. 11(b)], where high-rise building-dominated areas are primarily underestimated. For low-rise buildings, slight overestimations exist

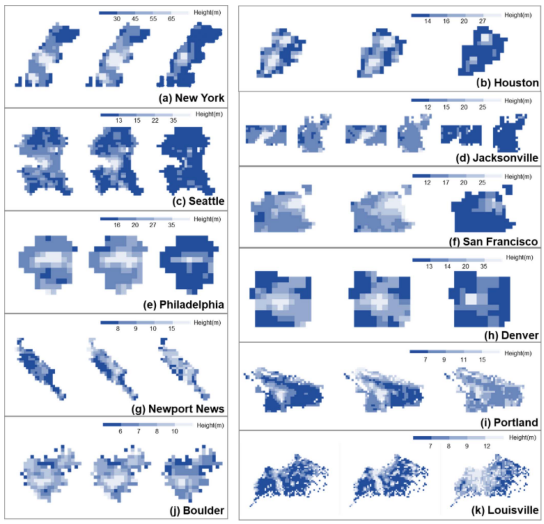

Fig. 11. Spatial patterns based on our results, reference heights, and heights in Li et al. [27] on a 1-km scale in 11 cities.

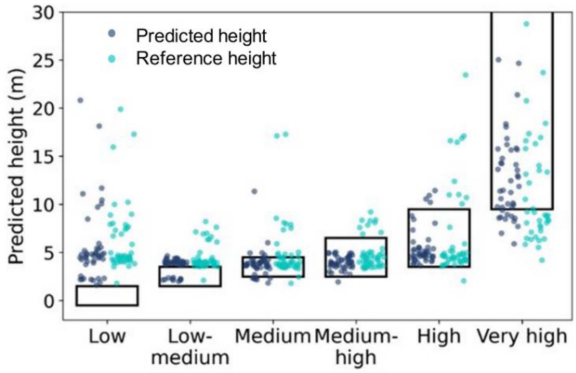

Fig. 12. Box plot of estimated and USGS heights.

in both our results and the height in Li et al. [27], although it is worth noting that the overestimation is more noticeable in Li et al. [27]. In addition, our results accurately depict the height from the CBDs to the surrounding areas in big and small cities, showing noticeable spatial heterogeneity across cities. For example, in Seattle, high-rise and low-rise building regions were similar to those in the reference heights [see Fig. 11(c)].

3) Validation at Block Scale: Our estimated results are consistent with the USGS product [34] for different height categories, suggesting the high accuracy of our predicted heights. The USGS building height dataset was acquired from the National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission. It provides a qualitative description of building height within each block in the US, including categories of "Low (primarily 1—2 story buildings)," "Low-medium (primarily 2–3 floor buildings)," "Medium (primarily 3–4 story buildings)," "Medium-high (primarily 3–5 story buildings)," "High (primarily 4–9 story buildings)," "Very high (buildings average 10 stories or higher)." We plotted the USGS categories for all blocks within 11 cities, assuming that the height of each story is 3 m [52], and compared USGS height to estimated heights.

Most of our estimated height results were in the corresponding height ranges identified by the USGS categories, especially in medium, medium-high, and high (see Fig. 12). However, some uncertainties existed in the low and very high categories. Our estimated heights are higher than those of the USGS products in the Low category. In addition, underestimations existed in the very high category. Notably, uncertainty was also observed in the low and very high categories in the comparison between the reference and USGS heights.

B. Spatial Transferability for National-Scale Mapping

Spatial transferability reflects the ability of models that train on a certain region and applied to a new region. The method proposed in this research has the potential for estimating building heights in different regions in large scale.

We ensure model transferability by combining synthetic features and a stratified sampling approach. First, we enhance spatial transferability by incorporating diverse feature categories, such as nDSM (DSM-DEM), building area, and building perimeter, directly reflecting building structure. These features significantly contribute to building height estimation (see Fig. 5). Second, we employ stratified sampling within each state to cover a wide range of height distributions, including samples from urban centers to rural areas. Our strategies efficiently capture spatial patterns and sample diversity, enhancing accuracy in supervised learning. Our approach offers a robust way to acquire training and testing data features, ensuring reliable accuracy.

In Section IV, we compared our estimated heights within different cities, height intervals, and urban scenarios (CBDs and low-rise areas). Our method's transferability is evident according to the results. First, estimated heights across cities of various sizes and patterns showed the consistency with reference height (i.e., R^2: 0.51–0.86, RMSE: 2.57–5.97 m). For a megacity like New York with approximately 800 skyscrapers, our method can also achieve an R^2 of 0.86. Second, estimated heights within height intervals closely matched reference data, showcasing our model's ability for high and low buildings across the U.S. Third, our model provided reliable height estimates across diverse cityscapes. In CBDs with high building density, our method achieves consistent results with reference data while traditional shadow-based methods often yield uncertain height estimates. It also performed well in sparser, low-rise buildings (e.g., buildings in city edges or rural areas).

To assess the applicability of our method for height estimation in urban areas with different building density levels, we calculated the model accuracy in different building density zones within the cities. We chose New York, Houston, and San Francisco for this experiment because buildings in these developed cities are more concentrated where shadows and layovers are more likely to introduce biases in estimation results. First, we converted building polygons into points and calculated the density for each building. Second, we categorized building densities into four levels according to quantiles; for example, building density in the 0.75–1.00 quantile range represents buildings in the areas with the highest building density. We calculated the

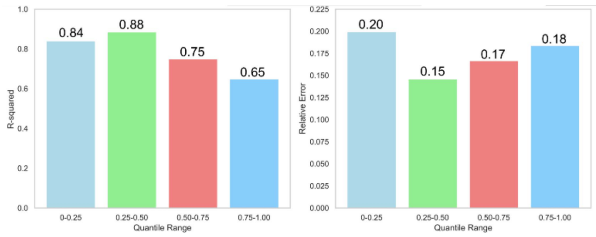

Fig. 13. Accuracy in different density levels within cities.

TABLE IV
BUILDING HEIGHT VALIDATION ON 10% LEFT SAMPLES IN OTHER FIVE CITIES

City	States	R^2	RMSE(m)
Baltimore	MD	0.78	3.71
Tempa	FL	0.62	4.40
Columbus	GA	0.58	2.89
New Orleans	LA	0.78	5.30
Syracuse	LA	0.79	3.31

R^2 and relative error (i.e., the proportionate difference between predicted value and the corresponding actual value) for buildings within each density quantile range (see Fig. 13). The results showed that in high-density areas, the R^2 is relatively low (0.65); however, the accuracy of the relative error is comparatively good (0.18). This indicates that although the accuracy in the high-density area is lower than that in other areas, because the influence of shade and layover, the overall estimation is acceptable. This may because that our multisource feature fusion method can complement and correct errors present in different data sources. For example, SAR data can compensate for the missing spectral information in optical data shadows and correct the errors caused by shadows [53]. At the same time, the night light data and population data we choose are not affected by the shadow, and can also mitigate the impact of the shadow and layover [54].

To further evaluate the transferability and the applicability to use for continental scale building height mapping, we tested the accuracy in five other cities in U.S. (see Table IV). Baltimore, Tempa, and New Orleans are relatively larger, with population of about 590 000, 399 700, and 383 000, respectively. And Columbus and Syracuse are relatively small, with population of about 196 500 and 142 000. The estimated heights are consistent with the reference height in these cities. The R^2 of these cities ranges from 0.58 to 0.80, with a highest R^2 reached 0.79 in Syracuse, Louisiana. And the RMSEs All cities have RMSE below 5.3 m, which is also relatively small.

VI. CONCLUSION

This study proposes a method for mapping building-scale heights over continental or regional areas. By incorporating multisource remote sensing features and calculating statistical values of features, our model can effectively estimate the heights of buildings. The model R^2 achieved 0.82 and an RMSE of 3.39 m when training with multi-source datasets (i.e., SAR, optical, terrain, social-economical, vector-based datasets). We also found the effectiveness of statistical values in building height estimating. especially for the higher buildings (i.e., the RMSE decreases by 3–8 m when estimating buildings that exceed 30 m). Our estimated heights are consistent with the reference height showed by the evaluation of model accuracy in in various cities, across different height intervals, and within distinct regions (i.e., CBDs and low-rise building dominated areas). The R^2 was approximately 0.80 and 0.60 in big and small cities, respectively.

We mapped the 3-D map of the 11 U.S. cities using our height estimating model and found a distinct roughness of the underlying surface. The mapping results also showed the superiority of our dataset comparing to datasets from Google Earth Pro, height in [27], and USGS height. Our estimated heights showed consistency (i.e., $R^2 > 0.70$ in big and small cities) with the heights manually measured from Google Earth Pro in the building-scale comparison. Regarding the 1 km scale validation, the estimated height better depicted the spatial pattern and was closer to the reference dataset than to the heights predicted by Li et al. [27]. And most of our estimated height results were within the corresponding height range identified by the USGS categories.

Our method can be used to acquire regional or global building height datasets, improving the fineness and the spatial extent of the three-dimensional urban morphology estimation. The results exhibited our method's satisfactory and accuracy of our method, indicating that the method based on multisource datasets has great potentials for building-scale height mapping at large scale. The method proposed in this study can build a global height estimation model. Especially, the method is useful for estimating building heights in nations with limited building-scale height datasets.

With the help of 3-D building information, we can further understand the complex interactions among urban form, climate, and human activities [55], [56], [57]. Using building height datasets, the upward and outward growth of cities can be depicted, providing a deeper understanding of the urbanization process. We can also analyze the relationship between urban morphology and local climate by calculating urban canopy parameters using height datasets [58], [59], [60], [61]. In addition, we can analyze anthropogenic activities and improve the accuracy of estimating socioeconomic factors (e.g., population, carbon emissions, and energy use) with the help of building height datasets [62], [63], [64], [65], [66].

REFERENCES

[1] J. Gao and B. C. O'Neill, "Mapping global urban land for the 21st century with data-driven simulations and shared socioeconomic pathways," *Nature Commun.*, vol. 11, no. 1, pp. 2302–2302, 2020, doi: 10.1038/s41467-020-15788-7.

[2] X. Liu et al., "High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015," *Nature Sustainability*, vol. 3, no. 7, pp. 564–570, 2020, doi: 10.1038/s41893-020-0521-x.

[3] Y. Chen, X. Li, X. Liu, and B. Ai, "Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation strategy," *Int. J. Geographical Inf. Sci.*, vol. 28, no. 2, pp. 234–255, 2014, doi: 10.1080/13658816.2013.831868.

[4] X. Liu et al., "A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects," *Landscape Urban Plan.*, vol. 168, pp. 94–116, 2017, doi: 10.1016/j.landurbplan.2017.09.019.

[5] G. Chen et al., "Global projections of future urban land expansion under shared socioeconomic pathways," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 537, doi: 10.1038/s41467-020-14386-x.

[6] C. Wang, M. Ferrando, F. Causone, X. Jin, X. Zhou, and X. Shi, "Data acquisition for urban building energy modeling: A review," *Building Environ.*, vol. 217, 2022, Art. no. 109056, doi: 10.1016/j.buildenv.2022.109056.

[7] X. Xu, J. Ou, P. Liu, X. Liu, and H. Zhang, "Investigating the impacts of three-dimensional spatial structures on CO2 emissions at the urban scale," *Sci. Total Environ.*, vol. 762, 2020, Art. no. 143096, doi: 10.1016/j.scitotenv.2020.143096.

[8] X. Zhao, Y. Zhou, W. Chen, X. Li, X. Li, and D. Li, "Mapping hourly population dynamics using remotely sensed and geospatial data: A case study in Beijing, China," *GIScience Remote Sens.*, vol. 58, no. 5, pp. 717–732, 2021, doi: 10.1080/15481603.2021.1935128.

[9] S. Shang, S. Du, S. Du, and S. Zhu, "Estimating building-scale population using multi-source spatial data," *Cities*, vol. 111, 2021, Art. no. 103002, doi: 10.1016/j.cities.2020.103002.

[10] L. Shao, W. Liao, P. Li, M. Luo, X. Xiong, and X. Liu, "Drivers of global surface urban heat islands: Surface property, climate background, and 2D/3D urban morphologies," *Building Environ.*, vol. 242, 2023, Art. no. 110581, doi: 10.1016/j.buildenv.2023.110581.

[11] Y. Sun, N. Zhang, S. Miao, F. Kong, Y. Zhang, and N. Li, "Urban morphological parameters of the main cities in China and their application in the WRF model," *J. Adv. Model. Earth Syst.*, vol. 13, no. 8, 2021, Art. no. e2020MS002382, doi: 10.1029/2020MS002382.

[12] Y. Zhou et al., "Satellite mapping of urban built-up heights reveals extreme infrastructure gaps and inequalities in the Global South," *Proc. Nat. Acad. Sci.*, vol. 119, no. 46, 2022, Art. no. e2214813119, doi: 10.1073/pnas.2214813119.

[13] Y. Cao and X. Huang, "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 Sens. Environ.*, vol. 264, 2021, Art. no. 112590, doi: 10.1016/j.rse.2021.112590.

[14] G. Liasis and S. Stavrou, "Satellite images analysis for shadow detection and building height estimation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 437–450, 2016, doi: 10.1016/j.isprsjprs.2016.07.006.

[15] X. Zhou and S. W. Myint, "Shadow pattern-enhanced building height extraction using very-high-resolution image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 180–190, 2023, doi: 10.1109/JSTARS.2022.3221146.

[16] Y. Zhou, T. Wei, X. Zhu, and M. Collin, "A parcel-based deep-learning classification to map local climate zones from Sentinel-2 images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4194–4204, 2021, doi: 10.1109/JSTARS.2021.3071577.

[17] G. Sun et al., "Combinational shadow index for building shadow extraction in urban areas from Sentinel-2A MSI imagery," *Int. J. Appl. Earth Observation Geoinf.*, vol. 78, pp. 53–65, 2019, doi: 10.1016/J.JAG.2019.01.012.

[18] Y. Huang, L. Zhuo, H. Tao, Q. Shi, and K. Liu, "A novel building type classification scheme based on integrated LiDAR and high-resolution images," *Remote Sens.*, vol. 9, no. 7, 2017, Art. no. 679, doi: 10.3390/rs9070679.

[19] J. Lao et al., "Retrieving building height in urban areas using ICESat-2 photon-counting LiDAR data," *Int. J. Appl. Earth Observation Geoinf.*, vol. 104, 2021, Art. no. 102596, doi: 10.1016/j.jag.2021.102596.

[20] S. K. Das, P. P. Joshi, R. S. Kokitkar, U. V. M. Krishna, H. A. Tanti, and A. C. Phadake, "Detection and validation of cloud top height from scanning Ka-band radar measurements using digital image processing technique," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1848–1856, 2021, doi: 10.1109/JSTARS.2020.3042868.

[21] M. Chini, R. Pelich, L. Pulvirenti, N. Pierdicca, R. Hostache, and P. Matgen, "Sentinel-1 InSAR coherence to detect floodwater in urban areas: Houston and hurricane harvey as a test case," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 107, doi: 10.3390/rs11020107.

[22] Y. Sun, Y. Hua, L. Mou, and X. X. Zhu, "CG-Net: Conditional GIS-aware network for individual building segmentation in VHR SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5201215, doi: 10.1109/TGRS.2020.3043089.

[23] X. X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.

[24] Y. Morishita, M. Lazecky, T. J. Wright, J. R. Weiss, J. R. Elliott, and A. Hooper, "LiCSBAS: An open-source InSAR time series analysis package integrated with the LiCSAR automated Sentinel-1 InSAR processor," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 424, doi: 10.3390/rs12030424.

[25] F. Sica, F. Calvanese, G. Scarpa, and P. Rizzoli, "A CNN-based coherence-driven approach for InSAR phase unwrapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4003705, doi: 10.1109/LGRS.2020.3029565.

[26] B. Zhu, Y. Wang, and H. Yu, "An algorithm measuring urban building heights by combining the PS-InSAR technique and two-stage programming approach framework," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7624–7635, 2023, doi: 10.1109/JSTARS.2023.3305890.

[27] M. Li, E. Koks, H. Taubenböck, and J. van Vliet, "Continental-scale mapping and analysis of 3D building structure," *Remote Sens. Environ.*, vol. 245, 2020, Art. no. 111859, doi: 10.1016/j.rse.2020.111859.

[28] X. Li, Y. Zhou, P. Gong, K. C. Seto, and N. Clinton, "Developing a method to estimate building height from Sentinel-1 data," *Remote Sens. Environ.*, vol. 240, 2020, Art. no. 111705, doi: 10.1016/j.rse.2020.111705.

[29] B. Tu, Q. Ren, J. Li, Z. Cao, Y. Chen, and A. Plaza, "NCGLF2: Network combining global and local features for fusion of multisource remote sensing data," *Inf. Fusion*, vol. 104, 2024, Art. no. 102192, doi: 10.1016/j.inffus.2023.102192.

[30] C. Yang and S. Zhao, "A building height dataset across China in 2017 estimated by the spatially-informed approach," *Sci. Data*, vol. 9, no. 1, 2022, Art. no. 76, doi: 10.1038/s41597-022-01192-x.

[31] X. Ma et al., "Mapping fine-scale building heights in urban agglomeration with spaceborne LiDAR," *Remote Sens. Environ.*, vol. 285, 2023, Art. no. 113392, doi: 10.1016/j.rse.2022.113392.

[32] B. Tu, Z. Wang, H. Ouyang, X. Yang, J. Li, and A. Plaza, "Hyperspectral anomaly detection using the spectral–spatial graph," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5542814, doi: 10.1109/TGRS.2022.3217329.

[33] D. Frantz et al., "National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series," *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112128, doi: 10.1016/j.rse.2020.112128.

[34] J. A. Falcone, "U.S. national categorical mapping of building heights by block group From Shuttle Radar Topography Mission Data: U.S. Geological Survey Data," 2016, doi: 10.5066/F7W09416.

[35] C.-J. Liu, V. A. Krylov, P. Kane, G. Kavanagh, and R. Dahyot, "IM2ELEVATION: Building height estimation from single-view aerial imagery," *Remote Sens.*, vol. 12, no. 17, 2020, Art. no. 2719, doi: 10.3390/rs12172719.

[36] C. Ren, M. Cai, X. Li, Y. Shi, and L. See, "Developing a rapid method for 3-dimensional urban morphology extraction using open-source data," *Sustain. Cities Soc.*, vol. 53, 2020, Art. no. 101962, doi: 10.1016/j.scs.2019.101962.

[37] D. Yu, S. Ji, J. Liu, and S. Wei, "Automatic 3D building reconstruction from multi-view aerial images with deep learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 171, pp. 155–170, 2021, doi: 10.1016/j.isprsjprs.2020.11.011.

[38] B. Tu, W. He, Q. Li, Y. Peng, and A. Plaza, "A new context-aware framework for defending against adversarial attacks in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5505114, doi: 10.1109/TGRS.2023.3250450.

[39] J. Arehart, F. Pomponi, B. D'Amico, and W. Srubar Iii, "A new estimate of building floor space in North America," *Environ. Sci. Technol.*, vol. 55, pp. 5161–5170, 2021, doi: 10.1021/acs.est.0c05081.

[40] A. F. Sandhini Putri, W. Widyatmanti, and D. A. Umarhadi, "Sentinel-1 and Sentinel-2 data fusion to distinguish building damage level of the 2018 Lombok Earthquake," *Remote Sens. Appl., Soc. Environ.*, vol. 26, 2022, Art. no. 100724, doi: 10.1016/j.rsase.2022.100724.

[41] K. Koppel, K. Zalite, K. Voormansik, and T. Jagdhuber, "Sensitivity of Sentinel-1 backscatter to characteristics of buildings," *Int. J. Remote Sens.*, vol. 38, pp. 6298–6318, 2017, doi: 10.1080/01431161.2017.1353160.

[42] F. Yuan and M. E. Bauer, "Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery," *Remote Sens. Environ.*, vol. 106, no. 3, pp. 375–386, 2007, doi: 10.1016/j.rse.2006.09.003.

[43] Microsoft, "US building footprints," 2018. [Online]. Available: https://tiles.arcgis.com/tiles/P3ePLMYs2RVChkJx/arcgis/rest/services/Microsoft_Building_Footprints/VectorTileServer

[44] B. Osmanoğlu, F. Sunar, S. Wdowinski, and E. Cabral-Cano, "Time series analysis of InSAR data: Methods and trends," *ISPRS J. Photogrammetry Remote Sens.*, vol. 115, pp. 90–102, 2016, doi: 10.1016/j.isprsjprs.2015.10.003.

[45] H. Huang et al., "Estimating building height in China from ALOS AW3D30," *ISPRS J. Photogrammetry Remote Sens.*, vol. 185, pp. 146–157, 2022, doi: 10.1016/j.isprsjprs.2022.01.022.

[46] X. Wu, J. Ou, Y. Wen, X. Liu, J. He, and J. Zhang, "Developing a data-fusing method for mapping fine-scale urban three-dimensional building structure," *Sustain. Cities Soc.*, vol. 80, May 2022, Art. no. 103716.
[47] G. Yu, Z. Xie, L. Xuecao, Y. Wang, J. Huang, and X. Yao, "The potential of 3-D building height data to characterize socioeconomic activities: A case study from 38 cities in China," *Remote Sens.*, vol. 14, 2022, Art. no. 2087,.
[48] T. Esch et al., "World settlement footprint 3D-A first three-dimensional survey of the global building stock," *Remote Sens. Environ.*, vol. 270, 2022, Art. no. 112877, doi: 10.1016/j.rse.2021.112877.
[49] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. London, U.K.: Chapman & Hall, 1984.
[50] R. Guida, A. Iodice, and D. Riccio, "Height retrieval of isolated buildings from single high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2967–2979, Jul. 2010.
[51] M. Batty, R. Carvalho, A. Hudson-Smith, R. Milton, D. Smith, and P. Steadman, "Scaling and allometry in the building geometries of Greater London," *Eur. Phys. J. B*, vol. 63, no. 3, pp. 303–314, 2008, doi: 10.1140/epjb/e2008-00251-5.
[52] T. Leichtle, T. Lakes, X. X. Zhu, and H. Taubenböck, "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.*, vol. 78, 2019, Art. no. 101372, doi: 10.1016/j.compenvurbsys.2019.101372.
[53] G. Sun, J. Cheng, A. Zhang, X. Jia, Y. Yao, and Z. Jiao, "Hierarchical fusion of optical and dual-polarized SAR on impervious surface mapping at city scale," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 264–278, 2022, doi: 10.1016/j.isprsjprs.2021.12.008.
[54] X. Chen, X. Jia, and M. Pickering, "A nighttime lights adjusted impervious surface index (NAISI) with integration of landsat imagery and nighttime lights data from international space station," *Int. J. Appl. Earth Observation Geoinformation*, vol. 83, 2019, Art. no. 101889, doi: 10.1016/j.jag.2019.05.022.
[55] J. Gao et al., "Dilution effect of the building area on energy intensity in urban residential buildings," *Nature Commun.*, vol. 10, no. 1, 2019, Art. no. 4944, doi: 10.1038/s41467-019-12852-9.
[56] X. Liu et al., "Global urban expansion offsets climate-driven increases in terrestrial net primary productivity," *Nature Commun.*, vol. 10, no. 1, 2019, Art. no. 5558, doi: 10.1038/s41467-019-13462-1.
[57] X. X. Zhu et al., "So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [Software and data sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 3, pp. 76–89, Sep. 2020.
[58] H. Li, F. Yuan, L. Shen, Y. Liu, Z. Zheng, and X. Zhou, "Improving the WRF/urban modeling system in China by developing a national urban dataset," *Geosci. Front.*, vol. 13, no. 4, 2022, Art. no. 101385, doi: 10.1016/J.GSF.2022.101385.
[59] M. Yu, X. Chen, J. Yang, and S. Miao, "A new perspective on evaluating high-resolution urban climate simulation with urban canopy parameters," *Urban Climate*, vol. 38, 2021, Art. no. 100919, doi: 10.1016/j.uclim.2021.100919.
[60] Z. Jandaghian and U. Berardi, "Comparing urban canopy models for microclimate simulations in weather research and forecasting models," *Sustain. Cities Soc.*, vol. 55, 2020, Art. no. 102025, doi: 10.1016/j.scs.2020.102025.
[61] J. Zhang, Z. Li, and D. Hu, "Effects of urban morphology on thermal comfort at the micro-scale," *Sustain. Cities Soc.*, vol. 86, 2022, Art. no. 104150, doi: 10.1016/j.scs.2022.104150.
[62] M. Ferrando, F. Causone, T. Hong, and Y. Chen, "Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches," *Sustain. Cities Soc.*, vol. 62, 2020, Art. no. 102408, doi: 10.1016/j.scs.2020.102408.
[63] M. K. Hossain and Q. Meng, "A fine-scale spatial analytics of the assessment and mapping of buildings and population at different risk levels of urban flood," *Land Use Policy*, vol. 99, 2020, Art. no. 104829, doi: 10.1016/j.landusepol.2020.104829.
[64] S. Shang, S. Du, S. Du, and S. Zhu, "Estimating building-scale population using multi-source spatial data," *Cities*, vol. 111, 2021, Art. no. 103002, doi: 10.1016/j.cities.2020.103002.
[65] Y. Zheng, J. Ou, G. Chen, X. Wu, and X. Liu, "Mapping building-based spatiotemporal distributions of carbon dioxide emission: A case study in England," *Int. J. Environ. Res. Public Health*, vol. 19, no. 10, 2022, Art. no. 5986, doi: 10.3390/ijerph19105986.
[66] B. Güneralp et al., "Global scenarios of urban density and its impacts on building energy use through 2050," *Proc. Nat. Acad. Sci.*, vol. 114, no. 34, pp. 8945–8950, 2017, doi: 10.1073/pnas.1606035114.

Yangzi Che received the B.S. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2021.

Her research interests include remote sensing, urban spatial analysis, and machine learning, especially machine learning to estimate building height.

Xuecao Li received the B.Sc. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2012, and the Ph.D. degree in ecology from Tsinghua University, Beijing, China, 2016.

He is currently a Professor with the College of Land Science and Technology, China Agricultural University, Beijing. He has authored more than 90 articles in international journals. His research interests include land use/cover change and modeling, global environmental change, and sustainable development.

Qian Shi (Senior Member, IEEE) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, in 2015.

She is an Associate Professor with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China. Her research interests include remote sensing image classification, including deep learning, active learning, and transfer learning.

Xiaoping Liu (Member, IEEE) received the B.S. degree in geography and the Ph.D. degree in remote sensing and geographical information sciences from Sun Yat-sen University, Guangzhou, China, in 2002 and 2008, respectively.

He is a Professor with the School of Geography and Planning, Sun Yat-sen University. He has authored 2 books and more than 100 articles. His research interests include image processing, artificial intelligence, and geographical simulation.