Reviewer #1 (Response to Reviewer)

General comments

This article uses a practical method based on multimodal data to construct the first global scale 3D information dataset of buildings. The data set provided by this study fills the gap in fine-grained building height data globally, which is of great significance for urban morphology research and climate change analysis. The model validation results are comprehensive and promising; however, a more detailed explanation of the technical methods would enhance the paper's clarity (see specific comments). Overall, the paper is well-structured, and the dataset is valuable to urban studies.

Response: thank you very much for your helpful comments and suggestions. We have carefully revised our manuscript and provided a point-by-point response below.

Specific comments

Comment #1: The article uses multiple sources of data for analysis, and it is recommended to add ablation experiments between different data to demonstrate the effectiveness of using the data.

Response: thank you for your question. We evaluate the contributions of features in different categories with added experiments.

We evaluated the impacts of input features using the US dataset to further indicate the efficiency of the synthetic-used multi-source datasets. We used single category (i.e., radar-only, optical-only, terrain-only) and different combinations of features as the model input to estimate the height of buildings (**Table R1**). According to the results, the R^2 and RMSE of radar-only model is 0.56 and 5.24m. The optical features provide effective information of building height estimation, with the model's R^2 increased 0.1 and RMSE decreased 0.7m compared with the radar-only model. The terrain features significantly increase the accuracy of estimation model, with model's R^2 increased 0.1 and RMSE decreased 0.8m compared with model2. Additionally, socioeconomic and vector features slightly improve the ability of model, with R^2 increased 0.05 and RMSE decreased 0.35m compared with model3. These results demonstrate that the synthetic use of these features is practicable and effective for building the height estimation model.

Table R1. Performance of models with different feature combinations.

	Feature combinations	R^2	RMSE (m)
Model1: radar-only	①②	0.56	5.24
Model2: radar + optical	①②③	0.67	4.57
Model3: radar + optical + terrain	①②③④⑤⑥	0.77	3.76
Model4: radar + optical + terrain + socioeconomic	①②③④⑤⑥⑦⑧	0.79	3.65
Model5: radar + optical + terrain + socioeconomic + vector	①②③④⑤⑥⑦⑧⑨	0.82	3.39

① Sentinel1_VV_mean, Sentinel1_VV_std, Sentinel1_VV_percentile5, Sentinel1_VV_percentile25, Sentinel1_VV_percentile50, Sentinel1_VV_percentile75, Sentinel1_VV_percentile95, Sentinel1_VH_mean, Sentinel1_VH_std, Sentinel1_VH_percentile5, Sentinel1_VH_percentile25, Sentinel1_VH_percentile50, Sentinel1_VH_percentile75, Sentinel1_VH_percentile95

② PALSAR_HH_mean, PALSAR_HH_std, PALSAR_HH_percentile5, PALSAR_HH_percentile25, PALSAR_HH_percentile50, PALSAR_HH_percentile75, PALSAR_HH_percentile95, PALSAR_HV_mean, PALSAR_HV_std, PALSAR_HV_percentile5, PALSAR_HV_percentile25, PALSAR_HV_percentile50, PALSAR_HV_percentile75, PALSAR_HV_percentile95

③ Sentinel2_Band2_mean, Sentinel2_Band2_std, Sentinel2_Band2_percentile5, Sentinel2_Band2_percentile25, Sentinel2_Band2_percentile50, Sentinel2_Band2_percentile75, Sentinel2_Band2_percentile95, Sentinel2_Band3_mean, Sentinel2_Band3_std, Sentinel2_Band3_percentile5, Sentinel2_Band3_percentile25, Sentinel2_Band3_percentile50, Sentinel2_Band3_percentile75, Sentinel2_Band3_percentile95, Sentinel2_Band4_mean, Sentinel2_Band4_std, Sentinel2_Band4_percentile5, Sentinel2_Band4_percentile25, Sentinel2_Band4_percentile50, Sentinel2_Band4_percentile75, Sentinel2_Band4_percentile95, Sentinel2_Band8_mean, Sentinel2_Band8_std, Sentinel2_Band8_percentile5, Sentinel2_Band8_percentile25, Sentinel2_Band8_percentile50, Sentinel2_Band8_percentile75, Sentinel2_Band8_percentile95

④ DEM_mean, DEM_std, DEM_percentile5, DEM_percentile25, DEM_percentile50, DEM_percentile75, DEM_percentile95

⑤ DSM_mean, DSM_std, DSM_percentile5, DSM_percentile25, DSM_percentile50, DSM_percentile75, DSM_percentile95

⑥ nDSM_mean, nDSM_std, nnDSM_percentile5, nDSM_percentile25, nDSM_percentile50, nDSM_percentile75, nDSM_percentile95

⑦ population_mean, population_std, population_percentile5, population_percentile25, population_percentile50, population_percentile75, population_percentile95

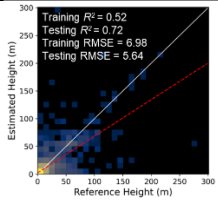
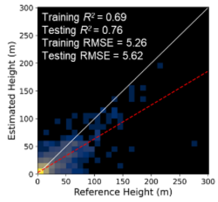
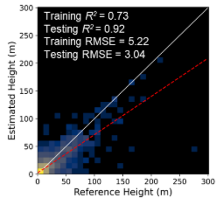
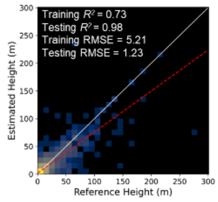
⑧ nighttimelight_mean, nighttimelight_std, nighttimelight_percentile5, nighttimelight_percentile25, nighttimelight_percentile50, nighttimelight_percentile75, nighttimelight_percentile95

⑨ building area, building perimeter

Comment #2: Why did you choose to use XGboost instead of random forest or support vector machine? Please provide additional experiments or explanations.

Response: thank you for your question. We have tested the performance of machine learning models in preliminary experiments, and finally used eXtreme Gradient Boosting (XGB) model due to its accuracy and efficiency (**Table R2**). Specifically, we compared the model performance of Decision Tree (DT), Random Forest (RF), Gradient Boosting Regression (GB), XGB, and Support Vector Machine (SVM), using 135291 training samples and 15033 testing samples in Africa. The results showed that the XGB model has the highest accuracy (testing R^2 = 0.733, RMSE = 5.213m). The performance of GB model is slightly inferior to that of the XGB model with testing R^2 of 0.731 and RMSE of 5.223m. The results of RF model showed overall underestimation with lower accuracy (testing R^2 = 0.688, RMSE = 5.625m). And the results of DT model show significant overestimation for building heights below 50m. The SVM model is ineffective to estimate height of buildings. These experiments were conducted using Intel® Core™ i9 with Python 3.9.

Table R2. Performance of models with different feature combinations.

Model	Model parameters	Training R^2	Training RMSE (m)	Testing R^2	Testing RMSE (m)	Training time	Scatter plot
DT	'max_depth': 10, 'min_samples_split': 2, 'min_samples_leaf': 1	0.723	5.643	0.519	6.988	10s	
RF	'n_estimators': 1000, 'max_depth': 10, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'bootstrap': True	0.758	5.266	0.688	5.625	514s	
GB	'n_estimators': 1000, 'max_depth': 10, 'min_samples_split': 2, 'min_samples_leaf': 1, 'learning_rate': 0.1, 'subsample': 0.8, 'max_features': 'sqrt'	0.919	3.041	0.731	5.223	67s	
XGB	'max_depth': 10, 'eta': 0.1," 'lambda': 30, 'alpha': 20, 'colsample_bytree':0.7, 'learning_rate' : 0.1, 'n_estimators': 1000	**0.986**	**1.234**	**0.733**	**5.213**	**46s**	
SVM	'kernel': 'rbf', 'C': 1.0, 'epsilon': 0.1	0.190	9.064	0.169	9.768	3179s	/

Comment #3: Please explain the specific operation of manual measurement in section 4.2 and the basis for the authenticity of manual height measurement values.

Response: thank you for your comment. We used manually measured building heights on Google Earth Pro as the reference height to evaluate the quality of our height results. We measured the length of the ridge lines of the 3D building models using a 3D measurement tool as the height of individual building (**Fig. R1**). The vertical information of these 3D building models is obtained using Google Street View, aerial photographs, and satellite images.

The selecting points in manual measurements can affect the precision of measured height results, especially for buildings with complex shapes. To minimize the error, we measured the smooth ridge lines of buildings. We tested the stability of the measurement result by repeatedly measuring the building heights five times. The measurement results of a 130m building fluctuate within 1m, indicating the measurement error is relatively small. Also, the measurement results of a 12.8m building fluctuate within 0.1m. This relative measurement error is less than 1%, suggesting the method is reliable to obtain reference heights.

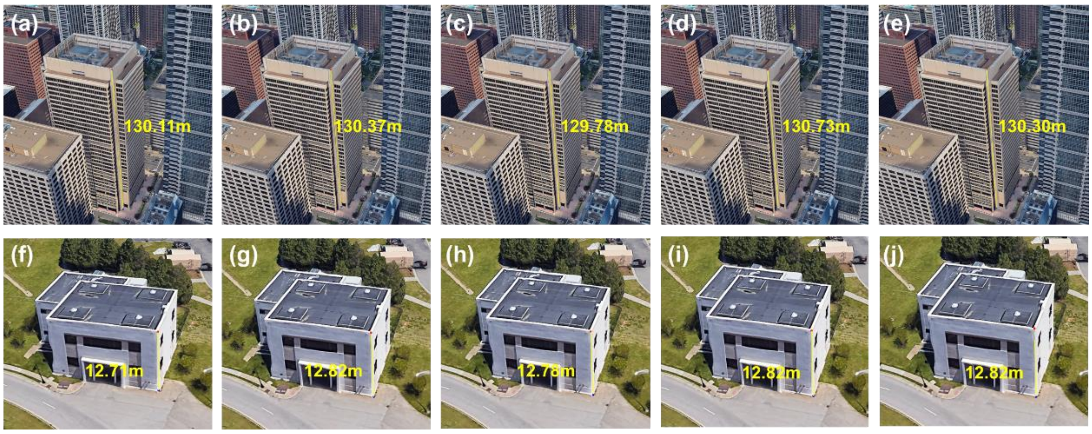


Figure R1. Results of measuring the same building five times using the Google 3D measurement tool.

Comment #4: What are the advantages of building scale 3D data over course scale 3D data set?

Response: thank you for your question. Building-scale 3D data enhances surface morphology with finest-scale details, which can support a wider range of urban studies compared to coarse-scale building height datasets.

Building-scale 3D data refines urban analysis and simulation by offering precise height information for building structures. The vector data scale avoids confusion with other surface objects, such as impervious surfaces (e.g., roads and parking lots). This allows for more detailed urban morphology analysis and simulations (e.g., digital twin) compared to coarse-scale height datasets, including building stock calculation, building carbon emission analysis, fine-scale population simulations, and thermal environment analysis. Moreover, building-scale 3D dataset can support a broader range of fine-scale analyses when combined with increasingly high-resolution datasets. As the resolution of remote sensing products have advanced, high-resolution datasets are now available, such as 1m land cover and 1m tree height datasets. Building-scale datasets can support urban analysis at these high-resolution levels. For example, building-scale 3D datasets and tree height dataset can be used to analyzed the cooling effects of shadows from buildings and trees in urban area *(Tolan et al., 2023)*. Additionally, building-scale dataset can also aggerated to other resolutions for analysis under difference scales. Relevant contents can be found in our manuscript as:

"Besides, our estimated heights accurately showed the spatial heterogeneity of building heights between densely high-rise buildings and low-rise buildings, benefiting from a finer resolution at the scale of individual buildings. Conversely, the resolution of the other three datasets is insufficient to reflect the spatial heterogeneity of building heights due to the significant differences in building height within each pixel." (page 14, line 312-315)

"The 3D-GloBFP map is the first individual building height dataset to depict the most detailed building three-dimensional morphology worldwide, offering great potential to support studies ranging from macro-scale global analyses to micro-scale investigations within urban areas. Our developed dataset provides precise height information and serve as the base input for urban analysis and simulations, such as climate modeling (He et al., 2019), population simulation (Zhao et al., 2021), building function classification (Zheng et al., 2024), and disaster assessments (Hossain and Meng, 2020). Moreover, our dataset also contributes to studies on the interaction

between human society and ecosystems (Zhong et al., 2021; Rodriguez Mendez et al., 2024; Güneralp et al., 2017; Arehart et al., 2022), such as Urban Heat Island (UHI) assessment (Li et al., 2020d), carbon footprint accounting (Li et al., 2024a), building shade studies (Watanabe et al., 2014), and building stock analysis (Frantz et al., 2023). These studies can further contribute to addressing environmental issues related to anthropogenic activities, thereby promoting the achievement of sustainable development." (page 24-25, line 489-498)

Reference:

Tolan, J., Yang, H.-I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tiecke, T., and Couprie, C.: Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar, Remote Sensing of Environment, 300, 113888, https://doi.org/10.1016/j.rse.2023.113888, 2024.

He, X., Li, Y., Wang, X., Chen, L., Yu, B., Zhang, Y., and Miao, S.: High-resolution dataset of urban canopy parameters for Beijing and its application to the integrated WRF/Urban modelling system, Journal of Cleaner Production, 208, 373-383, https://doi.org/10.1016/j.jclepro.2018.10.086, 2019.

Zhao, X., Zhou, Y., Chen, W., Li, X., Li, X., and Li, D.: Mapping hourly population dynamics using remotely sensed and geospatial data: a case study in Beijing, China, GIScience & Remote Sensing, 58, 717-732, 10.1080/15481603.2021.1935128, 2021.

Zheng, Y., Zhang, X., Ou, J., and Liu, X.: Identifying building function using multisource data: A case study of China's three major urban agglomerations, Sustainable Cities and Society, 108, 105498, https://doi.org/10.1016/j.scs.2024.105498, 2024.

Hossain, M. K. and Meng, Q.: A fine-scale spatial analytics of the assessment and mapping of buildings and population at different risk levels of urban flood, Land Use Policy, 99, 104829, https://doi.org/10.1016/j.landusepol.2020.104829, 2020.

Zhong, X., Hu, M., Deetman, S., Steubing, B., Lin, H. X., Hernandez, G. A., Harpprecht, C., Zhang, C., Tukker, A., and Behrens, P.: Global greenhouse gas emissions from residential and commercial building materials and mitigation strategies to 2060, Nature Communications, 12, 6126, 10.1038/s41467-021-26212-z, 2021.

Rodriguez Mendez, Q., Fuss, S., Lück, S., and Creutzig, F.: Assessing global urban CO2 removal, Nature Cities, 10.1038/s44284-024-00069-x, 2024.

Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P. L., Fragkias, M., Li, X., and Seto, K. C.: Global scenarios of urban density and its impacts on building energy use through 2050, Proceedings of the National Academy of Sciences, 114, 8945-8950, 10.1073/pnas.1606035114, 2017.

Arehart, J., Pomponi, F., D'Amico, B., and Srubar Iii, W.: A new estimate of building floor space in North America, Environmental Science & Technology, 55, 10.1021/acs.est.0c05081, 2021.

Li, Y., Schubert, S., Kropp, J. P., and Rybski, D.: On the influence of density and morphology on the Urban Heat Island intensity, Nature Communications, 11, 2647, 10.1038/s41467-020-16461-9, 2020d.

Li, M., Koks, E., Taubenböck, H., and van Vliet, J.: Continental-scale mapping and analysis of 3D building structure, Remote Sensing of Environment, 245, 111859-111859, https://doi.org/10.1016/j.rse.2020.111859, 2020a.

Watanabe, S., Nagano, K., Ishii, J., and Horikoshi, T.: Evaluation of outdoor thermal comfort in sunlight, building shade, and pergola shade during summer in a humid subtropical region, Building and Environment, 82, 556-565, https://doi.org/10.1016/j.buildenv.2014.10.002, 2014.

Frantz, D., Schug, F., Wiedenhofer, D., Baumgart, A., Virág, D., Cooper, S., Gómez-Medina, C., Lehmann, F., Udelhoven, T., van der Linden, S., Hostert, P., and Haberl, H.: Unveiling patterns in human dominated landscapes through mapping the mass of US built structures, Nature Communications, 14, 8014, 10.1038/s41467-023-43755-5,

2023.

Comment #5: Please explain in detail how to aggregate building scale height data to coarse resolution scales for validation.

Response: thank you for your question. We aggregated the building-scale height dataset to 1km-resolution to compare with other existing coarse-resolution products according to Eq1:

$$\overline{H} = \frac{\sum_{i=1}^{N} H_i}{N} \tag{1}$$

where $\overline{H}$ is the aggregated results in 1km pixel, N is the number of buildings in 1km pixel, and H_i is the height of buildings in the 1km pixel.

Specifically, we calculated the average height of all buildings located within each grid cell. In this process, each 1km pixel represents the height of buildings with other built-up regions excluded (e.g., roads and parking lots) in the computation. Details can be found in our manuscript as:

"We also aggregated the high-resolution data at 1 km resolution to align with the low-resolution data by calculating the average height of all buildings located within each grid cell. This approach allows us to compare the differences with the reference data at a consistent resolution across all datasets." (page 10, line 220-222)

Comment #6: What are the contributions of statistical values in the model?

Response: thank you for raising this concern. The statistical values can improve the accuracy of the models, especially for high-rise buildings. We compared the RMSEs for models trained with mean values alone versus models trained with all statistical values (i.e., standard deviation, 5%, 25%, 50%, 75%, and 95% quantiles) of pixels intersecting with each building boundary (**Fig. R2**). We evaluated the contributions of statistical values using RMSEs across different height intervals (i.e., <10m, 10-20m, 20-30m, 30-50m, 50-100m, >100m). The results show that the statistical values slightly enhance the accuracy of estimating results in low-rise buildings and significantly increased the accuracy of high-rise buildings. For instance, the model using statistical values reduce the RMSE by 3m in estimating buildings with height of 30-50m. Furthermore, these quantiles reduce the RMSE by approximately 6m and 7m in estimating buildings with height of 50-100m, and >100m. The statistical values can reflect the complex variations of pixel values within building boundaries of high-rise buildings, providing the model with comprehensive information and enhanced its accuracy. However, the statistical value does not significantly increase the model accuracy for low-rise buildings (i.e., buildings lower than 20m), due to the similar values of pixels within buildings.

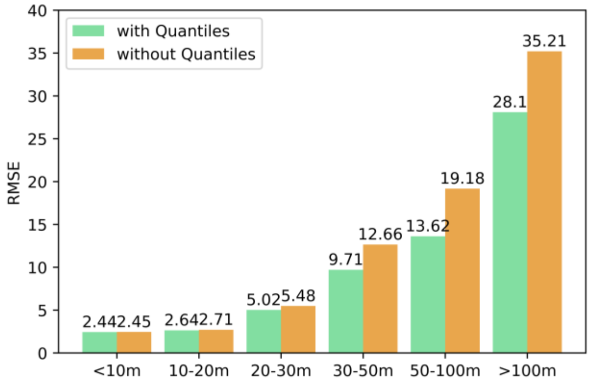


Figure R2. RMSEs of models with and without quantiles as input features.