Overall, this is a valuable and timely study. Long-term, annual building-height reconstruction remains an important gap in urban remote sensing, especially at fine spatial resolution and national scale. The proposed dataset has clear potential to support studies of three-dimensional urban growth and renewal in China. Nevertheless, several methodological and interpretive issues still deserve further clarification. Addressing these points would improve the robustness, transparency, and broader usability of the dataset.

We would like to express our sincere gratitude to the editor and the reviewers for their valuable time, constructive comments, and thoughtful suggestions on our manuscript. In the following sections, we address each of the reviewer's points in detail, organizing our responses in a question (black)-response (blue)-original (red) format to ensure clarity.

1. On Page 9, Lines 205–210 and Page 10, Lines 217–219, the manuscript states that reconstructed 1 km VV backscatter is used as an independent variable for XGB-ReVV during 1990–2014. Since this VV layer is reconstructed from Landsat-derived variables and terrain/DSM information, its uncertainty may be propagated into the subsequent 30 m building-height prediction. Although the authors report satisfactory performance of the reconstruction model in the Beijing–Tianjin–Hebei region, it remains unclear whether this performance can be generalized to the whole of China, particularly across regions with different urban forms, climate conditions, and surface characteristics. I suggest that the authors add a short discussion on the robustness and potential error propagation of the reconstructed VV variable, and clarify the extent to which the regional validation supports its nationwide application.

We sincerely thank the reviewer for this constructive suggestion. It is agreed that the nationwide robustness of the 1 km VV reconstruction and its potential error propagation need to be evaluated to ensure the reliability of building height predictions during 1990–2014. To verify whether the 1 km VV reconstruction model is generalizable beyond the Beijing–Tianjin–Hebei region, validation is conducted across 18 cities representing 9 distinct ecological and geographic regions across China, using the actual 2019 Sentinel-1 VV observations as the reference. These cities are characterized by highly diverse urban forms, local climates, and topographic conditions. The reconstruction accuracy metrics are calculated based on the actual 2019 Sentinel-1 VV observations. The results summarized in Table 1 demonstrate the high robustness of the reconstructed 1 km VV at the national scale. Regarding the metrics for different regions, the lowest RMSE is obtained in Lanzhou, with an RMSE of 1.57 dB and an R^2 of 0.90, whereas the highest RMSE is observed in Guangzhou, with an RMSE of 2.45 dB and an R^2 of 0.84. This indicates that although the VV regression model is generally reliable, its predictive ability varies across regions. The national pooled RMSE is 1.97 dB, which is slightly larger than the 1.5 dB RMSE reported for the Beijing–Tianjin–Hebei region in the original article, with an rRMSE of 36.7%, a mean absolute error (MAE) of 1.34 dB, and an R^2 of 0.89. These results confirm that the reconstruction model can be safely applied at the national scale.

Table 1 The VV regression model accuracy in different cities across China in 2019

Geographic Regions	**Representative Cities**	**Number of Pixels**	**RMSE (dB)**	**MAE (dB)**	**rRMSE**	**R^2**
Northern China Plain	Beijing	645	1.74	1.17	34.6%	0.90
	Shijiazhuang	398	1.91	1.29	27.4%	0.89
Northeast Plain	Harbin	451	1.92	1.35	28.9%	0.89
	Changchun	422	1.77	1.21	22.0%	0.87
Northwest Arid Zone	Lanzhou	160	1.57	1.26	34.6%	0.90
	Yinchuan	120	1.98	1.37	31.0%	0.85
Yangtze River Delta	Shanghai	548	1.98	1.29	28.6%	0.90
	Suzhou	463	1.97	1.26	34.1%	0.90
Yangtze River Basin	Wuhan	380	2.16	1.31	35.0%	0.89
	Changsha	320	2.34	1.41	38.1%	0.89
Sichuan Basin	Chengdu	493	1.97	1.29	38.1%	0.89
	Chongqing	179	1.74	1.17	30.7%	0.93
Southwest Highland	Kunming	130	1.68	1.18	40.1%	0.90
	Guiyang	250	1.81	1.31	31.3%	0.89
	Fuzhou	172	2.37	1.61	49.3%	0.82

Southeast Coastal	Xiamen	262	1.76	1.19	29.5%	0.92
Pearl River Delta	Guangzhou	267	2.45	1.35	56.1%	0.84
	Shenzhen	259	2.05	1.34	43.6%	0.86
National Pooled	All 59 cities	13997	1.97	1.34	36.7%	0.89

In addition to the nationwide validation results described above, the reconstructed VV backscatter is not used as an isolated predictor in the 30 m building-height model. Instead, it is integrated with the 30 m Landsat/Sentinel spectral bands and topographic variables to provide a complementary macro-scale structural constraint. In this framework, the 1 km reconstructed VV is introduced as a macro-scale spatial prior because its coarse spatial resolution mainly reflects regional built-up density and structural volume. By contrast, the 30 m Landsat/Sentinel spectral bands and topographic variables provide finer-scale spatial information, allowing local spatial and temporal details, such as building edges and height changes, to be captured (Frantz et al., 2021; Li et al., 2020; Dong et al., 2024). Therefore, potential local regression errors in the reconstructed VV are prevented from directly distorting the final 30 m building height maps through its integration with other remote-sensing variables. Although some false regressions may still exist and are reflected by the black squares in the confidence layer shown in Question 9(2) in the response letter to Reviewer #1 (Fig. 1 in this response letter), the overall spatial pattern of the final 30 m building height maps is preserved. Moreover, higher accuracy is achieved when the regressed VV is used as an input variable, as demonstrated in Fig. 9 of the original manuscript (Fig. 2 in this response letter).

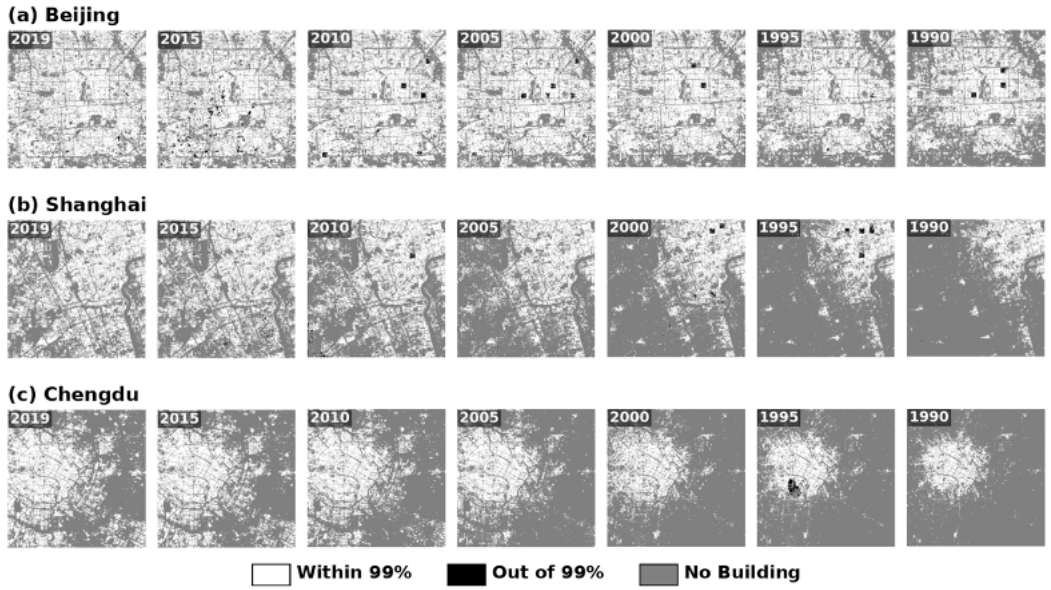

Fig. 1. Binary confidence layers for Beijing, Shanghai, and Chengdu from 1990 to 2019. Cropped from Fig. 13 in the response letter to Reviewer #1.

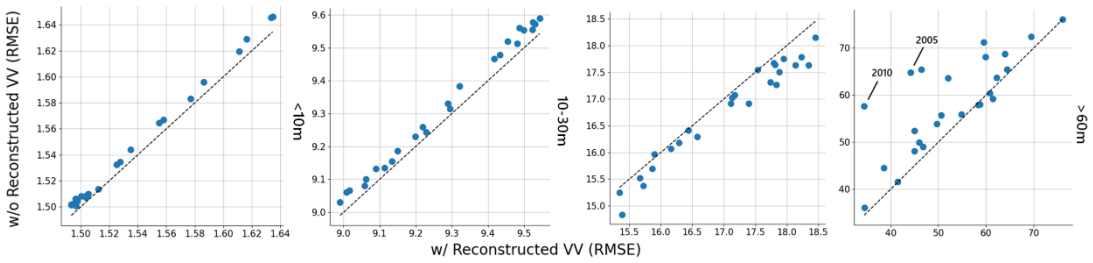

Fig. 2. RMSE differences between models with and without reconstructed VV. Cropped from Fig. 9 of the original manuscript.

We again thank the reviewer for this valuable suggestion. Section 3.1.2 of the manuscript is updated to include a brief discussion on the reliability of the VV regression:

> "…
> The regression framework is further extended nationwide in this study, and the regressed VV backscatter is utilized as an independent variable in subsequent XGBoost-based models for building height prediction. The nationwide robustness of the 1 km reconstructed VV backscatter is confirmed through validation across diverse geographic regions characterized by different urban forms, local climates, and topographic conditions. Detailed accuracy metrics are provided in Supplementary Table S3.
> …"

A detailed discussion is added in Section 4.6 of the manuscript to clarify the data uncertainty potentially introduced by the VV reconstruction:

> "…
> The joint use of reconstructed VV and spectral and topographic variables suppresses random reconstruction noise during building height prediction. Although localized inaccuracies may appear as discrete low-confidence areas in heterogeneous landscapes, they do not distort the overall spatial patterns of the final 30 m building height maps.
> …"

References:

[1] Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., and Hostert, P.: 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, 2021

[2] Li, X., Zhou, Y., Gong, P., Seto, K. C., and Clinton, N.: 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, 2020

[3] Dong, B., Zheng Q., Lin Y., Chen B., Ye Z., Huang C., Tong C., Li S., Deng J., and Wang, K.: Integrating physical model-based features and spatial contextual information to estimate building height in complex urban areas, International Journal of Applied Earth Observation and Geoinformation, 126, 103625, https://doi.org/10.1016/j.jag.2023.103625, 2024

2. On Page 7, Lines 148–151, the 2019 reference height data are derived from Baidu Maps building footprints, where floor counts are converted to height using a simple assumption of 3 m per floor. Although this approach is supported by Liu et al. (2021), the cited validation was based on a limited building-height assessment of only 519 buildings in Shenyang, with a reported mean height deviation of approximately 1 m and an accuracy of 86.8%. This raises the question of whether the same level of accuracy can be assumed for nationwide and multi-temporal building-height mapping. In addition, Section 4.1 would benefit from a clearer explanation of how the 2019 reference heights are transferred backward to 1990. In particular, if a pixel experienced in-situ height changes over the past three decades, such as demolition followed by reconstruction of a taller building, it would be helpful to clarify how such cases are handled during annual reference sample generation.

We thank the reviewer for these comments. Further scientific justification for the 3 m floor height assumption is provided, and the mathematical logic of our backward reference transfer mechanism is detailed in the response below.

Although the 86.8% accuracy is initially reported in Shenyang, the 3 m-per-floor conversion factor is widely validated and adopted as a highly stable vertical conversion parameter in national-scale urban morphology studies across China (Zhang, Zhao and Long, 2025; Wu et al., 2023). Importantly, the dataset is generated at a 30 m grid resolution, where each pixel represents an aggregated measure of building height derived from all sub-pixel structures within the corresponding cell. At this spatial scale, fine-grained architectural variations, such as high-ceiling spaces, rooftop decorations, and floor-level heterogeneity, are inherently smoothed through spatial aggregation, consistent with the scale-dependent nature of remote sensing observations (Woodcock and Strahler, 1987; Zhang et al., 2017). Consequently, the 3 m-per-floor assumption is used as an effective first-order approximation for large-scale gridded urban morphology representation, rather than as an exact building-level physical parameter.

To prevent historical reference labels from being contaminated by urban renewal such as in-situ demolition and taller reconstruction, the Continuous Change Detection and Classification (CCDC) algorithm is utilized as a temporal filter. For any target historical year t $(1990 \leq t \leq 2018)$ and pixel (x, y), the annual reference height $H_t(x, y)$ is determined by:

$$H_t(x,y) = \begin{cases} H_{2019}(x,y) & if\ T_{break}(x,y) = 0\ or\ T_{break}(x,y) < t \\ Excluded & otherwise \end{cases} \quad (1)$$

where $T_{break}(x, y)$ is the latest breakpoint year detected by the CCDC algorithm for that pixel, with 0 meaning no breakpoint detected. This method guarantees that the 2019 reference data set is only projected backward to year t if the land surface has remained spectrally stable and undisturbed since year t. If any spectral breakpoint is detected within the window $[t, 2019]$, the pixel is completely excluded from the training-validation pool for year t. This strictly prevents label mismatch or label pollution caused by urban renewal, ensuring high purity in our historical reference samples.

We thank the reviewer again for this valuable suggestion. Section 2.6 of the original manuscript is modified to include the following content:

"…

The 3 m-per-floor conversion is a statistically robust approximation for grid-level aggregation supported by different researches, although being unable to fully capture localized floor-height variations in specialized structures (Zhang, Zhao and Long, 2025; Wu et al., 2023).

…"

Section 3.1.1 of the original manuscript is updated, including the $H_t(x,y)$ formula to explain the annual reference extraction process using a more systematic approach:

"…

Given a target historical year t $(1990 \leq t \leq 2018)$ and pixel (x,y), the annual reference height $H_t(x,y)$ is determined by:

$$H_t(x,y) = \begin{cases} H_{2019}(x,y) & if\ T_{break}(x,y) = 0\ or\ T_{break}(x,y) < t \\ Excluded & otherwise \end{cases}$$

where $T_{break}(x,y)$ is the latest breakpoint year detected by the CCDC algorithm for that pixel, with 0 meaning no breakpoint detected. This method guarantees that the 2019 reference height is only projected backward to year t if the land surface has remained spectrally stable and undisturbed since year t. If any spectral breakpoint is detected within the window $[t, 2019]$, the pixel is completely excluded from the training-validation pool for year t.

…"

In addition, a comprehensive uncertainty analysis is added in Section 4.6, which quantifies and discusses potential error sources including those stemming from reference data (Question 2), CCDC-based temporal extraction (Question 3), and the use of spatial coordinates (Question 5). Furthermore, as detailed in our response to Reviewer #1 in Question 9, confidence layers are produced to represent these uncertainties spatially. These layers, along with a full methodological discussion, have been incorporated into the final manuscript to provide a transparent assessment of our model's predictive reliability.

References:

[1] Zhang, Y., Zhao, H., and Long, Y.: CMAB: A Multi-Attribute Building Dataset of China, Scientific Data, 12(1), 430, https://doi.org/10.1038/s41597-025-04730-5, 2025

[2] Wu, W. B., Ma, J., Banzhaf, E., Meadows, M. E., Yu, Z. W., Guo, F. X., Sengupta, D., Cai, X. X., and Zhao, B.: 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, 2023

[3] Kleman, J. and Fagerlund, E.: Influence of different nitrogen and irrigation treatments on the spectral reflectance of barley, Remote Sensing of Environment, 21.1, 1-14, https://doi.org/10.1016/0034-4257(87)90002-2, 1987

[4] Zhang, L., Weng, Q. and Shao, Z.: An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015, Remote sensing of environment, 201, 99-114, https://doi.org/10.1016/j.rse.2017.08.036, 2017

3. On Page 9, Lines 195–198, the authors state that applying the 1999 CCDC mask uniformly to 1990–1999 has "negligible effects" on reference data accuracy. However, Page 13, Lines 271–274 show that the validation accuracy of annual reference samples remains only around 65%–70% during 1990–2002. Therefore, I suggest replacing "negligible effects" with a more cautious statement, such as "limited but non-negligible uncertainty may remain for the pre-2000 period." The authors may also clarify in Section 4.1 or Section 6 that early-year estimates are more suitable for regional-scale or trend-level analyses, rather than for strong interpretation of pixel-level inter-annual changes.

We completely agree with the reviewer's insightful comment and sincerely appreciate the suggestion reviewer has provided. Acknowledging early-year uncertainties in a transparent manner improves the scientific rigor of our dataset. Section 3.1.1 of the manuscript is updated, replacing "negligible effects" with "limited but non-negligible uncertainty may remain for the pre-2000 period."

> "…
> which may introduce minor yet non-negligible uncertainty to the pre-2000 period, keeping the overall accuracy around 65%.
> …"

Furthermore, a clear guidance for data users is added in the revised Section 4.6. It is clarified that while the early-year estimates (1990–1999) are highly reliable for capturing regional-scale urban expansion, provincial building volume trends, and metropolitan-level development trajectories, users should exercise caution when conducting fine-grained, pixel-level, year-to-year causal interpretations in the 1990s due to the relative scarcity of CCDC change detections and historical Landsat observations in that era:

> "…
> The relative sparsity of CCDC-based change detections and the constrained availability of historical Landsat observations during the 1990s resulted in relative lower transfer accuracy of reference data in 1990-1999. The 1990–1999 estimates still demonstrate high reliability for capturing regional-scale urban expansion, provincial building volume trends, and metropolitan-level development trajectories, although caution is advised when conducting fine-grained, pixel-level, year-to-year causal interpretations for this period.
> …"

4. On Page 10–11, Lines 225–228, the manuscript states that the annual reference height samples are randomly split into 90% training and 10% testing sets. Since both subsets are derived from the same CCDC-filtered and 2019-reference-based sample generation procedure, this evaluation is better interpreted as testing the consistency of model fitting against CCDC-derived reference labels, rather than as a fully independent validation of spatial or temporal generalization. I suggest that the authors make this distinction clearer when presenting the accuracy results, so that the reported RMSE, MAE, and R^2 are not over-interpreted as completely independent validation metrics.

We sincerely appreciate this insightful methodological observation provided by the reviewer, which

encourages further refinement of the validation discussion. It is clarified that each year is treated as an independent entity in our modeling framework. Specifically, the training and validation sets for a given year *t* are derived exclusively from the CCDC-filtered reference pool for that year, ensuring that no temporal data leakage occurs between years. Although the reference labels originate from the same long-term CCDC-processed dataset, the labels for each year *t* are independently generated and are used only to train and validate the model created specifically for that corresponding year, rather than to fit a single model using all reference data generated across years. Consequently, the annual accuracy metrics, including RMSE, MAE, and R^2, reflect the model performance on year-specific independent data, as described in Section 3.2 of the original manuscript:

"XGBoost models are adopted due to their demonstrated computational efficiency and predictive accuracy in prior architectural height regression studies (Che et al., 2024; Stipek et al., 2024). For each year, **an independent XGBoost model is trained with year-specific variables extracted from remote sensing data**, paired with corresponding reference building heights derived from CCDC masks. This approach maximizes the utility of annual data characteristics in the estimation process."

Furthermore, independent validations of spatial and temporal generalization, in addition to the annual accuracy metrics, are performed and demonstrated in the original manuscript. For spatial independent validation, the annual estimates are validated using a completely independent external database of housing transaction records from Lianjia, which contains real community construction years, as shown in Fig. 3 below, which is also Fig. 1 in the Response to Reviewer #1. Every city validated using the Lianjia dataset is excluded from the sampled cities, ensuring minimum correlation with the training set. In this way, the spatial transferability of the models is verified.

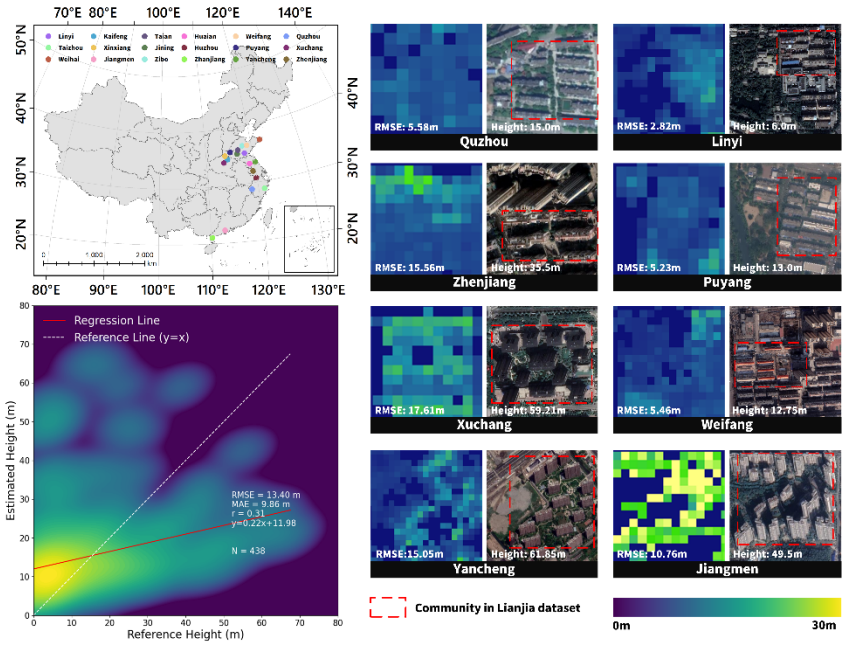

Fig. 3. Accuracy assessment of model performance in cities excluded from the reference samples using Lianjia real-estate data.

For temporal independent validation, historical visual audits are conducted in multiple non-sampled cities that are completely excluded from the training dataset, such as Handan, Jining, and Anyang, as shown in Fig. 4 below, which is also Fig. 6 in the Response to Reviewer #1 and originally Fig. 14 of the original manuscript. These audits demonstrate the temporal accuracy of the models. As shown in Fig. 4 in the response letter, the dataset matches the historical construction timeline and reflects the consistent dynamics of urban development.

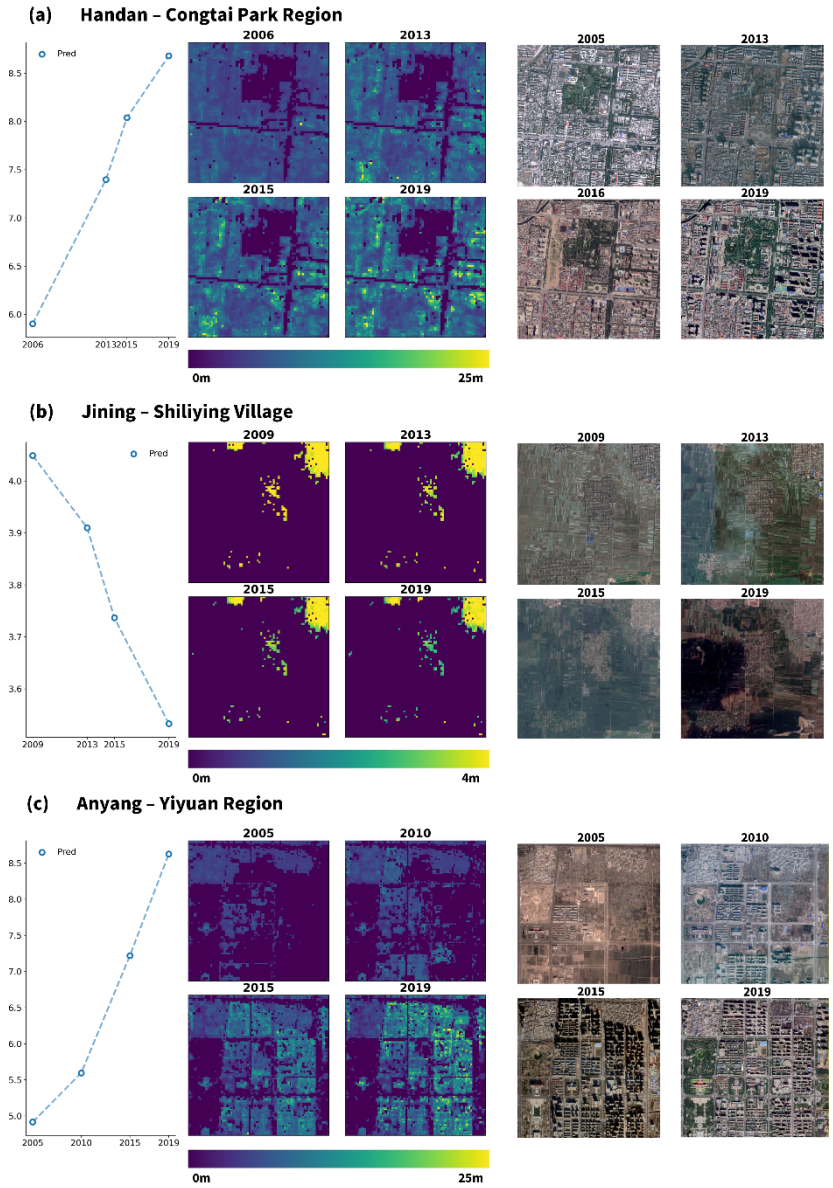

Fig. 4 (the revised Fig. 14 in the original manuscript). Comparison between mapped height changes and historical remote sensing images in non-sampled cities. (a) Congtai Park region, Handan; (b) Shiliying region, Jining; (c) Yiyuan region, Anyang. Remote sensing images are from © Google.

Finally, the results are benchmarked against entirely independent external datasets that do not share our reference data or modeling framework for cross-dataset validation, including WSF 3D, Ma's LiDAR-based height dataset and Wu's CNBH-10m as shown in Fig. 5 below, which is also Fig. 12 in the Response to Reviewer #1 (originally Fig. 10 of the original manuscript, updated due

to computational error). This cross-dataset validation strategy ensures the independence of our model evaluation and underscores the robustness of the product when compared against datasets derived from different sources.

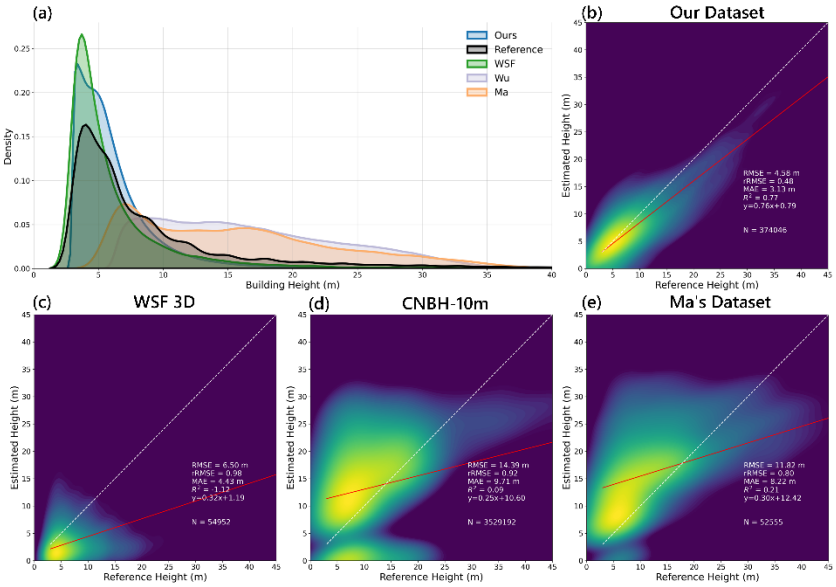

Fig. 5 (the revised Fig. 10 in the original manuscript). Comparison of reference height, our dataset and other datasets in Beijing in 2019. (a) Height distribution of different datasets; (b) Scatter plot of our dataset and reference heights; (c) Scatter plot of WSF 3D (Esch et al., 2022) and reference heights; (d) Scatter plot of CNBH-10m (Wu et al., 2023) and reference heights; (e) Scatter plot of Ma's dataset (Ma et al., 2024) and reference heights.

We thank the reviewer for this valuable comment. According to the reviewer's suggestion, the year-specific nature of the models is further clarified in the modified manuscript. For example, Section 3.3 is revised as follows:

"…
The accuracy of the three XGBoost-based models is further evaluated for three representative years: XGB-ReVV in 2014, XGB-S1 in 2017, and XGB-S1S2 in 2019 (Fig. 7).
…"

5. Page 16, Lines 335–342 show that latitude and longitude are the most important predictors, even exceeding most optical and SAR variables. Meanwhile, Page 15–16, Lines 318–326 report clear regional discrepancies in model performance, especially poorer performance in parts of southern China and the Pearl River Delta. This suggests that the model may partly rely on regional urban-form priors rather than only pixel-level remote-sensing signals. I recommend that the authors add a short discussion on whether the strong importance of geographic coordinates helps capture spatial heterogeneity, or whether it may introduce risks when extrapolating to regions or cities with limited representation in the training samples.

We thank the reviewer for raising this key geographic modeling question. In national-scale modeling,

latitude and longitude serve as spatial trend surface predictors that capture macro-geographic variations in urban morphology (Wu et al., 2023; Chen and Zhao, 2022; Meyer et al., 2019). In China, building forms and zoning regulations are heavily influenced by geographic locations. For example, northern cities have strict regulations for building spacing and heights to ensure winter solar access, resulting in lower building density and regular heights, whereas southern and coastal cities exhibit high-density, compact, and high-rise vertical patterns. (Yang, Xuan and Zhou, 2022; Wang et al., 2024). Introducing coordinates allows the XGBoost model to capture these regional urban-form priors and resolve spatial non-stationarity.

While these coordinates effectively account for spatial variations, their impact on temporal dynamics remains neutral. Specifically, since latitude and longitude are static variables over time, they cannot explain any temporal variations. Thus, all predicted annual height growth, demolition, and reconstruction dynamics over the 30-year period are strictly driven by the time-varying, physical remote sensing features, which are Landsat and Sentinel spectral composites and radar backscatter. This prevents coordinates from causing temporal extrapolation failures (Meyer et al., 2019)

Regarding the lower accuracy in parts of southern China including the Pearl River Delta, apart from coordinate overfitting, natural environmental physical noise is usually an inevitable cause. Frequent cloud cover and heavy precipitation in Southern China degrade the quality of annual optical composites (Zhang and Weng, 2016). Moreover, southern cities are characterized by dense sub-tropical evergreen vegetation, which interacts with high-density vertical structures, causing severe radar double-bounce and volume scattering noise in 30 m pixels (Li et al., 2020).

Upon model training, while highly autocorrelated coordinates can lead to overfitting, this risk is effectively controlled in our model, as presented in Fig. 2 in the Response to Reviewer #1 (Fig. 3 in this response letter). Our dataset is trained on approximately 12 million pixels across 59 geographically dispersed cities, which ensures a highly continuous and representative coordinate space.

To further evaluate the influence of geographic coordinates and mitigate potential concerns regarding spatial overfitting, an additional ablation experiment is conducted. A separate model is trained for the year 2019 using an identical training protocol, but excluding latitude and longitude as input predictors. The accuracy metrics of two models are presented in Table 2.

Table 2 Accuracy of model w/ and w/o coordinates in 2019

XGB-S1S2 in 2019	**RMSE (m)**	**MAE (m)**	**rRMSE**	**R^2**
w/ coords	**3.90**	**2.70**	**61.0%**	**0.80**
w/o coords	4.06	2.88	63.2%	0.78

The results indicated that while the model excluding coordinates maintained consistent predictive trends, the inclusion of geographic coordinates yielded superior performance across all evaluated metrics, having the model with coordinates acquiring a RMSE of 3.90m, MAE of 2.70m and R^2 of 0.80, the model without coordinates falling behind with a RMSE of 4.06m, MAE of 2.88m and R^2

of 0.78. This improvement confirms that geographic coordinates effectively capture regional spatial heterogeneity that are not fully represented by remote-sensing signals alone. Geographic coordinates function as a supportive regional prior, thereby reinforcing the model's reliability across diverse urban landscapes.

To clarify the role of geographic coordinates and address potential concerns regarding spatial overfitting, Section 4.3 has been revised, explicitly distinguishing between the role of coordinates as macro-scale spatial priors and the reliance on time-varying remote sensing signals for detecting annual changes, providing a more robust validation of our model's spatial-temporal consistency:

> "…
> Geographic coordinates serve as essential spatial trend surface predictors that allow the XGBoost model to resolve spatial non-stationarity across China's diverse urban morphologies. By functioning as supportive regional priors, these coordinates capture macro-scale variations that are not fully represented by remote-sensing signals alone (Chen and Zhao, 2022). On the other hand, as these variables are static over time, they remain decoupled from inter-annual temporal dynamics, ensuring that all detected height changes are driven solely by time-varying physical features (Meyer et al., 2019)."
> …"

Section 4.6 is updated as well to further discuss the uncertainty coordinates may introduce to the dataset:

> "…
> Although geographic coordinates are essential for capturing regional spatial trends, their may introduce potential risks of spatial overfitting. Because the training dataset covers most of China's topographical and climatic regions, the major patterns of urban morphology are effectively captured by the models. However, in relatively isolated areas where local zoning or morphological characteristics differ substantially from the regional distribution of reference samples, predicted values should be interpreted with caution.
> …"

References:

[1] Wu, W. B., Ma, J., Banzhaf, E., Meadows, M. E., Yu, Z. W., Guo, F. X., Sengupta, D., Cai, X. X., and Zhao, B.: 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, 2023

[2] Chen, Y. and Zhao, S.: A building height dataset across China in 2017 estimated by the spatially-informed approach, Scientific Data, 9.1, 76, https://doi.org/10.1038/s41597-022-01192-x, 2022

[3] Meyer, H., Reudenbach, C., Wöllauer, S. and Nauss, T.: Importance of spatial predictor variable selection in machine learning applications–Moving from data reproduction to spatial prediction, Ecological Modelling, 411,108815, https://doi.org/10.1016/j.ecolmodel.2019.108815, 2019

[4] Yang, G., Xuan, Y. and Zhou Z.: Influence of building density on outdoor thermal environment of residential area in cities with different climatic zones in China—Taking Guangzhou, Wuhan, Beijing, and Harbin as examples, Buildings, 12.3, 370, https://doi.org/10.3390/buildings12030370, 2022

[5] Wang, Y., Sun, G., Wu, Y. and Rosenberg, M. W.: Urban 3D building morphology and energy consumption: empirical evidence from 53 cities in China, Scientific Reports, 14.1, 12887, https://doi.org/10.1038/s41598-024-63698-1, 2024
[6] Zhang, L. and Weng, Q.: Annual dynamics of impervious surface in the Pearl River Delta, China, from 1988 to 2013, using time series Landsat imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 113, 86-96, https://doi.org/10.1016/j.isprsjprs.2016.01.003, 2016
[7] 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, https://doi.org/10.1016/j.rse.2020.111859, 2020

6. Several numerical values should be checked for consistency. The Abstract reports an RMSE range of 5.96–6.69 m, whereas Page 13, Lines 283–284 report 5.94–6.69 m, and Page 29, Lines 490–491 again report 5.96–6.69 m. Similarly, Page 27, Line 460 reports the 2019 national building volume as 884.69 km³, while Page 27, Lines 463–464 report 884.49 km³. I suggest that the authors harmonize these values throughout the manuscript to avoid giving the impression of inconsistent data versions or calculation procedures.

We apologize for these typographical inconsistencies. These discrepancies are remnants of earlier dataset versions and intermediate model runs. The entire manuscript is thoroughly audited, and all numerical statistics are harmonized:

- The national RMSE range is strictly standardized to 3.20-4.28 m in all sections of the manuscript. This value range is changed because new models are trained based on the adopted new sampling method, resulting in updated RMSE values.
- The 2019 national building volume is strictly standardized to 882.43 km³ across all pages.

All other figures, tables, and supplementary materials are double-checked to ensure complete numerical consistency throughout the revised manuscript.