Manuscript ID: ESSD-2025-327.R0

Title: GlobalBuildingAtlas: An Open Global and Complete Dataset of Building Polygons, Heights and LoD1 3D Models

Comments & Suggestions by REVIEWER#1
and our responses to them (shaded):

ESSD-2025-327.R0 "*GlobalBuildingAtlas: An Open Global and Complete Dataset of Building Polygons, Heights and LoD1 3D Models*"

Comments by Reviewer#1:

This paper produces a global building height dataset, which represents a substantial workload, and the dataset itself holds significant value.

My key concerns are: (1) the methodological innovation of this work can be better clarified; and (2) given that this data set completely relies on commercial satellite data, whether updates can be sustained in the future and whether other researchers can replicate the results remain unclear.

Response from Authors: We sincerely thank the reviewer for acknowledging the workload and the value of our dataset, and we appreciate the concerns raised. We address the key points as follows.

1. Methodological innovation:
Our work primarily aims to provide a global, high-resolution building height dataset rather than to propose novel methodologies. To the best of our knowledge, this is the first effort to generate high-resolution, pixel-wise height maps that enable building-wise height estimation using predicted building polygons. While we employ the off-the-shelf methods for both building footprint extraction and monocular height estimation, the focus of this work is on producing a practical, usable dataset and demonstrating the feasibility of global 3D mapping at scale. The distinction is already highlighted in the manuscript, which in our opinion also fits perfectly to the scope of ESSD.

2. Sustainability and replicability of the dataset
We use PLANET imagery for two main reasons:
- Resolution and cost: PLANET provides 3-meter resolution images, which are significantly finer than freely available options such as Sentinel-2 (10 m). This resolution meets the minimum requirement for building-level reconstruction while remaining affordable compared to very high-resolution commercial satellites like Maxar (~30 cm), which are prohibitively expensive for university research projects—especially when the goal is to open-source the results. Planet thus offers the best balance: imagery of sufficient resolution for individual building reconstruction, global availability, and a manageable cost.
- Global coverage and revisit frequency: PLANET's constellation ensures frequent revisits and comprehensive global coverage, making it feasible to update the dataset efficiently in the future at affordable cost.

Moreover, PLANET provides free access to imagery through its education and research programs for students and faculty at accredited universities (https://www.planet.com/industries/education-and-research/), which further supports the dataset's sustainability.

We acknowledge that reliance on commercial satellite imagery may limit direct replication by some researchers. To mitigate this, we release the full code of our pipeline under an MIT license alongside the dataset. This allows researchers to adapt the pipeline to alternative satellite imagery sources, thereby ensuring reproducibility and the continued usability of the methodology.

Page 4, Line 101: You mention that current raster-scale building height data often suffer from low resolution and poor quality. However, state-of-the-art raster-scale building height data can already achieve resolutions of 2.5 meters (e.g., Cao et al. A deep learning-based super-resolution method for building height estimation at 2.5 m spatial resolution in the Northern Hemisphere), and many of them, after contour optimization, rival instance-level products in structural detail. Compared to these studies, where does the advantage of your dataset lie?

Response from Authors: We thank the reviewer for raising this important point. We are aware of the recent work by Cao & Weng (2024), which indeed represents an impressive advance in raster-scale building height estimation. Our dataset differs in three key aspects:
1. Pixel-wise coverage: GBA.Height provides height estimation for all pixels, including both building and non-building areas. In contrast, Cao & Weng focus only on building pixels.
2. Geographic coverage: GBA offers truly global coverage, while Cao & Weng cover only three continental regions (Europe, North America, and China).
3. Performance: In overlapping test areas, we achieve both higher completeness and lower building-wise RMSE.

Product	Completeness	Building-wise RMSE [m]
Cao & Weng, 2024	0.55	4.17
GBA.LoD1 (ours)	0.99	3.80

Thus, while both efforts are valuable, our dataset provides more comprehensive pixel-level height information, global-scale coverage, and improved accuracy in building height estimates.

Page 6, Line 130: What type of LiDAR did you use—airborne or spaceborne? Which countries are covered? This information is essential, even if it appears in supplementary materials. Furthermore, Figure 1 shows no 3D labels for Africa or South America. Did you rely on training from other continents and generalize to these area? How was the accuracy of this extrapolation validated?

Response from Authors: We thank the reviewer for this question.

1. LiDAR data
We used airborne LiDAR data, and it covers 168 city-scale regions around the globe. We revised Sect. 3.3 and added Appendix A to provide more details.

3.3 LiDAR

As the most precise 3D measurement available, we used aerial LiDAR data covering a wide range of geolocations (see Fig. 1) in developed countries. A comprehensive list of the covered countries is provided in Appendix A. These data are publicly released by governments for open use. Due to the high operational cost of LiDAR observations, no such data are available in Africa. LiDAR observations were primarily acquired in 2019 to ensure temporal consistency with the PSR data. When 2019 acquisitions were not available, data from adjacent years were utilized. Since the coverage is concentrated in developed regions, where building changes are generally limited, the temporal discrepancies between the PSR data and the 3D labels are expected to be negligible. We processed the LiDAR data into normalized digital surface models (nDSMs), which represent the heights of ground objects. These nDSMs served as reference data for training our neural networks.

Appendix A: Additional Information of the 3D Training Set

The 3D training set introduced in Sect. 3.3 comprises a total of 168 cities (Table A1). The majority of samples originate from Europe (109 cities), followed by North America (39) and Oceania (17). In contrast, Asia and South America contribute only two and one cities, respectively, while no publicly available LiDAR data were identified for Africa that could be used to supervise monocular height estimation models.

In total, the training set contains 187,239 paired samples of PSR imagery and corresponding nDSM patches, each with a spatial size of 256 × 256 pixels at 3 m resolution, corresponding to a ground coverage of approximately 110,438 km^2.

Table A1. List of Countries with Numbers of City-scale Regions in the 3D Training Set

Europe			**109**	**North America**	**39**	**Asia**	**2**
Austria	5	Norway	8	Canada	19	China	1
Belgium	8	Poland	1	Mexico	1	Philippines	1
Denmark	5	Ireland	1	United States	19		
Estonia	2	Slovakia	1			**South America**	**1**
France	3	Slovenia	2	**Oceania**	**17**	Brazil	1
Germany	25	Spain	18	Australia	11		
Latvia	2	Switzerland	3	New Zealand	6	**Africa**	**0**
Luxembourg	1	United Kingdom	17				
Netherlands	7			**TOTAL**		**168**	

2. Generalization to Africa and South America
We acknowledge the absence of 3D labels in Africa as a current limitation (Sect. 5.4). In this region, model performance necessarily relies on extrapolation from training in other continents. This limitation, however, highlights an important opportunity: since no suitable open-access LiDAR datasets currently exist for Africa, our dataset can serve as a valuable starting point to stimulate further validation efforts. We are actively exploring options, including potential acquisition of commercial datasets, to support future validation and refinement in Africa.

For South America, although only a small number of 3D labels are available, they were included for both training and validation. As shown in Table 3, the results in this region a re acceptable, suggesting that the model can generalize reasonably well even with limited local supervision.

Section 3.1: PSR is a commercial satellite, can other researchers replicate this study at low cost or update the results in the future?

Response from Authors: We acknowledge the reviewer's concern regarding the use of Planet Scope (PSR) imagery, which is commercial. However, reproducibility and sustainability are ensured in several ways - sorry for partially repeating our arguments that are already mentioned before for the sake of comprehensiveness: First, Planet provides free access for students and faculty through its Education and Research program (see https://www.planet.com/industries/education-and-research/), lowering the barrier for academic replication. Second, compared to very-high-resolution commercial data, Planet imagery strikes a unique balance of resolution (3 m) and affordability, which is essential for building-level reconstruction at the global scale. Third, we release the full pipeline under the MIT license with Commons Clause, enabling other researchers to apply our methodology to alternative satellite imagery sources (e.g., Sentinel-2, Landsat, or future open datasets). This ensures that our approach remains reproducible, even for those without direct access to PSR, and that updates to the dataset can be efficiently performed in the future when suitable imagery is available. In addition, we plan to update the datasets with our access to PSR data on a regular basis and make it available to the community, and related efforts are already underway.

Section 4.3.1: When constructing the training set, the access time of all labels and imagery should be emphasized.

Response from Authors: Thank you for this suggestion. We agree that the timing of label and imagery acquisition is important, and we emphasize this aspect in the revised Sect. 3.3. Specifically, the PSR imagery used in this study was acquired in 2019. We selected LiDAR observations from the same year whenever possible. When 2019 LiDAR data were unavailable, we used acquisitions from adjacent years.

Because most of our training cities are located in developed regions such as Europe and North America, where building activity is relatively modest, we expect temporal discrepancies in building heights to be negligible. We now make this point explicit in the manuscript to clarify the temporal consistency of our training set.

3.3 LiDAR

As the most precise 3D measurement available, we used aerial LiDAR data covering a wide range of geolocations (see Fig. 1) in developed countries. A comprehensive list of the covered countries is provided in Appendix A. These data are publicly released by governments for open use. Due to the high operational cost of LiDAR observations, no such data are available in Africa. LiDAR observations were primarily acquired in 2019 to ensure temporal consistency with the PSR data. When 2019 acquisitions were not available, data from adjacent years were utilized. Since the coverage is concentrated in developed regions, where building changes are generally limited, the temporal discrepancies between the PSR data and the 3D labels are expected to be negligible. We processed the LiDAR data into normalized digital surface models (nDSMs), which represent the heights of ground objects. These nDSMs served as reference data for training our neural networks.

Page 6, Line 130: You mention adding an extra FCN head. From which layer of UperNet is this FCN head connected—PPM block or fused layer? Were ablation studies conducted to demonstrate accuracy improvements? Otherwise, the addition of an FCN head seems arbitrary, especially since the pyramid structure already captures deep semantic information. Furthermore, if an additional head for supervision is necessary, FCN is an overly simplistic choice.

Response from Authors: Thank you for pointing out this issue. We made a mistake in describing the network architecture in the previous manuscript. In fact, we employ an auxiliary loss similar to that used in PSPNet. Specifically, the feature map from the third layer of ConvNeXt is passed through a convolutional layer and supervised with a cross-entropy loss against the ground truth. This auxiliary branch is only used during training, and its prediction is discarded during inference. We have revised the paper to correct this description.

4.3.2 Building Map Extraction

A building map extraction network was designed to map the input PSR images to binary building masks. This followed an encoder–decoder architecture based on UPerNet (Xiao et al. (2018)), with ConvNeXt-Tiny (Liu et al. (2022)) serving as the backbone. The network is supervised by cross-entropy loss. Following the practice in Zhao et al. (2017), an auxiliary loss is calculated using the third-layer features extracted from ConvNext backbone to enhance the feature representation. Before being fed into the network, image patches were upscaled by a factor of 4 and cropped to a size of 512×512. The network was trained for 160,000 iterations with a batch size of 8, using the AdamW optimizer (Loshchilov and Hutter (2017)). The global inference was built-up on the pipeline for generating GlobalBuildingMap, as described in Zhu et al. (2024).

Figure2 Why were two separate models chosen for height and contour estimation? Numerous studies have shown that joint training is more efficient and improves accuracy.

Response from Authors: Thank you for this insightful question. We agree that joint training can often improve efficiency and accuracy. In our case, however, we chose separate models for two main reasons:

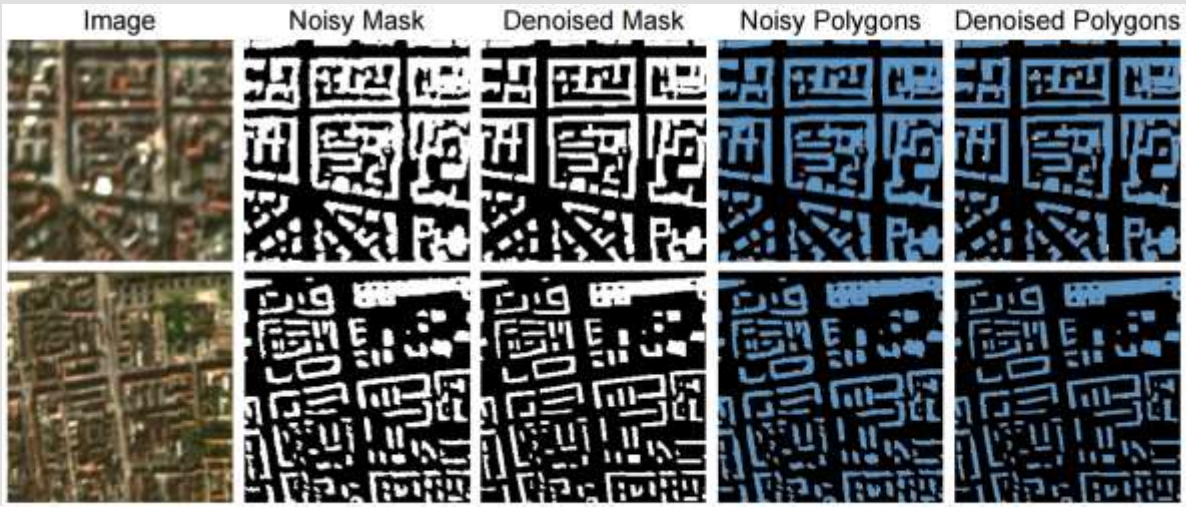
1. Label availability: The coverage of labels for 2D building footprints and 3D building heights is not identical. A joint training strategy would restrict the footprint model to a smaller subset of areas where both labels coexist, limiting its generalization ability.
2. Engineering considerations: Maintaining separate models allows for easier verification, parallel development, and flexible updates (e.g., updating building polygons or heights independently), which is advantageous for large-scale production.

Thus, while joint training is valuable in other contexts, our separation strategy better fits the practical requirements and data constraints of this global-scale project.

Page 8, Line 180: You trained the model using labels with randomly added noise, but the patterns of this artificial noise may differ significantly from the actual noise introduced by the model itself. How effective is the resulting denoising model in practice? Furthermore, since there are already su perior denoising models available—such as those based on adversarial learning or diffusion models—wouldn't these approaches be more suitable alternatives?

Response from Authors: Yes, we agree that the distribution of the artificially added noise may differ from the real noise, and that more advanced approaches such as GAN-based or diffusion-based models could potentially enhance the regularity and fidelity of the generated building polygons. In fact, our preliminary experiments also confirm their effectiveness in addressing this issue. However, since our work focuses on processing building footprints at a global scale, simplicity and computational efficiency are of primary importance. This is the main reason why we adopted a relatively lightweight yet effective denoising pipeline.

In future work, we plan to explore more sophisticated methods for building instance extraction and shape regularization, particularly for building masks derived from medium-resolution satellite imagery. To give an impression of the effectiveness of our current strategy, we provide some visualization examples below, from which one can already observe that the regularity of building masks has been significantly improved.



Page 9, Line 205: According to Figure 1, 3d samples from mainland China appear extremely limited (only three cities). Moreover, the data source is not specified. To my knowledge, building height labels in China typically only include floor counts, not precise meter-level measurements. What specific processing was applied?

Response from Authors: Thank you for pointing this out. You are correct that building height labels in mainland China generally provide only floor counts rather than precise meter-level measurements. For this reason, we did not include any mainland Chinese cities in our 3D training set.

The cities shown in Figure 1 (e.g., Harbin, Shanghai, and cities in Tibet) were used only as part of the 2D training set. The only city in China included in both the 2D and 3D training sets is Hong Kong, where accurate 3D building height data are publicly available through the Hong Kong government portal (https://3d.map.gov.hk/).

Page 10, Line 251: Additional buildings in auxiliary data could easily be false positives. While you attempt to remove false positives in your own data, how do you ensure that false positives in auxiliary sources do not compromise the final results?

Response from Authors: Thank you for this comment. In our own model outputs, we observed that false positives occur primarily in non-urban areas, which is why we applied explicit filtering to reduce them.

For auxiliary building footprint sources, the situation is different: these products are generally derived from higher-resolution or higher-quality data and, in our observations, contain far fewer false positives in non-urban regions. In urban areas, we employ a quality-guided footprint fusion strategy, which controls the contributions of different sources according to their reliability. This process minimizes the inclusion of false positives from auxiliary datasets and ensures that only the most consistent and credible footprints are retained in the final product.

Page 11, Line 277: Asia has the largest number of buildings, yet according to the paper and Figure 1, it has the fewest 2D and 3D labels. Could this affect model performance in Asia?

Response from Authors: Thank you for this important question. It is true that Asia currently has the fewest 2D and 3D labels in our training set, and this may affect model performance in some regions. Nevertheless, our evaluation in Japan—where accurate 3D reference data are available—shows that the results remain acceptable for practical applications.

We fully agree that expanding the availability of high-quality 2D and 3D labels in Asia would further improve performance. However, such data are not yet widely accessible. We see this as an important direction for future work and plan to incorporate additional datasets as they become available, in order to strengthen model generalization across Asia.

For the present dataset, we have exhausted all open labels currently available and are strived to achieve the best possible performance under the current label data constraints. We hope the reviewer will agree that, despite these limitations, the dataset already provides substantial value to the community..

Table 3: Instance-level RMSE should theoretically be much lower than 3-meter raster resolution, yet anomalies appear in Asia, Africa, and South America. Please explain. The RMSE for building height estimation in Oceania is reported as merely 1.5 meters. However, existing studies suggest that current building label data itself contains inherent inaccuracies. If this 1.5-meter error is potentially smaller than the intrinsic error of the reference labels themselves, the validity and meaningfulness of such accuracy evaluation become scientifically questionable.

Response from Authors: Thank you for this thoughtful comment.

1. On apparent “anomalies” across continents:

The differences you observed are not anomalies but rather reflect the evaluation protocols used for raster versus vector products. For raster products (e.g., GBA.Height), evaluations are conducted against ground-truth building polygons; in other words, the reference instances are ground-truth buildings. For vector products (e.g., GBA.LoD1), evaluations are performed against the building instances defined in the vector product itself. These two protocols are designed to allow fair comparison within each product category (raster vs. vector), but they are not intended for cross-comparison between categories.

If one does attempt a cross-comparison, it is important to note that completeness differs across products. This means the RMSE values are calculated on different sets of buildings, and thus cannot directly reflect the relative advantages or disadvantages of raster versus vector representations.

2. On the 1.5 m RMSE in Oceania

The validation in Oceanis, namely two cities, Launceston (https://nre.tas.gov.au/land-tasmania/aerial-photography/elevation-data) and Geelong (https://www.geelongdataexchange.com.au/pages/digital-twin-v2/), relies on airborne LiDAR datasets, which are maintained to have a vertical accuracy of approximately ±0.2 m. This is substantially smaller than the 1.5 m RMSE reported in our results, indicating that the evaluation is reliable and that the observed error reflects model performance rather than inaccuracies in the labels.

Comments & Suggestions by REVIEWER#2
and our responses to them (shaded):

ESSD-2025-327.R0 “*GlobalBuildingAtlas: An Open Global and Complete Dataset of Building Polygons, Heights and LoD1 3D Models*”

Comments by Reviewer#1:

The study provided high-quality, consistent, and global building data in 2D and 3D form, which are helpful to the urban management and Sustainable Development Goal. The proposed method is innovative and dataset is of high accuracy. However, there are still some problems that deserve to solve before publications.

Response from Authors: Many thanks for reviewer’s acknowledgement of our contribution. We answer the concerns as follows.

(1) In the related work part, the author mentioned more details about data and methods used for building footprint and height estimation, including SAR/InSAR data, deep learning methods.

Response from Authors: Thank you for your comment. In the related works section, our main focus is to introduce existing large-scale building footprint and building height products. While we acknowledge that some studies have leveraged SAR/InSAR and deep learning methods to derive building attributes, these works are often limited to a few specific regions. Including all of them may dilute the focus of our discussion on large-scale building height products.

(2) More details about the PSR data should be introduced in section 3.1, including acquisition time, data quantity, etc. The detailed description of LiDAR data used in this study is lacking. Is it the satellite-based data or the ground-based data?

Response from Authors: In our previous manuscript, the detailed information on the PSR data was provided in Section 4.2. The purpose of Section 3.1 is to give a brief overview of all the data sources used in this study. Since the acquisition and preprocessing of PSR data are part of the overall pipeline illustrated in Figure 2, we prefer to keep the detailed description in Section 4.2.

Additionally, we refine Sect. 3.3 to provide more details of the LiDAR data.

3.3 LiDAR

As the most precise 3D measurement available, we used aerial LiDAR data covering a wide range of geolocations (see Fig. 1) in developed countries. A comprehensive list of the covered countries is provided in Appendix A. These data are publicly released by governments for open use. Due to the high operational cost of LiDAR observations, no such data are available in Africa. LiDAR observations were primarily acquired in 2019 to ensure temporal consistency with the PSR data. When 2019 acquisitions were not available, data from adjacent years were utilized. Since the coverage is concentrated in developed regions, where building changes are generally limited, the temporal discrepancies between the PSR data and the 3D labels are expected to be negligible. We processed the LiDAR data into normalized digital surface models (nDSMs), which represent the heights of ground objects. These nDSMs served as reference data for training our neural networks.

More information that could be interesting are detailed in the new Appendix A.

Appendix A: Additional Information of the 3D Training Set

The 3D training set introduced in Sect. 3.3 comprises a total of 168 cities (Table A1). The majority of samples originate from Europe (109 cities), followed by North America (39) and Oceania (17). In contrast, Asia and South America contribute only two and one cities, respectively, while no publicly available LiDAR data were identified for Africa that could be used to supervise monocular height estimation models.

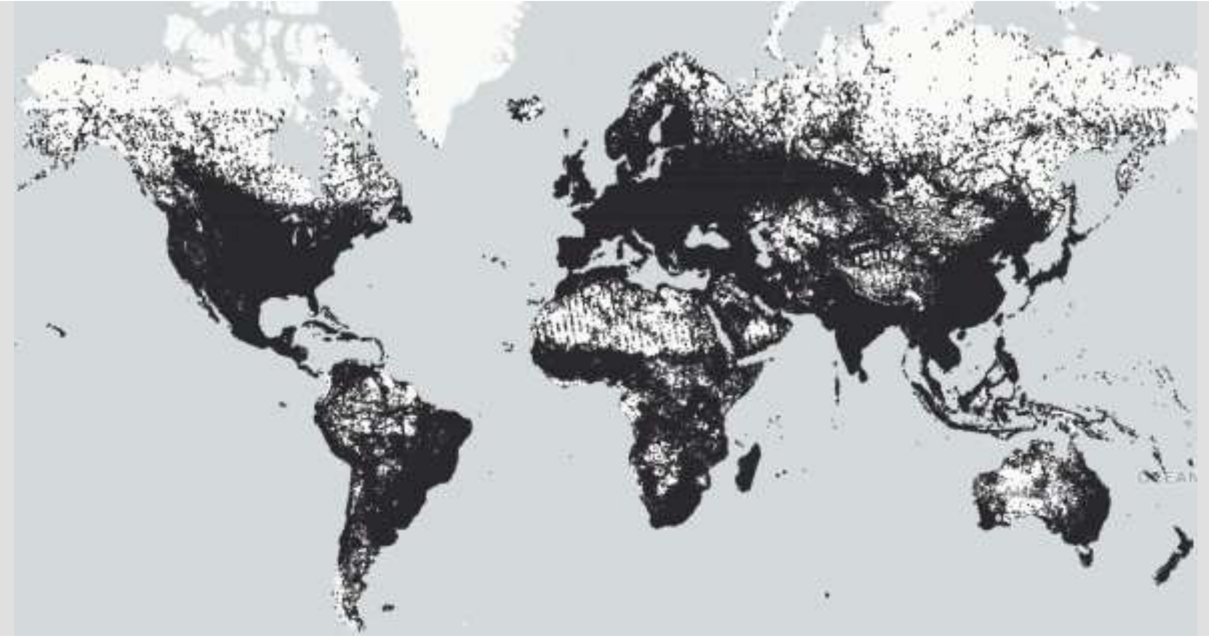
In total, the training set contains 187,239 paired samples of PSR imagery and corresponding nDSM patches, each with a spatial size of 256 × 256 pixels at 3 m resolution, corresponding to a ground coverage of approximately 110,438 km^2.

Table A1. List of Countries with Numbers of City-scale Regions in the 3D Training Set

Europe			**109**	**North America**	**39**	**Asia**	**2**
Austria	5	Norway	8	Canada	19	China	1
Belgium	8	Poland	1	Mexico	1	Philippines	1
Denmark	5	Ireland	1	United States	19		
Estonia	2	Slovakia	1			**South America**	**1**
France	3	Slovenia	2	**Oceania**	**17**	Brazil	1
Germany	25	Spain	18	Australia	11		
Latvia	2	Switzerland	3	New Zealand	6	**Africa**	**0**
Luxembourg	1	United Kingdom	17				
Netherlands	7			**TOTAL**		**168**	

(3) To reduce the calculation amount, the author could extract areas with building and remove areas without building in advance.

Response from Authors: Yes, to reduce computational cost, we already apply a region-of-interest filtering strategy as a preprocessing step. As described in Section 4.2, inference is performed only on areas identified as urban regions according to the Global Urban Footprint (GUF) dataset. This allows us to avoid unnecessary computation on non-urban regions that contain no buildings or only very few. The figure below illustrates the urban regions where the inference process was carried out. We also provided this information in our published dataset.



(4) The LiDAR data, satellite imageries and extracted polygons may have some biases. The author could introduce more about rectification and the elimination of these biases. Meanwhile, I wonder if there are LiDAR data, why should the author use the deep learning method?

Response from Authors: Thanks for the question. We answer the questions from the below two aspects.

1. We acknowledge that biases between LiDAR data, satellite imagery, and extracted polygons do exist, particularly due to potential spatial misalignments between LiDAR and PSR data. While applying a registration process to align the two data sources could be beneficial, it would also require a highly robust algorithm to handle large-scale training data, which we leave for future work. Nevertheless, given the scale of our training dataset, we argue that the impact of such misalignments between LiDAR and PSR data is likely to be minimized.

2. On the use of deep learning despite LiDAR availability:Airborne LiDAR data does not offer global coverage. In our work, we used airborne LiDAR data covering 168 city-scale regions around the globe – that's all open data we could find. Since airborne LiDAR offers extremely high height accuracy, we paired them with Planet satellite imagery in part to build up the training dataset to train a deep learning model in order to infer global building heights from Planet satellite imagery alone (explicitly for the areas where LiDAR data is not available) and in part to evaluate the height accuracy of satellite-imagery-derived building heights. In short, deep learning pipeline is necessary to achieve a first ever global building height estimation from satellite imagery, while LiDAR data rather solely serves as training and evaluation data in this study.

(5) The author presented the estimation error of the building height and volume. I wonder if the accuracy of building footprint in this study can be estimated.

Response from Authors: Many thanks for the reviewer's efforts in improving this paper. In Table 3, we not only reported the accuracy of our height estimation model, but also the building footprint quality. Specifically, related metrics include AP, AR, N-ratio and IoU.

(6) There are many data in the section 5.1 and the author should consider their relations. For example, the RMSE of the building height and building volume are 8.9m and 586.8m3/m2. For simple calculation, the estimation error of the building footprint is about 66m2 per 100m2.

Response from Authors: Thank you for this insightful comment. We agree that relating the different error metrics can help interpretation. However, the ratio of building volume error to building height error does not directly correspond to the footprint estimation error, because building volume depends on both the footprint area and the height.

Formally, the building volume error can be expressed as $\Delta V = \hat{V} - V$, where $\hat{V}$ and V are predicted building volumes and ground truth building volumes, respectively. It can be written as $\Delta V = \hat{V} - V = \hat{S}\hat{h} - Sh$, where S and h denote ground truth building area and height, while the notations with heads denote the estimated parameters. Then

$$\frac{\Delta V}{\Delta h} = \frac{\hat{V}-V}{\hat{h}-h} = \frac{\hat{S}\hat{h}-Sh}{\hat{h}-h} = \frac{\hat{S}\hat{h}-Sh+S\hat{h}-S\hat{h}}{\hat{h}-h} = \frac{S(\hat{h}-h)}{\hat{h}-h} + \frac{(\hat{S}-S)\hat{h}}{\hat{h}-h} = S + \hat{h}\frac{\Delta S}{\Delta h}$$

It turns out that the ratio of building volume error to building height error is the summation of ground truth building area and a factor related to building area and height error.

(7) The author should simply discuss the reason for the performance difference of the proposed method in different continents.

Response from Authors: Thank you for this comment. We have addressed this point in Sect. 5.1. In brief, the observed differences in performance across continents are primarily due to variations in building morphology, the representation of regional characteristics in the training data, and comprehensiveness and quality of validation data.

The building height and volume estimation accuracy of the GlobalBuildingAtlas dataset across continents is mainly influenced by several factors: (1) variations in building morphology, (2) the representation of regional characteristics in the training data, and (3) the comprehensiveness and quality of the validation data. The global average height RMSE of GBA.LoD1 stands

(8) The section 5.4 and section 6 could be merged in to a new discussion part to improve the readability. The high correlation with grid population data can be seen as a kind of precision verification.

Response from Authors: Thank you for this suggestion. While we appreciate the idea of merging the sections, Sect. 5.4 focuses on the strengths and limitations of the dataset, whereas Sect. 6 presents real-world application showcases. To maintain clarity and proper emphasis, we have renamed Sect. 5.4 to "*Strengths and Limitations*" instead of merging it.

Regarding the high correlation with grid population data, we agree that the two are correlated. However, as shown in Fig. 7, even in Europe—where development is relatively uniform—the building volume per capita exhibits substantial variation. Therefore, while population data provide some indication, they cannot serve as a definitive measure of the precision of our building height and volume estimates.

(9) There are still some grammatical and lingual problems, and authors should make a thorough revision.

Response from Authors: Thank you for your suggestion. We have carefully revised the manuscript to address grammatical and linguistic issues, improving clarity, readability, and overall presentation.