Responses to comments from anonymous referee #2

This manuscript presents AGPC, an annual 500 m gridded population dataset for mainland China covering 1990–2020. By incorporating time-varying 3D building information together with building functional attributes, the dataset provides an improved representation of population distribution in vertically developed urban environments. The long temporal coverage and nationwide consistency make the dataset valuable for studies of urbanization, environmental change, and human–environment interactions. Overall, the topic is interesting, and the manuscript is well organized, and the dataset represents a useful contribution to the ESSD community. However, several minor comments should be addressed before reconsideration for publication. These comments are attached below.

Response:

We appreciate the reviewer's careful evaluation of our manuscript and recognition of the value of AGPC. The comments have helped us identify several aspects that required further clarification and improvement, particularly regarding the model training and temporal transfer strategy, the geographical representativeness of the validation data, and the comparison with existing population datasets. In response, we have clarified the methodological workflow, substantially expanded the township-level validation samples, added representative spatial comparisons with existing datasets, and revised the relevant text, figures, tables, and equations throughout the manuscript. Detailed responses to each comment are provided below.

1. The Introduction mentions that previous studies have rarely incorporated building height or volume information into gridded population mapping. However, several recent studies closely related to this topic are not discussed, for example, Lei et al. (2024, doi.org/10.3390/ijgi13090335) and Tu et al. (2026, doi.org/10.3389/feart.2026.1748599). Although these studies focus on either a single year or limited regions, citing and briefly discussing them in the Introduction would provide a more comprehensive review of the current state of the field and better position AGPC relative to existing work.

Response:

Thank you for this helpful suggestion. We agree that recent advances in incorporating detailed building information into gridded population mapping should be more comprehensively discussed. Following the reviewer's suggestion, we have expanded the relevant literature review in the Introduction by including Lei et al. (2024) and Tu et al. (2026), together with two additional closely related studies (Chen et al., 2024; Yan et al., 2025).

Based on this expanded review, we have also reorganized and refined the research gaps. Rather than suggesting that 3D building information has rarely been incorporated, we now acknowledge recent progress and more specifically highlight the remaining limitations in the availability of large-scale 3D-based population datasets, the representation of temporal changes in 3D urban structure, and the integration of detailed building functional information. Accordingly, the positioning of AGPC has been revised to emphasize its use of multi-temporal 3D building characteristics and building functional information for long-term population reconstruction. These revisions provide a more comprehensive review of recent developments and more clearly distinguish AGPC from existing studies. The corresponding revisions in the Introduction are as follows:

(In Lines 93–102)

"…Recent studies have begun to address this limitation by incorporating 3D building information into population mapping. At the national scale, Chen et al. (2024) and Lei et al. (2024) incorporated building height and building volume, respectively, together with other geospatial covariates to produce 100 m gridded population estimates for mainland China. At the urban scale, Yan et al. (2025) and Tu et al. (2026) introduced building height information for fine-scale population estimation in selected Chinese cities, with Tu et al. (2026) further distinguishing residential and non-residential buildings. These studies demonstrate the potential of 3D building information to improve population mapping beyond conventional 2D indicators. Nevertheless, the incorporation of 3D building information into gridded population mapping is still at a relatively early stage, with only a limited number of studies conducted at large spatial scales. Moreover, publicly available and systematically validated gridded population datasets explicitly incorporating 3D building information remain scarce, limiting their broader application and evaluation."

(In Lines 113–118)

"…This temporal limitation is even more pronounced for 3D building information, as the few existing 3D-based population mapping studies have generally relied on static 3D building data for individual reference years. Consequently, key processes associated with urban development, including horizontal expansion, increasing building density, and vertical intensification, remain insufficiently represented in long-term population reconstruction, making it difficult to characterize corresponding changes in population-carrying capacity over time. Temporally consistent 3D built-environment data therefore remain critical for improving long-term population mapping."

(In Lines 131–137)

"…Although recent studies have begun to distinguish residential and non-residential buildings (e.g., Tu et al., 2026), the explicit integration of spatially detailed building functions with 3D building structure remains limited, particularly for nationwide population mapping. Explicitly incorporating building functional types can therefore provide complementary information beyond POI-based indicators and better account for the heterogeneous effects of different urban functions on population allocation."

Studies cited in this response:

"Chen, Y., Xu, C., Ge, Y., Zhang, X., and Zhou, Y.: A 100 m gridded population dataset of China's seventh census using ensemble learning and big geospatial data, Earth System Science Data, 16, 3705-3718, https://doi.org/10.5194/essd-16-3705-2024, 2024.

Lei, Z., Zhou, S., Cheng, P., and Xie, Y.: Improved population mapping for china using the 3d building, nighttime light, points-of-interest, and land use/cover data within a multiscale geographically weighted regression model, ISPRS International Journal of Geo-Information, 13, 335, https://doi.org/10.3390/ijgi13090335, 2024.

Tu, M., Ji, Q., and Sun, Y.: Enhancing the accuracy of urban population spatialization at 100-m grid scale through differentiating residential and non-residential buildings, Frontiers in Earth Science, 14, 1748599, https://doi.org/10.3389/feart.2026.1748599, 2026.

Yan, Z., Ma, L., Wang, X., Kim, Y., and Zhang, L.: High-precision population estimates by remote

sensing big data and advanced transformer deep learning model, Remote Sensing Applications: Society and Environment, 39, 101638, https://doi.org/10.1016/j.rsase.2025.101638, 2025."

2. The description of the model training and validation procedure is currently not sufficiently clear and may lead to confusion. Lines 133–137 state that five-fold cross-validation was performed and that "in each fold, the data were randomly split into 70% for model training and 30% for validation." In contrast, Lines 410–415 describe a fixed 70/30 training–validation split, followed by five-fold GridSearchCV performed on the training subset. Please clarify the exact workflow and revise the corresponding description to avoid misunderstanding.

Response:

Thank you for pointing out this ambiguity. We apologize for the inconsistent description of the model training and validation procedure in the original manuscript. The exact workflow was as follows. First, the full dataset was randomly split into 70% for model training and hyperparameter tuning and 30% as an independent validation set. The 30% validation subset was held out and was not involved in model training or hyperparameter selection. Within the 70% training subset, five-fold cross-validation was implemented using GridSearchCV for hyperparameter tuning. In each cross-validation iteration, four folds (80% of the training subset) were used for model fitting and the remaining fold (20%) for internal validation. After the optimal hyperparameters were identified, the model was refitted using the entire 70% training subset with the selected hyperparameters and subsequently evaluated using the independent 30% validation subset.

We have revised the corresponding descriptions in the manuscript to clearly distinguish the five-fold cross-validation used for hyperparameter tuning from the independent validation used for final model evaluation, as follows:

(In section 2.1, Line 180–184)

"*...For model development and evaluation, the county-level census data were randomly divided into a 70 % training subset and a 30 % independent validation subset. The training subset was used for model fitting and hyperparameter optimization, while the validation subset was held out from the entire training and tuning process and used exclusively for the final evaluation of model performance.*"

(In section 4.1.1, Lines 499–506)

"*...To systematically optimize model performance and reduce the risk of overfitting, hyperparameter tuning was performed on the 70% training subset using GridSearchCV with five-fold cross-validation. In each iteration, four folds (80 % of the training subset) were used for model fitting and the remaining fold (20 %) for internal validation. Based on the cross-validation results, the optimal hyperparameter configuration consisted of 120 trees, a learning rate of 0.05, a maximum tree depth of 5, and a subsample ratio of 0.7, representing a balance between model complexity and generalization capability. The model was then refitted using the entire 70 % training subset with the selected hyperparameters and subsequently evaluated on the independent 30 % validation subset, which had not been involved in either model fitting or hyperparameter tuning.*"

3. The manuscript states that separate baseline models were developed for 2010 and 2020, and then subsequently extended to other years. However, it is unclear how these models were

applied to each five-year time point. Please specify which baseline model was used for each historical year (1990, 1995, 2000, 2005, 2015, etc.), or explain the criteria used to determine the temporal transfer strategy.

Response:

We agree that, although the temporal transfer strategy was illustrated in the workflow figure (Fig. 3), it was not sufficiently explained in the main text. Specifically, the 2010 baseline model was used to estimate population allocation weights for all historical years from 1990 to 2015, whereas the independently trained 2020 model was applied only to 2020. The primary consideration for this strategy was that the 2010 model provides a more appropriate baseline for reconstructing the historical period than the terminal-year 2020 model, thereby reducing the potential influence of population–environment relationships specific to more recent urban development. In addition, using a common baseline model for the historical years helps maintain methodological consistency across the reconstructed time series and avoids potential discontinuities caused by switching between independently trained baseline models.

It should also be clarified that transferring the 2010 model does not mean directly reproducing the 2010 population distribution in other years. The model was used to estimate relative population allocation weights based on the predictors corresponding to each historical year. Time-varying predictors, particularly dynamic 3D building information, were updated for each time point, and the resulting spatial weights were further constrained by year-specific census/statistical population totals.

We have revised the relevant text to explicitly describe this temporal transfer strategy and its rationale, as follows:

(In section 3.5, Lines 424–433)

"*…Under this assumption, the trained XGBoost model serves as a transferable spatiotemporal predictor. Specifically, the 2010 baseline model was applied to the five-year time points from 1990 to 2015 (i.e., 1990, 1995, 2000, 2005, 2010, and 2015) to estimate population distribution weights, whereas the independently trained 2020 model was used for 2020. The 2010 model was selected as the common baseline for historical reconstruction because it provides a more appropriate reference for the pre-2020 period than the terminal-year 2020 model, thereby reducing the potential influence of population–environment relationships specific to more recent urban development. Using a common baseline model also helps maintain methodological consistency across the historical time series. Multi-temporal population distribution weights were estimated at five-year intervals using the time-varying built-up area and 3D building characteristics corresponding to each time point.*"

4. The finest-level validation data (township-level census records) appear to represent only a subset of townships rather than complete national coverage. As shown in Fig. 1, the spatial distribution of these samples is uneven, with some regions being sparsely represented. If additional township-level census records are available, it would be beneficial to include more validation samples from these underrepresented regions to improve the geographical representativeness of the evaluation.

Response:

We agree that broader geographical coverage of township-level validation data can improve the representativeness and robustness of the evaluation. The township-level census records were originally available mainly in tabular form without explicit spatial information and had to be manually matched to administrative boundaries. This process requires considerable manual verification because of duplicated administrative names, name changes, and boundary adjustments, which limited the number and geographical coverage of validation samples included in the original manuscript.

In response to this suggestion, we substantially expanded the township-level validation datasets for both 2010 and 2020, with particular attention to regions that were sparsely represented in the original dataset. Additional records were collected, spatially matched, and verified to ensure that all provinces across mainland China were represented. As a result, the revised datasets now contain approximately 10,000 township-level census records for each of 2010 and 2020, providing substantially improved geographical coverage compared with the original validation datasets. Accordingly, all township-level validation analyses were recalculated using the expanded datasets, and the corresponding validation statistics (Tables 4), figures (Fig.1, Fig. 4, and Fig. 7), and related descriptions have been updated throughout the revised manuscript. Fig. 1 has also been updated to show the improved spatial distribution of the validation samples. The corresponding revisions in the manuscript are as follows:

(In section 2.1, Lines 191–199)

"…Approximately 10,000 township-level census records were compiled for each of 2010 (Fig. 1-c, 10,834 records) and 2020 (Fig. 1-d, 10,336 records). As the original township-level census records were primarily available in tabular form without explicit spatial information, they were spatially matched to township administrative boundaries based on administrative names, with ambiguous records manually verified to account for duplicated names, administrative name changes, and boundary adjustments. The resulting validation datasets provide broad geographical coverage across mainland China, with all provinces represented. These township-level records, representing a finer spatial scale than the county-level training data, were used to independently evaluate the accuracy of the gridded population estimates, with a particular focus on the model's ability to capture intra-county spatial heterogeneity."

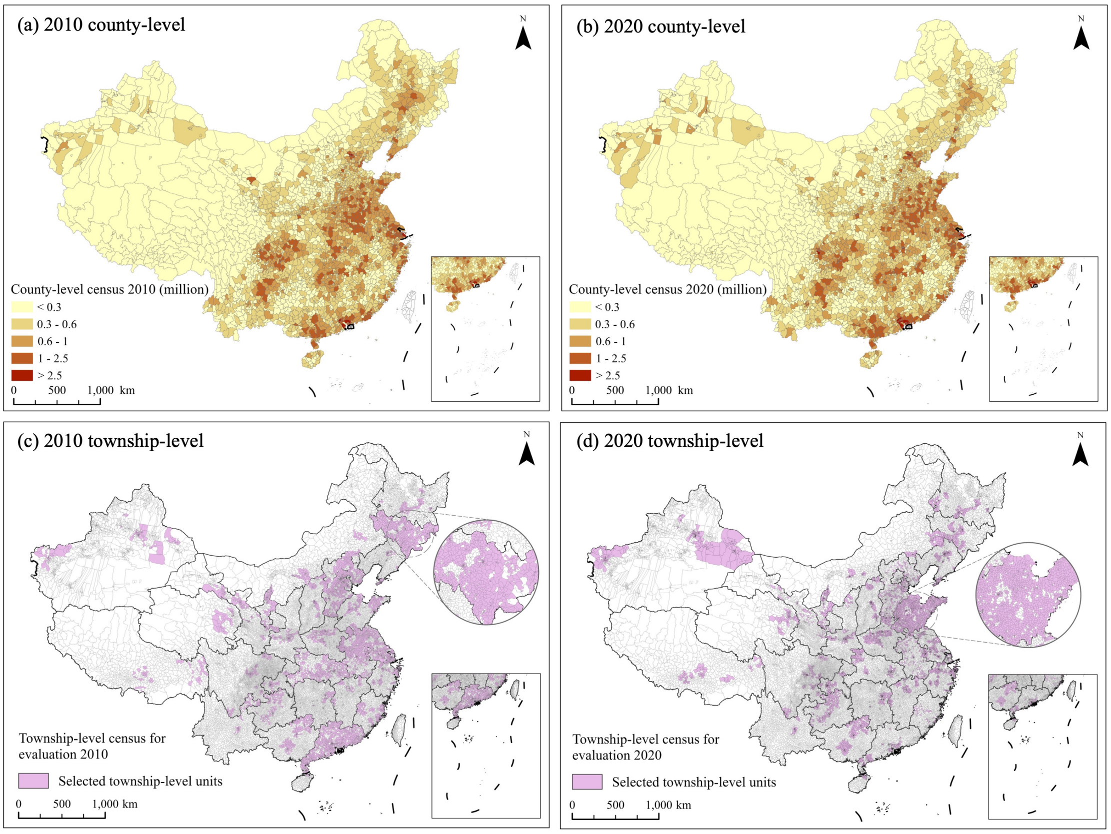

Fig. 1 Population census data from Sixth and Seventh National Population Census of China, including county-level population data for 2010 (a) and 2020 (b), and the spatial distribution of selected township-level census units for 2010 (c) and 2020 (d)

5. The comparison with existing gridded population datasets is currently limited to statistical metrics. Adding several representative map comparisons with existing datasets (e.g., WorldPop or other widely used products) would provide a more intuitive evaluation of the dataset. Visual comparisons over typical urban and rural regions would help readers better understand the spatial differences among the datasets and more clearly demonstrate the advantages of AGPC, particularly in densely built urban areas where the incorporation of dynamic 3D building information is expected to provide greater improvements.

Response:

We agree that map-based comparisons can provide a more intuitive assessment of the spatial differences among gridded population datasets. In response, we have added a spatial comparison of AGPC with WorldPop and PopSE for representative regions in Beijing, Nanjing, Wuhan, and Shenzhen in 2020 (Fig. 13). The comparison covers both densely developed urban areas and peripheral or sparsely built-up areas, revealing clear differences among the products in the spatial extent and intra-urban patterns of population allocation. In particular, AGPC shows a closer correspondence with the built environment and better differentiates population distribution associated with variations in 3D building structure and building functions. We have added the corresponding discussion to the revised manuscript to complement the statistical evaluation and provide a more intuitive assessment of the spatial characteristics of AGPC relative to existing products. The corresponding revisions in the manuscript are as follows:

(In section 4.6, Lines 852–868)

"o further examine the spatial differences among gridded population products, representative urban regions in Beijing, Nanjing, Wuhan, and Shenzhen were selected for a map-based comparison of population distributions in 2020 (Fig. 13). To ensure comparability across products with different native spatial resolutions, all datasets were aggregated to a common 500 m grid before comparison, thereby minimizing differences arising solely from spatial resolution. Existing products generally show a broader spatial extent of population allocation, with scattered population estimates extending into peripheral regions and sparsely built-up rural areas, whereas AGPC exhibits a closer correspondence between population distribution and the underlying built environment. Within densely developed urban areas, WorldPop tends to produce a smoother population surface, with smaller differentiation among areas characterized by contrasting urban forms and building characteristics. In contrast, AGPC better captures the spatial differentiation of population distribution associated with variations in 3D building structure and building functions, enabling population allocation to more closely reflect differences in population-carrying capacity. Although PopSE also incorporates building-related information, it exhibits stronger localized population concentrations. Some of these high-value clusters appear to correspond more closely to areas of intensive commercial activity than to residential population distribution, suggesting that a stronger influence of commercial activity indicators (e.g., commercial POIs) relative to residential-related information may contribute to such spatial concentration. Overall, these spatial comparisons demonstrate that incorporating 3D urban structure and functional characteristics enables AGPC to provide a more physically informed representation of intra-urban population distribution, particularly in densely developed cities where population concentration cannot be fully characterized by conventional 2D indicators."

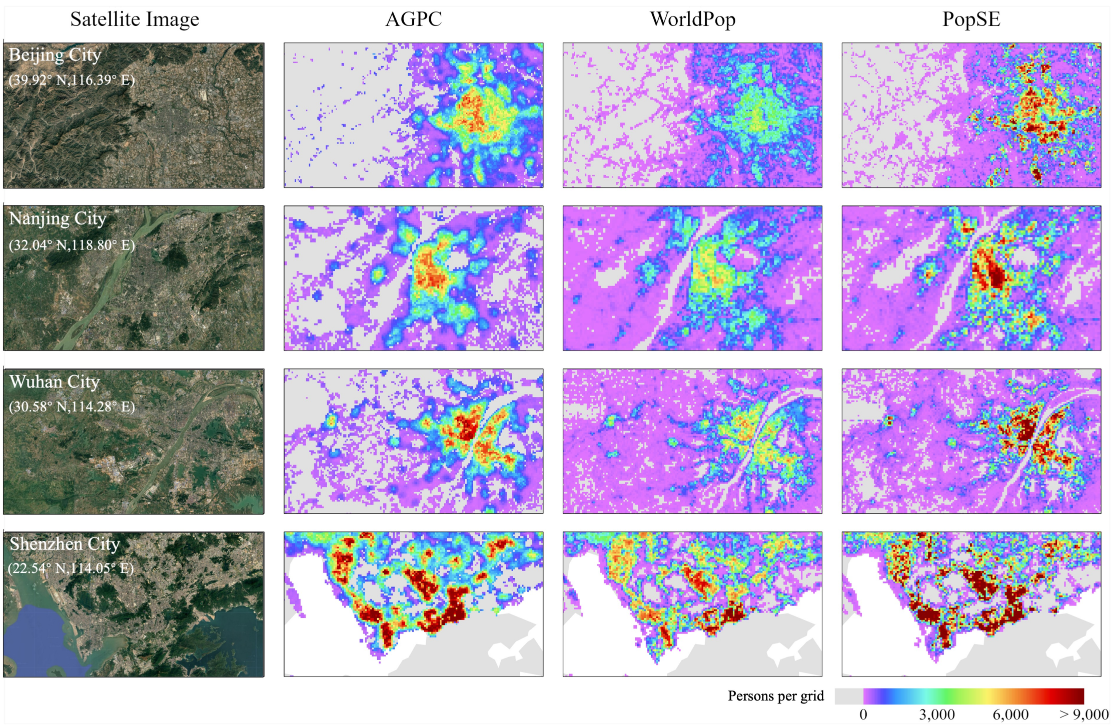

Fig. 1 Comparison of gridded population distributions in 2020 among AGPC, WorldPop, and PopSE in selected areas

6. Lines 459–460 state that "six five-year intervals between 1990 and 2020 were systematically

validated." This wording is potentially ambiguous because the study actually validates seven time points separated by six five-year intervals.

Response:

Thank you for pointing out this ambiguity. We agree that the original wording could be misinterpreted as referring to six validation time points, whereas the validation was actually conducted at seven time points separated by six five-year intervals. We have revised the sentence in the revised manuscript, as follows:

(In section 4.1.3, Lines 560–562)

"To assess the reliability of the multi-temporal population distribution results, the predicted population maps at seven time points from 1990 to 2020 (i.e., 1990, 1995, 2000, 2005, 2010, 2015, and 2020) were systematically validated."

7. The caption of Fig. 9 should be population counts per grid rather than population density.

Response:

Thank you for pointing out this inconsistency. We have corrected the caption of Fig. 9 to accurately reflect the variable displayed in the figure. In addition, Fig. 9 has been further adjusted to improve its readability, as follows:

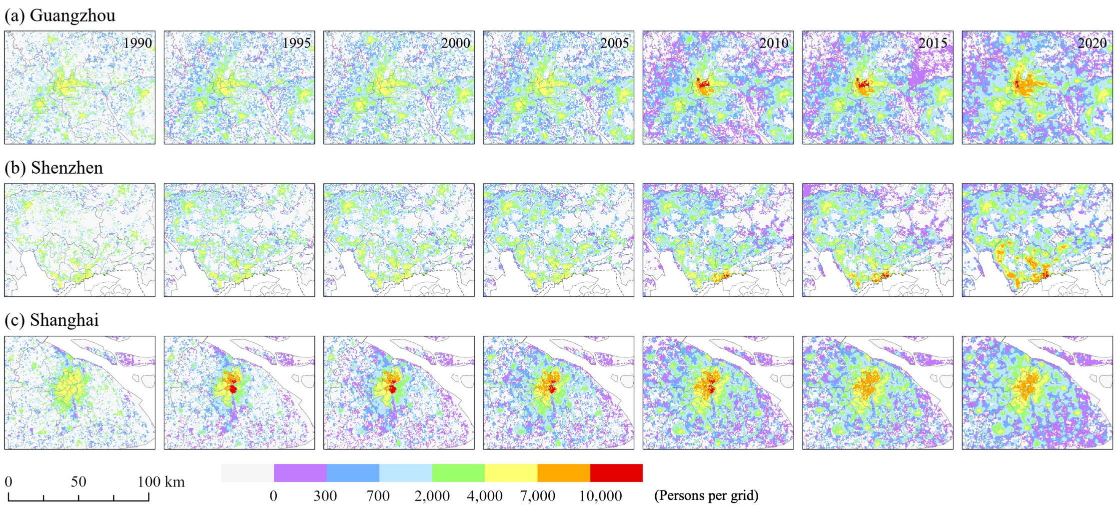

Fig. 9 Changes in the spatial distribution of population (persons per grid) in Guangzhou (a), Shenzhen (b), and Shanghai (c) during 1990–2020

8. Correct editorial inconsistencies throughout the manuscript.

- "Grided" should be corrected to "Gridded" throughout the manuscript.
- Line 445: delete "on".
- Line 514: "that" → "than".
- Line 338: remove the extra period following Section 3.5.
- The last column of Table 5 under rRMSE (%) should be labelled "Validation" instead of "Training".
- Line 660: "3D built-up coverage" should be "2D built-up coverage".

- Line 701: “conventional 3D population products” should be “conventional 2D population products”.
- In Fig. 11, “dependency” should be “dependence”.
- Eq. (8) appears to contain a notational error, where an unnecessary subscript i is included and should be removed.

Response:

Thank you for carefully identifying these editorial and notational inconsistencies. We have corrected all the issues listed above throughout the revised manuscript, including the spelling and grammatical errors, section formatting, column label in Table 5, terminology in the main text and Fig. 11, and the notation in Eq. (8). We have also carefully checked the manuscript for similar editorial inconsistencies. Please refer to the revised manuscript for the corresponding corrections.