Response to Reviewers' comments

We sincerely thank both reviewers for their careful evaluation and constructive comments. In response, we have substantially revised the manuscript, with particular attention to the methodological description, validation robustness, data transparency, interpretation of the results, and comparison with existing gridded population datasets. Relevant text, figures, tables, and equations have been revised accordingly.

Detailed point-by-point responses to Reviewer #1 and Reviewer #2 are provided below. The reviewers' original comments are presented in black, our responses are shown in blue, and all changes made to the revised manuscript are highlighted in red.

Reviewer #1

This manuscript presents AGPC, a new annual 500-m gridded population dataset for mainland China covering the period 1990–2020. A key strength of the study is that it explicitly incorporates three-dimensional building volume dynamics into the population mapping framework, thereby addressing the density-saturation limitations of conventional two-dimensional population products in vertically developed urban environments. The authors provide comprehensive validation using census and statistical data at multiple administrative levels, and the results demonstrate that AGPC achieves improved spatial representation compared with several widely used population datasets. Overall, the manuscript is well organized, and the proposed dataset will be valuable for studies of population dynamics, urbanization, disaster risk assessment, and sustainable development. I have several comments and concerns that I believe will help further improve the clarity, reproducibility, and interpretation of the manuscript.

Response:

We thank the reviewer for the positive assessment of our manuscript and for recognizing the value of the AGPC dataset. We also appreciate the constructive comments, which have helped us improve its quality and clarity. Accordingly, we have substantially expanded the township-level validation samples, clarified the availability of the GUS-3D dataset and the temporal consistency of the POI and impervious-surface data, and refined the interpretation of SHAP importance across different years. We also clarified the county-level population constraints and corrected the terminology, and editorial issues identified throughout the manuscript. These revisions have improved the robustness, reproducibility, and clarity of the manuscript. Our point-by-point responses are provided below.

1. Section 2.1 indicates that approximately 6,800 townships were used for validation in 2010, whereas only about 3,830 were available for 2020. Please explain the reason for this substantial reduction (e.g., administrative boundary changes or data availability) and discuss whether the smaller validation sample affects the robustness and comparability of the reported validation results.

Response:

We agree that the unequal sample sizes may affect the comparability of the validation results. The difference in the number of township-level validation samples between 2010 and 2020 in the original manuscript was mainly due to data availability and difficulties in spatially matching tabular census records to township boundaries. In particular, duplicated township names, administrative name changes, and boundary adjustments required substantial manual verification. In response, we substantially expanded and manually verified the township-level validation datasets, with particular attention to improving geographical coverage and maintaining comparable sample sizes between the two years. The revised township-level validation samples now contain 10,834 records for 2010 and 10,336 records for 2020, with all provinces in mainland China represented. Accordingly, all township-level validation analyses, statistics, figures, and tables have been recalculated and updated. The expanded and more balanced datasets provide a more robust and comparable basis for evaluating AGPC across the two benchmark years. This response can be found in the revised manuscript, 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."

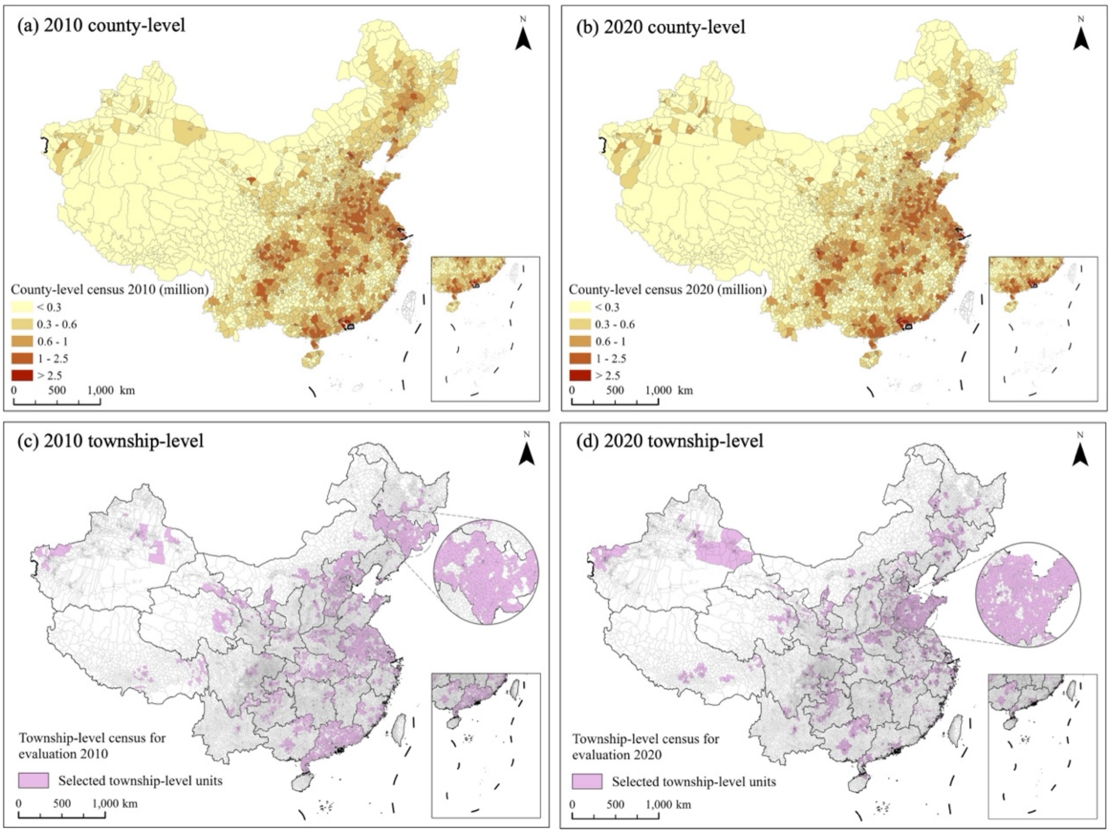

Fig. 1 Population census data from Sixth and Seventh National Population Census of China, including county-level population data for 2010 (a) and 2020 (b), and the spatial distribution of selected township-level census units for 2010 (c) and 2020 (d)

2. In Section 2.2.1 and the reference list, the GUS-3D dataset is cited as Liu et al. (2025) and Wu et al. (2026), with the latter marked as "Undergoing Review." Please clarify its current availability, public access protocol, or expected release timeline so that the reproducibility of AGPC can be properly assessed.

Response:

Thank you for raising this point regarding data availability and reproducibility. The manuscript describing the GUS-3D dataset (Wu et al., 2026) is currently under review and has therefore not yet been formally published. Nevertheless, to facilitate the reproducibility and assessment of AGPC, the GUS-3D data used in this study can currently be accessed through a private data link: https://figshare.com/s/9b7d1d40792daa9ab707. This link provides access to the dataset used for the reconstruction of the historical population distribution in AGPC. Upon formal publication of the GUS-3D manuscript and dataset, the current access information will be replaced or supplemented with the permanent public repository and corresponding DOI citation. The corresponding revisions can be found in the data description section and data availability section of the revised manuscript, as follows:

(In section 2.2.1, Lines 240–247)

"*…The 3D building variables were derived from the spatiotemporal GUS-3D dataset developed by our research team (Liu et al., 2025; Wu et al., 2026), which provides global coverage at a 500 m spatial resolution and includes building volume, building footprint ratio, and footprint-*

area–weighted mean building height for each grid cell. The dataset spans 1990–2020 at five-year intervals, enabling dynamic characterization of the spatiotemporal evolution of urban 3D structure. The manuscript describing the complete GUS-3D dataset is currently under review (Wu et al., 2026); however, the data used in this study are currently accessible through a private Figshare link provided in the Data Availability section. A permanent public repository and corresponding DOI citation will be provided upon formal publication of the GUS-3D dataset."

(In Data Availability, Lines 894–895)

"*…The GUS-3D dataset used to characterize the spatiotemporal evolution of urban 3D structure is currently accessible through a private link: https://figshare.com/s/9b7d1d40792daa9ab707*."

3. Please clarify the temporal consistency of the POI dataset. Section 2.2.2 states that the POI data were collected from Gaode Map but does not specify their acquisition year. These POIs are subsequently treated as temporally invariant throughout the study. If they represent recent urban conditions (e.g., 2022–2023), their application to earlier years such as 1990 or 1995 may introduce temporal inconsistency, as also acknowledged in Section 4.1.3. Please specify the acquisition year of the POI dataset and further discuss the potential influence of this temporal mismatch on the early-period population estimates.

Response:

Thank you for pointing out this issue. The POI dataset used in this study was collected in 2020 through the Gaode Map API, which has now been explicitly stated in Section 2 of the revised manuscript. Because comparable historical POI datasets are unavailable, the 2020 POI density surfaces were treated as temporally invariant covariates throughout the reconstruction period. We acknowledge that this assumption may introduce temporal inconsistency, particularly for earlier years and areas that experienced substantial changes in urban functions. In our framework, this potential influence is partly mitigated by the use of time-varying built-up areas and 3D building characteristics, together with year-specific population constraints. Nevertheless, the lack of temporally consistent POI data remains a limitation in reconstructing historical changes in urban functions.

In the revised manuscript, we have therefore clarified the acquisition year and temporal treatment of the POI data in Section 2.2.2, and expanded the limitation discussion in Section 4.7, including potential improvements using multi-temporal POI archives or other temporally explicit functional data, as follows:

(In section 2.2.2, Lines 252–262)

"POIs serve as spatial proxies for urban functional facilities and socioeconomic activities and can partially reflect the intensity and type of human activities within cities. POI data were collected from Gaode Map (https://amap.com) through its API in 2020, comprising more than 34 million records. To better capture major urban activity patterns, the original 23 POI categories defined by Gaode Map were aggregated into eight functional types: residential, recreation, life services, education, commercial, office, medical, and open space. Details of the reclassification and record counts are provided in Table 2. Kernel density estimation (KDE) was applied to each POI category to generate density surfaces at a 500 m spatial resolution. A

quartic kernel function (Silverman, 2018) with a uniform bandwidth of 2,000 m was used for all KDE calculations. Because comparable historical POI archives extending back to the 1990s and early 2000s are generally unavailable, the 2020 POI density surfaces were treated as temporally invariant covariates in the historical reconstruction. It should also be noted that the acquisition year of a POI record does not necessarily correspond to the establishment year of the associated facility."

(In section 4.7, Lines 872–886)

"...First, some explanatory variables, such as POI density and road network characteristics, were assumed to be temporally static due to the lack of long-term historical datasets. In particular, the POI data were obtained from a consistent 2020 dataset and applied throughout the reconstruction period, which may introduce temporal inconsistency for earlier years. Although POI density primarily represents urban functional intensity and spatial organization rather than the exact historical presence of individual facilities, the lack of temporally consistent historical POI data limits the ability to fully reconstruct past changes in urban functions. This mismatch may cause historical urban functions to be represented by their more recent spatial configuration, potentially affecting fine-scale population allocation in areas that experienced substantial functional transformation or rapid urban development. Nevertheless, historical population changes are also represented by time-varying built-up area and 3D building characteristics and constrained by year-specific statistical population totals, reducing the dependence of the reconstruction on temporally invariant POI information alone. Future studies could integrate multi-temporal POI archives, historical business registration data, or other temporally explicit functional land-use information to better capture the evolution of urban functions and further improve the temporal consistency of population reconstruction."

4. Section 3.1 states that the GAUD impervious surface dataset was combined with GUS-3D to delineate inhabited areas. However, the cited GAUD product only covers 1985–2015. Please specify how inhabited areas were derived for 2020 and clarify whether the 2015 GAUD layer was directly used as a proxy or whether an alternative dataset or updating strategy was adopted for the post-2015 period.

Response:

Thank you for pointing out this issue. The 2015 GAUD layer was not used as a proxy for 2020. Instead, the inhabited-region mask for 2020 was constructed using the corresponding 2020 impervious surface layer extracted from the 30 m Annual Global Land Cover (AGLC) dataset (Li et al., 2026), together with building information from GUS-3D. The AGLC dataset extends the temporal coverage of the impervious surface information originally provided by GAUD (Liu et al., 2020) to 2022, allowing year-specific impervious surface information to be used throughout the study period.

We have revised Section 3.1 to clarify this procedure and the relationship between the GAUD and AGLC datasets. The citation and Data Availability section have also been updated to include the AGLC dataset, which is publicly accessible through the National Tibetan Plateau Data Center (https://doi.org/10.11888/Terre.tpdc.302477). The corresponding revisions in the manuscript are as follows:

(In section 3.1, Lines 312–335):

"...Inhabited regions were primarily identified based on building presence, supplemented with impervious surface data to capture small and fragmented built-up areas. To capture the spatiotemporal dynamics of inhabited regions over the study period (1990–2020), building footprint ratio data from the GUS-3D dataset (Liu et al., 2025; Wu et al., 2026) were integrated with annual impervious surface information extracted from the 30 m Annual Global Land Cover (AGLC) dataset (Li et al., 2026). The AGLC dataset extends the temporal coverage of impervious surface information originally provided by the Global Annual Urban Dynamics (GAUD) dataset (Liu et al., 2020) to 2022, enabling year-specific impervious surface layers to be used throughout the study period. Because small and fragmented impervious features were partially filtered during the construction of GUS-3D, impervious surface fractions were recalculated at the 500 m grid level from the corresponding annual 30 m AGLC layers and merged with the GUS-3D building footprint ratios to construct a more complete inhabited-region mask for each year. In addition, fragmented settlements not captured by the AGLC dataset, such as small rural villages in mountainous areas (e.g., parts of Sichuan Province), were manually delineated using high-resolution satellite imagery from Google Maps. Building height and volume for these supplementary settlements were assigned based on visual interpretation and surrounding settlement patterns. This procedure further improves the spatial completeness and continuity of inhabited regions, particularly in sparsely populated and topographically complex areas."

(In Data Availability, Lines 932–933)

"The annual global land-cover dataset AGLC used in this study is publicly accessible through the National Tibetan Plateau Data Center at https://doi.org/10.11888/Terre.tpdc.302477."

The citation of AGLC dataset was added to the reference list accordingly:

"Li, B., Xu, X., Zhuang, H., Zhang, H., Che, Y., Zheng, Y., Cai, Y., Liu, X., and Li, X.: Annual global land cover mapping at 30 m resolution from 1985 to 2022 with high temporal consistency, Sci Bull, 71, 2726-2730, https://doi.org/10.1016/j.scib.2026.01.027, 2026."

5. Section 4.4 shows that the SHAP importance of CommercialPOI_den increases from approximately 28% in 2010 to more than 35% in 2020. Since the POI layer is assumed to be temporally invariant, please clarify how this increase should be interpreted. Does the higher SHAP importance in 2020 reflects a stronger dependence of the model on commercial POIs? Please provide a clearer mechanistic explanation to avoid misinterpretation.

Response:

Thank you for this important comment. We agree that the increase in the SHAP importance of CommercialPOI_den from 2010 to 2020 requires careful interpretation, particularly because the same 2020 POI layer was used in both models. The higher SHAP importance of CommercialPOI_den in the 2020 model should not be directly interpreted as evidence that population distribution became more strongly dependent on commercial POIs over time. SHAP importance is model- and data-specific: it quantifies the average contribution of a feature to the predictions of a particular model given the distribution of all predictors and the response variable, rather than representing an intrinsic or temporally comparable importance of the feature itself. In our study, the 2010 and 2020 models were trained independently using population data and other predictors corresponding to their respective years. Although the

CommercialPOI_den layer remains unchanged, the population distribution, 3D building characteristics, and distributions and interactions of the other predictors differ between the two years. These differences can change how the independently trained models use CommercialPOI_den and consequently result in different SHAP importance values. Therefore, the increase from approximately 28% in 2010 to more than 35% in 2020 indicates that CommercialPOI_den accounts for a larger share of the model attribution in the 2020 model under its specific predictor and population distributions, rather than demonstrating a temporal strengthening of the underlying population–commercial POI relationship. In particular, because the POI layer is temporally invariant, the difference in SHAP importance cannot by itself be used to infer changes in the actual influence of commercial activities on population distribution between 2010 and 2020.

We have revised Section 4.4 to clarify this interpretation and explicitly caution against interpreting differences in absolute SHAP importance as temporal changes in the underlying population–POI relationship. Instead, the cross-year comparison focuses on the general consistency of feature rankings, contribution patterns, and the directionality of SHAP values. The corresponding revision in Section 4.4 is as follows:

(In Lines 721–752)

"To improve the interpretability of the XGBoost-based dasymetric mapping model, the SHAP framework was employed to quantify the contribution of each feature and characterize its relationship with population density. The SHAP summary plots for 2010 and 2020 (Fig. 10) exhibit generally consistent feature rankings and contribution patterns despite the models being trained independently for the two years. Although differences in relative SHAP importance are observed, these values should not be interpreted as direct temporal changes in feature importance. SHAP values quantify the average contribution of each feature to predictions for a specific model and dataset, rather than representing an intrinsic importance of the feature itself. For example, the relative SHAP importance of commercial POI density increases from approximately 28 % in 2010 to more than 35 % in 2020. Because the same POI layer was used for both years, this increase does not necessarily indicate a stronger population dependence on commercial POIs in 2020, but may reflect differences in population distribution, other temporally varying predictors and their interactions, and how the independently trained models utilize these predictors We therefore focus on the consistency of feature rankings, contribution patterns, and the directionality of SHAP values rather than directly comparing their absolute magnitudes between years.

Commercial POI density is ranked as the most influential feature in both models, with higher densities consistently associated with positive SHAP values, suggesting a strong relationship between commercial intensity and population concentration. This contribution may partly reflect the high spatial representativeness of commercial POIs, which are generally more comprehensively documented than residential population carriers. Moreover, POI density primarily serves as a proxy for the spatial organization of urban functions rather than a direct indicator of the historical evolution of individual facilities, whose establishment times are often unavailable. Its importance should therefore be interpreted as representing urban functional intensity rather than the temporal development of commercial activities.

In contrast, slope generally exhibits negative SHAP values, implying that steeper terrain constrains population concentration, although a limited number of high-slope samples show positive contributions. 3D building volume consistently ranks among the most influential predictors, with larger building volumes corresponding to positive SHAP contributions. Building-function variables further distinguish different urban land-use characteristics: higher residential building fractions generally increase predicted population density, whereas commercial and industrial building fractions tend to contribute negatively.

These results highlight the importance of incorporating vertical urban morphology and functional characteristics into fine-scale population mapping. The persistent contribution of 3D building volume indicates that population-carrying capacity in dense urban environments cannot be adequately represented by conventional 2D built-up indicators alone. Building-function variables also provide complementary information beyond POI-based indicators. The relatively limited explanatory power of residential POI density may result from the incomplete representation of residential communities in POI datasets despite their dominant role as population carriers. Incorporating building-based functional information therefore helps better characterize residential land use and reduce potential biases in population spatialization."

6. Figure 3 includes a "District-level Census Data" component under the Population Constraints module, yet district-level census data are rarely mentioned elsewhere in the manuscript. Please clarify whether district-level constraints were incorporated into the baseline population allocation.

Response:

Thank you for pointing out this inconsistency. The label "District-level Census Data" in Fig. 3 was not sufficiently precise and may have caused confusion. The population constraints were derived from county-level administrative units in China, which include counties, districts, county-level cities, and other county-level administrative divisions. To ensure consistency with the terminology used throughout the manuscript and to more accurately reflect the administrative level of the census data, we have revised the label in Fig. 3 from "District-level Census Data" to "County-level Census Data." The figure and the corresponding text have been updated accordingly, as follows.

(In section 2.1, Lines 173–177):

"……Here, county-level administrative units refer collectively to counties, urban districts, county-level cities, and other administrative divisions at the equivalent level in China. Population statistics for 2,845 such county-level units across mainland China were digitized and linked to corresponding county boundary data acquired from the Tianditu platform (www.tianditu.gov.cn)"

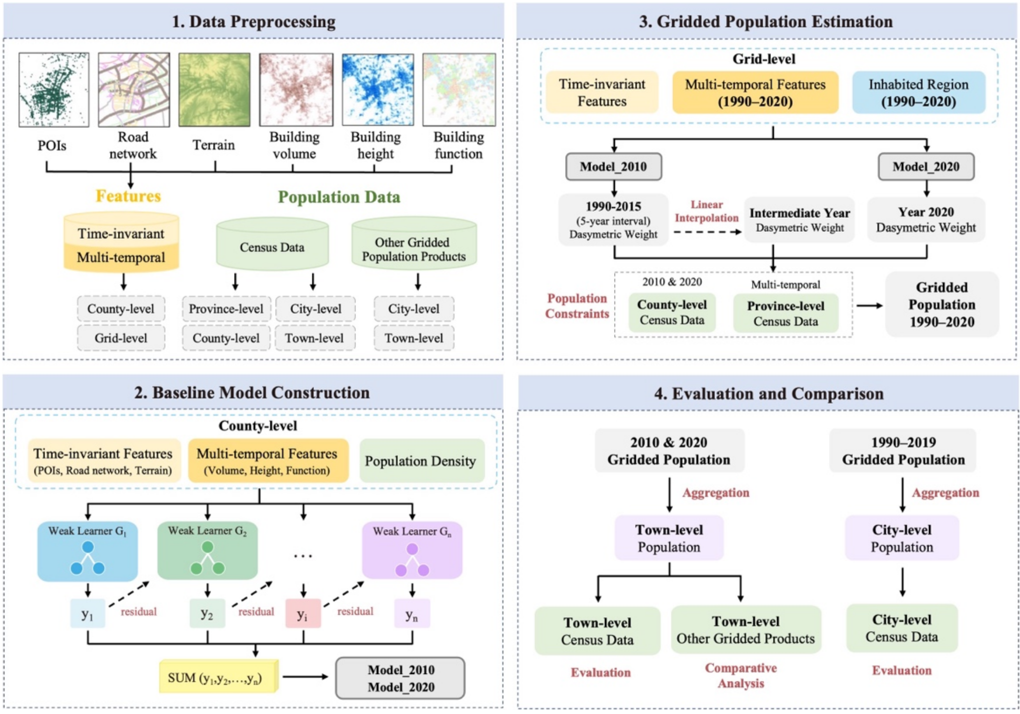

Fig. 3 The overall workflow of deriving the AGPC dataset

7. The word "Grided" appears repeatedly in the title, abstract, and throughout the manuscript (e.g., "500 m Grided Population"). The correct spelling is "Gridded". Please correct this throughout the manuscript.

Response:

Thank you for pointing out this spelling error. We have corrected "Grided" to "Gridded" in the title, abstract, and throughout the manuscript. Please refer to the revised manuscript for more details.

8. In the reference list, "China Statics Press" should be corrected to "China Statistics Press".

Response:

We have corrected "China Statics Press" to "China Statistics Press". Please refer to the updated reference list in the revised manuscript.

Reviewer #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."

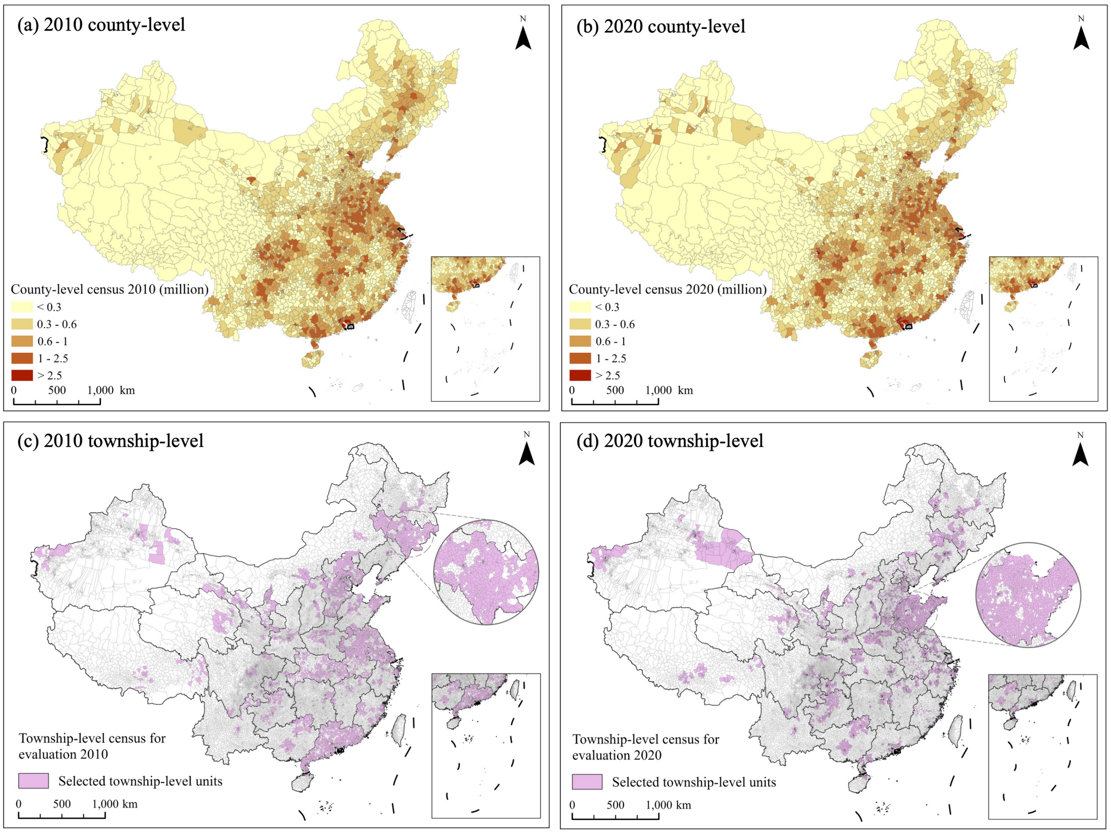

Fig. 1 Population census data from Sixth and Seventh National Population Census of China, including county-level population data for 2010 (a) and 2020 (b), and the spatial distribution of selected township-level census units for 2010 (c) and 2020 (d)

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."

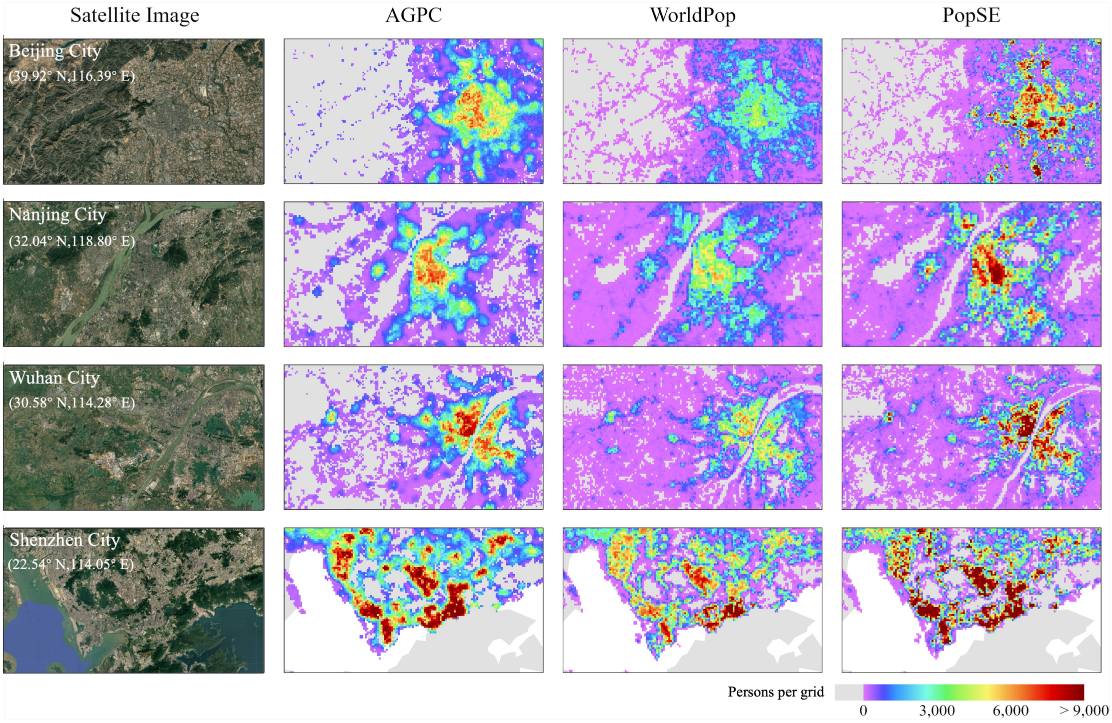

Fig. 1 Comparison of gridded population distributions in 2020 among AGPC, WorldPop, and PopSE in selected areas

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:

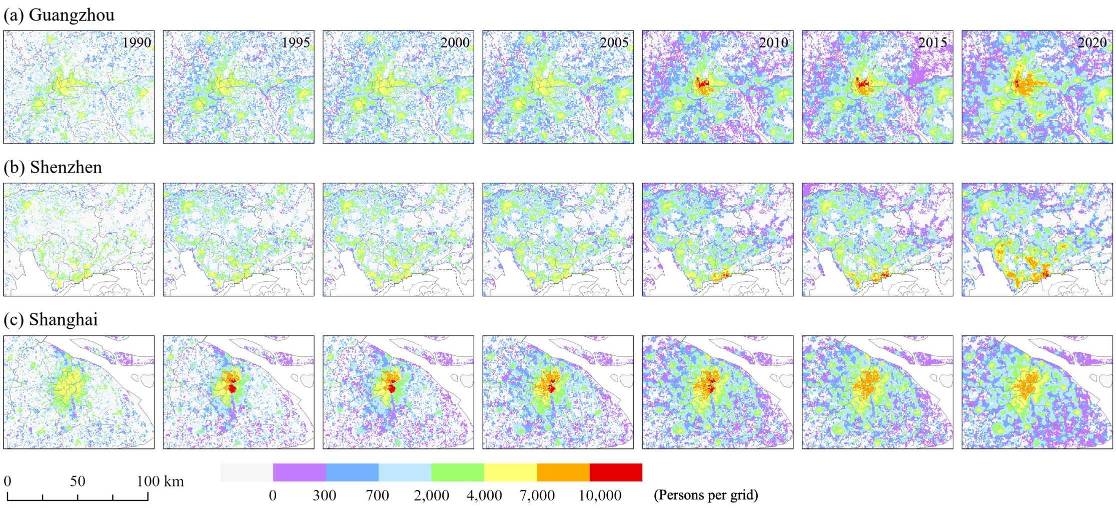

Fig. 9 Changes in the spatial distribution of population (persons per grid) in Guangzhou (a), Shenzhen (b), and Shanghai (c) during 1990–2020

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.