Responses to Editor

Dear Editor,

Thank you very much for your helpful comments. The manuscript has been fully revised according to your suggestions and the reviewers' comments. The following is a point-to-point response to your comments, and responses are in blue.

Comments from the Editors and Reviewers:

Dear Q. L. Feng,

We are pleased to inform you that the open discussion of your following ESSD manuscript was closed:

Title: Global-PCG-10: a 10-m global map of plastic-covered greenhouses derived from Sentinel-2 in 2020
Author(s): Bowen Niu et al.
MS No.: essd-2024-538
MS type: Data description paper

No more referee comments and short comments will be accepted. Now the public discussion shall be completed as follows:

You - as the contact author - are requested to respond to all referee comments (RCs) by posting final author comments (ACs) on behalf of all co-authors no later than 23 Jul 2025 (final response phase). Please log in using your Copernicus Office user ID at: https://editor.copernicus.org/essd-2024-538/final-response

When replying to the referee comments (RCs) it is sufficient to post one author comment (AC) by starting a new discussion thread. Please also consider replying to community comments (CCs) from the scientific community.

After your AC posts, you have to explicitly finalize the final-response form through the button "Finalize". You will then receive a separate email asking you to prepare and submit your revised manuscript for peer-review completion and potential final publication in ESSD.

Preparation and submission of a revised manuscript is encouraged only if you can satisfactorily address all comments and if the revised manuscript meets the high quality standards of ESSD (https://www.earth-system-science-data.net/peer_review/review_criteria.html). In case of doubt, please ask the handling topic editor directly whether they would encourage submission of a revised manuscript or not.

Please note also that the submission of a revised manuscript does not ensure publication in ESSD. The topic editor will carefully assess your revised manuscript in view of the interactive public discussion and may forward it to the original or new referees for further commenting.

You are invited to monitor the processing of your manuscript via your MS overview at: https://editor.copernicus.org/ESSD/my_manuscript_overview

Thank you very much in advance for your cooperation. In case any questions arise, please do not hesitate to contact me.

Kind regards,

The editorial support team
Copernicus Publications
editorial@copernicus.org

Response:

We have modified the manuscript carefully and now we resubmit our revised version. Thank you again for your support on our research.

Yours sincerely,
Bowen Niu, Quanlong Feng, China Agricultural University
on behalf of all the co-authors

Responses to RC #1

Dear Reviewer,

Thank you very much for your helpful comments. The manuscript has been fully revised according to your suggestions. The following is a point-to-point response to your comments, and responses are in **blue.**

General comments.

This very interesting paper use machine learning and deep learning on Sentinel-2 10 m GSD images to obtain a global map of plastic-covered greenhouses (PCGs). Really, it is not the first global PCG maps since Tong et al. (2024) already published other global PCG map derived from PlanetScope images, a commercial satellite with 3 m GSD, but using also in the first steps Sentinel-2. Although both works have a similar objective (i.e., to attain a global PCG map), the strategies used were quite different.

Thanks to the attained global PCG map, Niu et al. (2025) give interesting data about the area of PCG around the world, the major concentrations, spatial distribution, etc.

The manuscript is well written and it is worth being published. However, a few specifics comments should be taken into account.

Response:

Thank you for your encouragement and support on our study. We have revised the manuscript carefully according to your suggestions. The following is a point-to-point response to each suggestion.

Specific comments.

1. Some cites in the manuscript appears with an extra comma. For example, in Page 2, Line 64, the cites "Aguilar et al., (2016) and Yang et al., (2017) independently developed…" should be "Aguilar et al. (2016) and Yang et al. (2017) independently developed…" Similarly, "Zhang et al., (2022a)" in Page 2, Line 66, should be "Zhang et al. (2022a)". Please, correct this issue throughout the manuscript.

2. Page 3, Line 71. The cite (Zhang et al., 2024) should be (Zhang et al., 2024a).

3. Page 3, Line 76. Zhang et al., 2023 should be 2023a or 2023b. Please, review it.

Response:

Thank you for your careful and detailed review. We have thoroughly checked the manuscript and corrected all instances of extra commas in citations to the proper format (e.g., changing "Aguilar et al., (2016)" to "Aguilar et al. (2016)"). Additionally, we have updated specific citations for Zhang et al. to include the correct suffixes, changing "(Zhang et al., 2024)" to "(Zhang et al. 2024a)" on Page 3, Line 71 and clarifying the year suffixes for Zhang et al. 2023 as appropriate (e.g., 2023a or 2023b) to ensure

consistency throughout the manuscript.

4. Page 5, Line 139-140. “Actually, Sentinel-2 is a constellation consisting of two satellites, i.e., Sentinel-2A and Sentinel-2B, which are in the same sun-synchronous orbit while phased at 180° to each other”. In fact, there is a new Sentinel-2C. You should speak a little about it.

Response:

We sincerely thanks for your valuable comment. We have noted the latest development of the Sentinel-2 constellation. As you pointed out, in addition to Sentinel-2A and Sentinel-2B, Sentinel-2C has also been successfully launched on March 3rd, 2024 by ESA to ensure the continuity of Earth observation capability. In the revised manuscript, we have updated the relevant description to reflect the current composition of the Sentinel-2 constellation and have briefly supplemented information about the role of Sentinel-2C. The revised content is as follows:

Sentinel-2 multispectral images were used in this study. As the important part of ESA’s Copernicus Programme, Sentinel-2 aims to provide global Earth Observation data at a fine scale with 10 meters captured by MultiSpectral Instrument (MSI) with a total of 13 bands and a swath width of 290 km. Sentinel-2 is a satellite constellation initially composed of Sentinel-2A and Sentinel-2B, which operate in the same sun-synchronous orbit but are phased 180° apart to ensure a high revisit frequency. In addition, Sentinel-2C, the third satellite in the constellation, was successfully launched in March 2024. It serves as a replacement unit to ensure data continuity and system redundancy throughout the mission duration. Several reg-edge bands that are very sensitive to vegetation have been designed in Sentinel-2, which could capture a more detailed conditions of vegetated regions than other satellites such as Landsat and MODIS.

The above contents have been added in Section 2.2 “Satellite datasets” of the manuscript (**see *Lines 135 ~ 143* for details**).

5. Page 6, Line 163. In Figure 2 (Stage 2) the train/validation ratio is 7:3, and in the manuscript you wrote 8:2. Is it a mistake in the Figure?

6. Page 12, Line 273. In Figure 2 (Stage 2) the train/validation ratio is 7:3, and in the manuscript you wrote 8:2. Please review it.

Response:

Thank you very much for pointing out this inconsistency. We feel sorry for our careless mistake. Actually, there is a mismatch between the figure and the description in the manuscript. We have revised Figure 2 to ensure that the train/validation ratio is consistent with what is reported in the manuscript (8:2). We sincerely appreciate your careful review and helpful comment.

7. Page 7, Line 175. In the caption of Figure 3, you should clarify that the size of the reference samples (512×512) are pixels and not meters.

8. Page 11, Line 237. You should clarify also in the manuscript that the size of the reference samples (512×512) are pixels and not meters.

Response:

Thank you for your valuable suggestion. We have revised the manuscript at Page 7, Line 175 and Page 11, Line 237 to explicitly clarify that the size of the reference samples (512×512) refers to pixels and not meters to avoid any misunderstanding. Specifically, we have updated the relevant expressions to "512 × 512 pixels" throughout the manuscript for consistency and clarity.

9. Page 9, Line 200. In the caption of Figure 4, it is written Multiple-temporal NDVI. Is not more appropriated multi-temporal NDVI?

Response:

Thank you for pointing this out. We agree that "multi-temporal NDVI" is the more appropriate term. We have corrected the caption of Figure 4 accordingly.

10. Page 9, Line 205. "1> Spectral features". Strange login method.

11. Page 10, Line 212. "2> Textural features". Strange login method.

Response:

Thank you for your careful review. We totally agree with you that the notation "1>" and "2>" in these lines are not appropriate and may be confusing. We have revised them to standard numbering (e.g., "(1) Spectral features" and "(2) Textural features") to improve clarity and consistency.

12. Page 13, Line 286. Fu et al. (2021) is not in reference section.

Response:

Thank you for pointing this out. As suggested, we have added the missing reference for Fu et al. (2021) to the reference section to ensure completeness and consistency.

13. Page 15, Line 332-335. There are some numbers without thousands separation (e.g., 9874.51 km2, 2530.56 km2, 8224.90 km2).

Response:

Thank you for your careful inspect. We have revised these numbers in Line 332–335 to include thousands separation (e.g., 9,874.51 km², 2,530.56 km², 8,224.90 km²) to improve readability and consistency.

14. Page 16, Line 344. Figure 8a is not cited in the manuscript, and it should be.

Response:

Thank you for bringing this to our attention. We have revised the manuscript to include the appropriate citation of Figure 8a to ensure that all figures are properly referenced in the manuscript.

15. Page 18, Line 375. Why 20500 points for GH and 20500 for Non-GH. Justify this figure.

Response:

Thank you for this good question. Initially, we selected 20,500 test samples each for PCG and Non-PCG equally, with the primary goal of ensuring statistical stability for calculating the overall accuracy (OA) and evaluating the classification performance for the PCG category. However, through further literature review and methodological refinement, we have realized that this balanced sampling approach did not fully consider the effects of class imbalance on accuracy assessment.

To address this issue, we referred to the methodology proposed by Olofsson et al. (2014), and drew on best practices from land use and land cover classification studies such as Wang et al. (2023) and Tian et al. (2025) to re-sample PCG and non-PCG and re-calculate the confusion matrix. In the updated method, we strictly followed the stratified random sampling strategy recommended by Olofsson et al. (2014), in which samples were selected in proportion to the actual mapped area of each class within the study region.

However, since PCG covers less than 1% of the global area, a strictly proportionate sampling approach would yield an insufficient number of PCG samples, making it difficult to effectively assess its classification accuracy. To address this issue, we adopted the approach used in the aforementioned studies and increased the proportion of PCG samples in the test dataset to approximately 10%. Now that the number of PCG is 6,000 while Non-PCG is 40,000. This adjustment could enhance the evaluation capability for this minority class (PCG) and ensures the scientific rigor and representativeness of the final accuracy estimates.

References

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. and Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote sensing of Environment*, *148*, pp.42-57.

Tian, F., Wu, B., Zeng, H., Zhang, M., Zhu, W., Yan, N., Lu, Y. and Li, Y., 2025. GMIE: a global maximum irrigation extent and central pivot irrigation system dataset derived via irrigation performance during drought stress and deep learning methods. *Earth System Science Data*, *17*(3), pp.855-880.

Wang, M., Mao, D., Wang, Y., Xiao, X., Xiang, H., Feng, K., Luo, L., Jia, M., Song, K. and Wang, Z., 2023. Wetland mapping in East Asia by two-stage object-based Random Forest and hierarchical

decision tree algorithms on Sentinel-1/2 images. *Remote Sensing of Environment*, *297*, p.113793.

16. Page 19, Line 381. Table 1 shows the confusion matrix where OA, User Accuracy (UA) and Producer Accuracy (PA) are depicted. Really, UA=Recall and PA=Precision, so, Table 2 is not necessary. The only data useful in Table 2 is F1 Score. I think that you should rewrite the methods and results about the accuracy assessment. Furthermore, Why is the classification so biased? For example, UA is 99.99% and PA is 86.30% for Non-GH and, UA is 84.18% and PA is 99.99% for GH.

Response:

We totally agree that there was redundant information between Table 1 and Table 2. Following your suggestion, we have removed Table 2 and redesigned Table 1 accordingly.

In addition, the current description regarding the sampling strategy, sample proportion design and the reliability of the test samples in the confusion matrix was insufficient. We have revised this section based on the reconstructed confusion matrix and now provide a detailed explanation of the test sample collection process. The specific modifications are as follows.

To further quantitatively evaluate the reliability of the Global-PCG-10 dataset, we constructed a dedicated test sample set. The spatial distribution of test samples is shown in Figure 10.

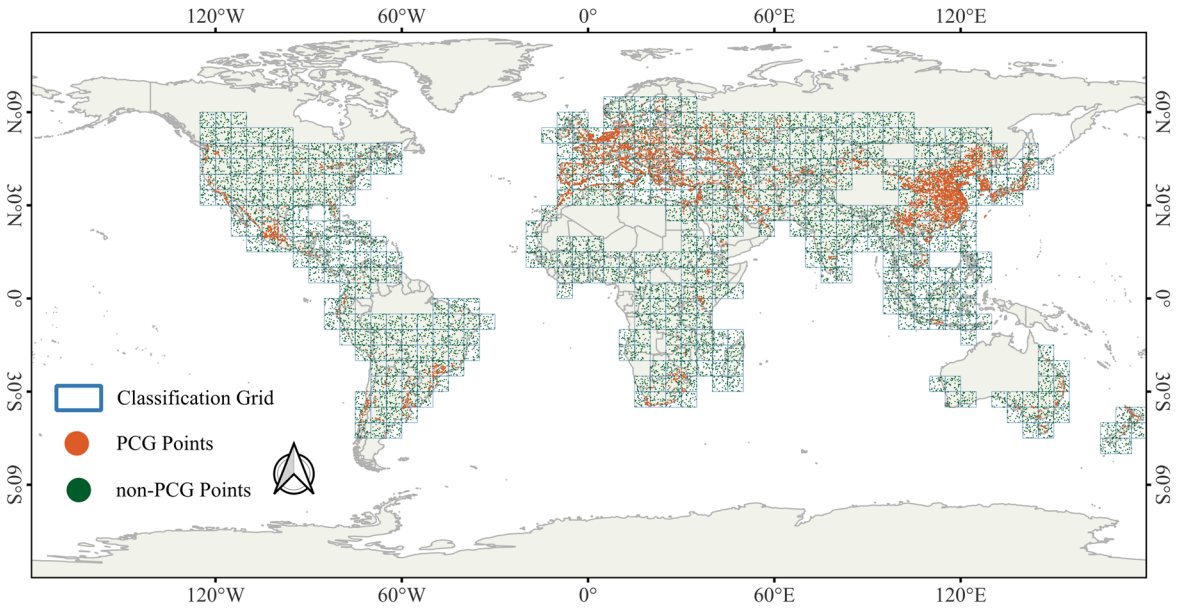

Figure 10. Spatial distribution of global test samples.

The dataset includes two categories, PCG and non-PCG. Based on previous research practices (Olofsson et al., 2013, 2014; Tian et al., 2025; Wang et al., 2023), we followed the stratified random sampling strategy recommended by Olofsson et al. (2014), in which samples were drawn in proportion to the mapped area of each class within the actual mapping region. However, since the global coverage of PCG is less than 1%, strictly proportional sampling would result in too few PCG samples to support a

statistically robust accuracy assessment. To address this issue, and consistent with the approaches adopted in the above studies, we moderately increased the proportion of PCG samples in the test set to approximately 10%. This adjustment could enhance the evaluation capability for this minority class.

As shown in Table 1, the total number of test samples is 46,000, with 6,000 PCG samples and 40,000 non-PCG samples. To ensure the validity, we applied separate sampling strategies for each category. As for PCG, test samples were derived from the global 3-meter PCG dataset in 2019 developed by Tong et al. (2024), and manually verified through Google Earth visual interpretation. Since the Global-PCG-10 dataset is for the year 2020, and considering that PCGs typically have long lifespans and stable structures, the 2019 dataset by Tong provides a reliable reference. Additionally, we performed a second round of verification using historical Google Earth imagery in around 2020 to confirm their existence and status, minimizing sampling bias from prior knowledge. And for non-PCG, due to the large quantity required, manual sampling was impractical. We thus randomly sampled non-PCG from the GLC_FCS30D dataset to ensure independence and randomness. All samples were also verified through visual interpretation of historical Google Earth imagery in around 2020 to ensure label correctness.

Table 1. Confusion matrix.

Confusion Matrix	**Reference**: Non-PCG	**Reference**: PCG	*UA (%)*
Predicted: Non-PCG	39,991	893	97.82 ± 0.13
Predicted: PCG	9	5,107	99.82 ± 0.11
PA (%)	99.98 ± 0.01	85.12 ± 0.90	
F1-score (%)	-	91.88 ± 2.71	
OA (%)			98.04 ± 0.12

Note: PA, Producer's Accuracy; UA, User's Accuracy; OA, Overall Accuracy.*

Based on this test dataset, Global-PCG-10 achieved a PA of 85.12% ± 0.90%, a UA of 99.82% ± 0.11%, an F1-score of 91.88% ± 2.71% and an overall accuracy of 98.04% ± 0.12% (Table 1). In the revised confusion matrix, the bias for non-PCG has been effectively reduced. However, PCG still exhibits a gap between precision and recall, characterized by a high precision but a low recall. This may be caused by missed detections of small PCG patches. Unlike PlanetScope, Sentinel-2 has lower spatial resolution with 10 meters, and small PCG often spans only a few mixed pixels, making it difficult to extract meaningful spectral features for accurate PCG classification. The high precision, on the other hand, is likely due to post-processing applied to the initial classification results. Among these steps, the Sieve Filter method played a key role by removing small, erroneous regions through multi-level filtering, thereby improving the quality of PCG predictions and enhancing precision.

The above contents have been added in Section 4.2 "Reliability of Global-PCG-10" of the

manuscript (**see *Lines 423 ~ 455* for details**).

References

Olofsson, P., Foody, G.M., Stehman, S.V. and Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote sensing of environment*, *129*, pp.122-131.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. and Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote sensing of Environment*, *148*, pp.42-57.

Tian, F., Wu, B., Zeng, H., Zhang, M., Zhu, W., Yan, N., Lu, Y. and Li, Y., 2025. GMIE: a global maximum irrigation extent and central pivot irrigation system dataset derived via irrigation performance during drought stress and deep learning methods. *Earth System Science Data*, *17*(3), pp.855-880.

Wang, M., Mao, D., Wang, Y., Xiao, X., Xiang, H., Feng, K., Luo, L., Jia, M., Song, K. and Wang, Z., 2023. Wetland mapping in East Asia by two-stage object-based Random Forest and hierarchical decision tree algorithms on Sentinel-1/2 images. *Remote Sensing of Environment*, *297*, p.113793.

17. Page 22, Line 421-422. "… and in May 2024, the University of Copenhagen published a global 3-m PCGs dataset also in 2019". Please, you should cite Tong et al. (2024) here.

18. Page 22, Line 430. You should cite Tong et al. (2024) properly in the caption of Figure 12.

Response:

Thank you for your suggestion. We have now cited Tong et al. (2024) appropriately in the manuscript to reference the global 3-m PCG dataset published by the University of Copenhagen. Additionally, we have added the proper citation in the caption of Figure 12 to ensure full attribution.

19. Page 23, Line 434. "Tong et al., (2024) acquired from 3-m …". Again, this cite appears with an extra comma.

Response:

Thank you for pointing out the punctuation error. We have corrected the citation on Page 23 by removing the extra comma, so it now reads "Tong et al. (2024) acquired from 3-m …" for proper formatting.

20. Page 30, Line 667. "Zhang, X., Liu, L., and Chen, X.: Global annual wetland dataset Data Descriptor at 30 m with a fine classification system from 2000 to 2022, Sci. Data, https://doi.org/10.1038/s41597-024-03143-0, 2024c". This reference do not appear in the manuscript.

Response:

Thank you for your careful review. We have now included the citation for Zhang et al. (2024c) within the manuscript where relevant to ensure consistency between the references and the main manuscript.

Final Comments:

It is very important that the global PCG map and the code are accessible to researchers. I have tested that the code for generating the initial labels of PCGs is publicly available via the following link on Google Earth Engine: https://github.com/MrSuperNiu/Greenhouse_Classification_GEE. It consists of feature extraction, RF classification, etc. Additionally, the code of APC-Net is accessible through the following link: https://github.com/MrSuperNiu/APCNet. The Global-PCG-10 dataset is stored on figshare, and can be downloaded here: https://doi.org/10.6084/m9.figshare.27731148.v2 (Niu et al., 2024).

Response:

Thank you for your careful check on our open-access dataset and code. We are happy to share with the community our global PCG map and hope it helps for other researchers.

Thank you again for your comments. They are valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our studies.

Yours sincerely,

Bowen Niu, Quanlong Feng

on behalf of all the co-authors

Responses to RC #2

General comments：

The manuscript presents the development of a global plastic greenhouse distribution map (Global-PCG-10) based on Sentinel-2 10 m imagery. Weak labels were generated via a random forest classifier, and these were refined through a deep learning framework integrated with an active learning strategy. The spatial patterns, extent, and proportional coverage of global PCGs were analyzed, yielding results of clear scientific merit and practical utility. However, several issues and deficiencies must be addressed prior to publication:

Response:

Thank you for your encouragement and support on our study. We have revised the manuscript carefully according to your suggestions. The following is a point-to-point response to each suggestion.

1. Some in-text citations do not conform to the journal's formatting requirements. For example, in line 64 the citation "Aguilar et al., (2016)" incorrectly includes a comma. Please review and standardize all reference notations.

Response:

We appreciate your careful attention on the manuscript. As suggested, we have thoroughly checked all in-text citations and have revised to ensure consistency with the journal's formatting requirements. Specifically, we have corrected instances such as "Aguilar et al., (2016)" to "Aguilar et al. (2016)" and made similar adjustments wherever necessary.

2. In Figure 2 (Phase 2) the train/validation split is stated as 7:3, whereas in the main text it is described as 8:2. Please verify and ensure consistency in both places.

Response:

Thank you very much for pointing out this inconsistency. We feel sorry for our careless mistake. Actually, there is a mismatch between the figure and the description in the manuscript. We have revised the figure to ensure that the train/validation ratio is consistent with what is reported in the manuscript (8:2). We sincerely appreciate your careful review and helpful comment, which helped us improve the clarity and accuracy of our work.

3. In lines 162-163 authors state that the initial labels were generated via GEE and RF classification. Was any quality control applied to these labels (e.g., manual verification rate, error-rate assessment)? Please clarify.

Response:

We appreciate your concern regarding the quality control of the initial labels. Actually, prior to training the deep learning model, we implemented multiple strategies to ensure the accuracy and reliability of initial PCG labels, as detailed below.

(1) Collection of high-confidence samples.

To construct the training samples required for GEE-based Random Forest (RF) classification, we conducted field surveys in key greenhouse-intensive regions in China (e.g., Weifang in Shandong, Kunming in Yunnan, and Lishu in Jilin). During the surveys, we also consulted local farmers to confirm the locations and types of PCG. Considering that PCG typically remain in use for around 10 years or more with relatively high stability (Ou et al., 2021), we performed systematic manual visual interpretation of historical high-resolution imagery from Google Earth in multiple global regions to obtain high-confidence samples. For areas outside China, we additionally referred to published literature, meta-analyses and online sources for auxiliary identification. All samples were further verified using Sentinel-2 imagery to ensure their actual presence in the year 2020. We also refined PCG and non-PCG labels based on the RF classification results within each grid to enhance overall labeling accuracy.

(2) Quality assessment and selection of RF classification results.

Based on the collected samples, we trained a RF model with GEE, using a split between training and validation sets. A confusion matrix was constructed to evaluate the classification accuracy, where the validation set was excluded from the training process and used solely for accuracy assessment. Only those classification maps with an overall accuracy (OA) greater than 95% and a user accuracy (UA) for the PCG class above 90% were selected as candidate label maps for training the deep learning model.

(3) Final screening of training labels for the deep learning model.

All candidate label maps were undergone further visual inspection. Each 512×512 pixels image patch was visually checked to ensure high annotation quality, and only the regions with the most reliable classification results were retained for deep learning model training. This process ensured that the final training labels used in the deep learning model were of high reliability.

The above contents have been added in Section 3.1 “Stage-1: PCG weak label generation” of the manuscript (**see *Lines 178 ~ 200* for details**).

4. In lines 165-166 (Stage 3) mention “post-processing was applied to the PCGs classification results to eliminate isolated noises” without specifying the algorithm or parameter settings. Please provide details of the post-processing method so that the results are reproducible.

Response:

We sincerely thank you for the valuable suggestion. In this study, we adopted a post-processing step

that used a pixel-connected component-based Sieve Filter to eliminate isolated noises in the initial classification results. Specifically, we used the gdal.SieveFilter() function from the GDAL library (invoked in the Python environment) to perform the filtering. An 8-connected neighborhood was adopted, and a set of hierarchical thresholds for the minimum number of connected pixels (10 / 20 / 50) was applied. This multi-level threshold setting was designed to accommodate variations in noise distribution and mapping requirements across different regions.

The above contents have been added in Section 4.2 "Reliability of Global-PCG-10" of the manuscript (**see *Lines 455 ~ 458* for details**).

5. The manuscript alternates between "PCG" and "PCGs" (e.g., line 166 "PCGs classification results" vs. Figure 2 "PCG classification results"). Please unify the terminology throughout.

Response:

Thank you for your valuable comments and suggestions. We have carefully reviewed and revised the usage of "PCG" and "PCGs" throughout the manuscript. The terminology has been standardized as PCG accordingly.

6. In lines 239-243 outline the overall network architecture but omit implementation details of the MDCN and non-local modules. Schematic diagrams or detailed descriptions of these submodules are required.

Response:

We appreciate your valuable suggestion. In response, we have added detailed descriptions of both the MDCN and non-local modules in the revised manuscript. Additionally, we have provided schematic diagrams illustrating the structures of these submodules to improve clarity (see new Figure 5). We believe these additions could enhance the readability and completeness of the network architecture description.

The detailed reply is as follows.

In this study, we employed a deep semantic segmentation model, APC-Net (Niu et al., 2023a), as the core model to generate the final PCG classification map in a coarse-to-fine manner. APC-Net effectively integrates local and global features through multi-scale feature learning, thereby enhancing its classification capability under complex global terrain conditions.

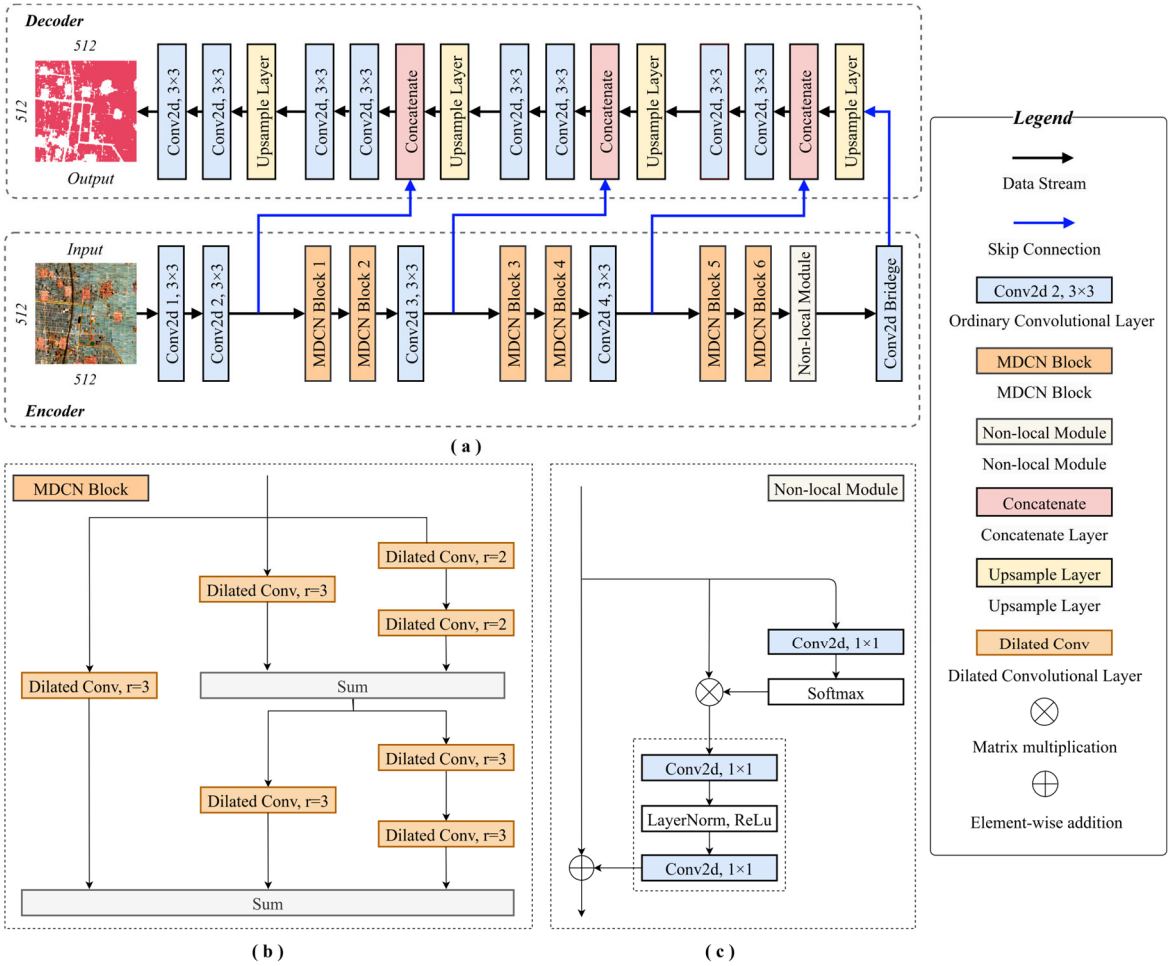

Figure 5. Overview of the proposed APC-Net model

Specifically, APC-Net consists of two main components, an encoder and a decoder (see Figure 5a). The encoder, which is the core of the network, takes a 512×512 remote sensing image patch as input and extracts highly representative features through a multi-layer structure. This not only enhances intra-class consistency but also improves inter-class separability. The encoder includes convolutional layers, an MDCN (Multi-scale Dilated Convolutional Network) module and a non-local module. The MDCN module (Figure 5b) integrates multi-scale dilated convolutions to effectively capture multi-scale local features, addressing scale variation issues that are common in PCG classification. The non-local module (Figure 5c) focuses on capturing global contextual information, thereby improving the model's capability for overall scene understanding. The decoder is responsible for restoring spatial information from the downsampled feature maps generated by the encoder and producing the final segmentation map. It employs bilinear interpolation for upsampling and skip connections to fuse encoder features, further refining the representation. The final PCG classification output maintains the same spatial resolution (512×512) as the input image.

The above contents have been added in Section 3.1.2 "APC-Net model" of the manuscript (**see *Lines***

***255 ~ 271* for details**).

7. Section 3.2.2 provides an overview of the active-learning strategy but lacks specifics on the number of iterations, per-iteration sample-selection criteria, and stopping conditions. Please supply quantitative stopping rules rather than the current qualitative phrase "until performance stabilizes or results are satisfactory."

Response:

Thanks for your valuable comment. In this study, the active learning process was conducted for up to five iterations. Each iteration involved a complete model training procedure, followed by performance evaluation using a validation set, focusing primarily on overall accuracy (OA) and mean Intersection over Union (mIoU). To determine whether to proceed with additional iterations, we applied a quantitative stopping criterion as follows. If the improvement in both OA or mIoU between two consecutive iterations was less than 1%, the model was considered to have reached performance stability, and the iterative process was terminated.

Regarding the sample selection strategy for each iteration, we identified candidate samples for manual review by comparing the model predictions with the initial weak labels. Specifically, an image patch was selected for further review if the Patch-level IoU between the model prediction and the weak label was less than 0.6. This threshold was determined empirically based on our observation of preliminary results and manual verification, aiming to balance classification accuracy and the efficiency of the active learning process.

The Patch-level IoU served as a metric to quantify the degree of disagreement between the model and the weak labels within a single patch, helping to identify potential labeling errors or uncertain regions. This patch-wise screening strategy enabled each iteration to focus on the most informative samples, thereby effectively improving both training data quality and model performance.

The above contents have been added in Section 3.2.2 "Active Learning strategy" of the manuscript (**see *Lines 284 ~ 300* for details**).

8. In lines 253-254, authors state that "if results are unsatisfactory, significantly altered initial labels are selected for further training," but "expected standards" and "significant alteration" are not quantified. Please define the thresholds and the human-involvement workflow in detail.

Response:

Thank you for your valuable comments. In response to your concern regarding the quantification of "expected standards" and "significant alteration," we have provided further clarification as follows.

(1) Expected Standards. In practice, we did not predefine rigid accuracy thresholds as expected

standards during model training. Instead, we primarily relied on visual interpretation of the PCG classification results to evaluate model performance. This approach is grounded in the following considerations. First, an early stopping mechanism was applied during deep learning model training to prevent overfitting. Second, the training and validation sets were split in an 8:2 ratio to maximize the use of sample data and improve model performance. Third, the training process involved iterative optimization of initial labels, allowing the model to learn more representative features and thus achieve improved PCG classification results.

Since it is difficult to predict the final performance of the model prior to training, we evaluated the results based on empirical experience, classification accuracy metrics and spatial distribution consistency. According to previous experience, an Overall Accuracy (OA) above 90% and a mean Intersection over Union (mIoU) above 0.6 are generally considered acceptable for binary classification tasks and served as our practical reference thresholds.

(2) Significant Alteration. "Significant alteration" refers to cases where there is a clear discrepancy between the initial labels and the model's predictions. Specifically, we quantified this by defining an image patch as significantly altered if the Patch-level IoU between the initial labels and model predictions is below 0.6. In such cases, the image patch is flagged for further review and potential label updating.

(3) Human-involvement Workflow. Once the model identifies image patches requiring label updates, the following process is initiated. The model calculates the Patch-level IoU between its predictions and the initial labels for each image patch. If the IoU is below 0.6, the image patch is flagged for manual review. Human annotators examine these image patches to verify whether the predicted labels align with the actual conditions. If the labels are found to be incorrect, they are manually corrected and returned for model retraining. The updated labels are then incorporated into the training dataset for the next iteration. This iterative process continues until either (a) the model reaches the expected performance standards, or (b) the improvement in either OA or mIoU across two consecutive iterations is less than 1%, the training process is terminated.

The above contents have been added in Section 4.2 "Reliability of Global-PCG-10" of the manuscript (**see *Lines 284 ~ 300* for details**).

9. In lines 270–275, authors note the use of the Adam optimizer and data partitioning but omit critical hyperparameters such as total epochs, batch size, and learning-rate decay strategy. Please provide these training details.

Response:

Thank you for the helpful comment. We have provided the missing training details, including the total number of epochs, batch size, learning rate, learning-rate decay strategy and early stopping criteria

in the revised manuscript. Specifically, the model was trained for 200 epochs with a batch size of 8, using an initial learning rate of 1e-4 with a step decay schedule. In addition, early stopping was applied with a patience of 10 epochs based on the validation loss to prevent overfitting. We believe these additions could improve the transparency and reproducibility of our work.

All of the above contents have been added in Section 3.2.3 "Training Details" of the manuscript (**see *Lines 314 ~ 318* for details**).

10. In lines 272-273, the dataset is described as comprising 14 825 training images and 3 707 validation images. Please explain how the sample was divided (randomly or by region? Is global diversity guaranteed?).

Response:

Thank you for your valuable comment. Considering that PCG areas in China account for more than two-thirds of the global total, an overrepresentation of Chinese samples in the training dataset may reduce the model's ability to accurately identify PCG in other regions. To mitigate this issue, we divided the samples into two subsets, one for China and the other for non-China regions, allowing the model to learn region-specific PCG features separately. To further enhance the global representativeness of the dataset, we included PCG distribution areas across multiple continents, including Asia, Europe, North America, South America, Africa and Oceania during the sample construction stage. This could ensure sufficient geographic diversity in both the training and validation sets, thereby improving the model's generalization capability.

Specifically, we collected 10,230 samples from China and 8,302 samples from other regions. Each subset was then randomly divided into training and validation sets using an 8:2 split to train two separate regional PCG classification models. The predictions of the two models were finally combined to generate the Global-PCG-10 dataset, a global 10-meter resolution PCG classification product.

All of the above contents have been added in Section 3.2.3 "Training Details" of the manuscript (**see *Lines 319 ~ 327* for details**).

11. In line 292, the term "false positives (FP)" is incorrectly written as "false negatives." This typo may confuse the metric definitions. Please verify that all terminology and formulas in the text and tables are consistent.

Response:

We appreciate your careful reading and helpful comment. We have corrected the typo at line 292, changing "false negatives" to "false positives (FP)" as intended. We have also carefully reviewed the entire manuscript, including all metric definitions, formulas and tables to ensure consistency and accuracy.

12. Section 4.2 presents only examples with no significant false alarms, which seems inconsistent with the overall recall rate of 84%. Please include error - case analysis and discuss the causes of missed detections.

Response:

We sincerely thank you for the constructive suggestion to include an analysis of missed detection (omission) cases. In response, we have added a detailed discussion of typical omission scenarios. The relevant description is as follows.

Based on the test samples and a systematic comparison with the 3-meter resolution PCG data provided by Tong et al. (2024), we identified two main types of omission errors in the current Global-PCG-10 dataset during the PCG extraction process, as detailed below.

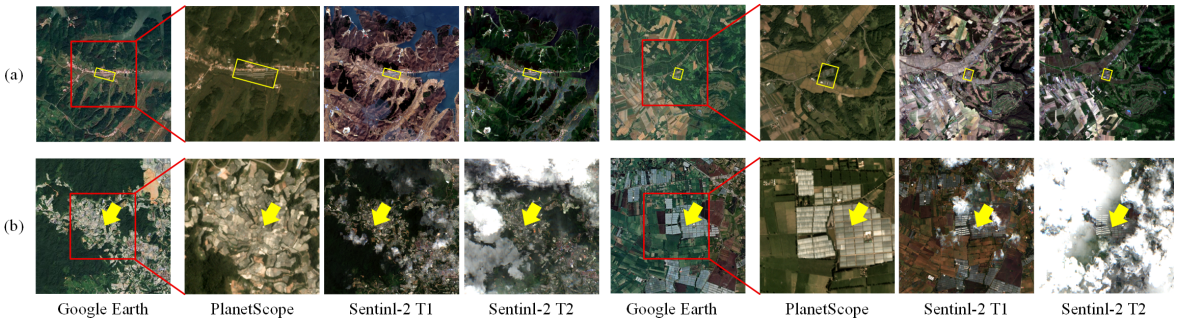

Figure 11. Bad case analysis

(Note: From left to right, the image shows the Google Earth imagery from 2020, 3-meter spatial resolution PlanetScope imagery from 2019 and Sentinel-2 imagery from spring and summer, respectively.)

As shown in Figure 11a and b, due to the relatively coarse spatial resolution (10 meters) of Sentinel-2 imagery compared to higher resolution sources such as PlanetScope or Google Earth (3 meters or finer), small-scale PCG targets often occupy only a few to a dozen pixels. These pixels are usually mixed pixels that contain spectral information from multiple surrounding land cover types. As a result, the model finds it difficult to extract PCG's distinct spectral and texture features, which impairs its ability to accurately detect small and visually inconspicuous PCG. For instance, in the area shown in Figure 11a, the PCG can be roughly identified in the high-resolution image, with some observable texture patterns. However, in the corresponding Sentinel-2 image at 10-meter resolution, the PCG contours are blurred and lack clear geometric and textural features, leading to missed detections.

Meanwhile, the Global-PCG-10 dataset is derived using multi-temporal Sentinel-2 imagery from spring and summer, organized by 1° grid tiles. However, due to cloud contamination and limited observation opportunities, it is challenging to obtain cloud-free images for both seasons in some regions (Figure 11b). This could limit the model's ability to extract consistent temporal features, thereby increasing the likelihood of omission errors. Figure 11b presents a typical case, although the overall cloud coverage is relatively low, even thin clouds can affect surface reflectance values and interfere with the

model's classification performance.

In summary, misclassification errors in PCG classification primarily arise from two aspects: (1) the presence of mixed pixels in medium-resolution imagery when detecting small-scale PCG, which weakens the model's ability to learn effective spectral and textural representations; and (2) limitations in the spatial and temporal availability of remote sensing data, particularly due to cloud cover and long revisit intervals, which may result in missing key seasonal observations and reduce classification accuracy.

All of the above contents have been added in Section 4.2 "Reliability of Global-PCG-10" of the manuscript (**see *Lines 459 ~ 484* for details**).

13. In lines 374-375, authors apply balanced sampling between PCG and non-PCG in the test set, which does not reflect their real-world scarcity and may lead to overestimated performance. Please discuss the impact of balanced sampling on evaluation or provide results under the true class distribution.

Response:

Thank you for this good question. Initially, we selected 20,500 test samples each for PCG and Non-PCG equally, with the primary goal of ensuring statistical stability for calculating the overall accuracy (OA) and evaluating the classification performance for the PCG category. However, through further literature review and methodological refinement, we have realized that this balanced sampling approach did not fully consider the effects of class imbalance on accuracy assessment.

To address this issue, we referred to the methodology proposed by Olofsson et al. (2014), and drew on best practices from land use and land cover classification studies such as Wang et al. (2023) and Tian et al. (2025) to re-sample PCG and Non-PCG and re-calculate the confusion matrix. In the updated method, we strictly followed the stratified random sampling strategy recommended by Olofsson et al. (2014), in which samples were selected in proportion to the actual mapped area of each class within the study region.

However, since PCG covers less than 1% of the global area, a strictly proportionate sampling approach would yield an insufficient number of PCG samples, making it difficult to effectively assess its classification accuracy. To address this issue, we adopted the approach used in the aforementioned studies and increased the proportion of PCG samples in the test dataset to approximately 10%. Now that the number of PCG is 6,000 while Non-PCG is 40,000. This adjustment could enhance the evaluation capability for this minority class (PCG) and ensures the scientific rigor and representativeness of the final accuracy estimates.

The above contents have been added in Section 4.2 "Reliability of Global-PCG-10" of the manuscript (**see *Lines 426 ~ 443* for details**).

References

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. and Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote sensing of Environment*, *148*, pp.42-57.

Tian, F., Wu, B., Zeng, H., Zhang, M., Zhu, W., Yan, N., Lu, Y. and Li, Y., 2025. GMIE: a global maximum irrigation extent and central pivot irrigation system dataset derived via irrigation performance during drought stress and deep learning methods. *Earth System Science Data*, *17*(3), pp.855-880.

Wang, M., Mao, D., Wang, Y., Xiao, X., Xiang, H., Feng, K., Luo, L., Jia, M., Song, K. and Wang, Z., 2023. Wetland mapping in East Asia by two-stage object-based Random Forest and hierarchical decision tree algorithms on Sentinel-1/2 images. *Remote Sensing of Environment*, *297*, p.113793.

14. In lines 381-382 the confusion matrix shows that the PCG class has a UA of 84.18% (FN = 3 243), yet no discussion of the false-negative causes is provided. Please analyze the reasons for these missed detections.

Response:

We sincerely thank you for the valuable comments. We apologize for the misunderstanding caused by the representation of the confusion matrix in the preprint version. The issue stemmed from an unclear explanation of the test sample construction. In the original dataset, 20,500 test samples were independently selected for each of the PCG and Non-PCG classes based on reference labels. However, in the preprint version, the confusion matrix failed to correctly reflect this setup due to misplacement of the values "2" and "3243", resulting in incorrect sample counts under each reference class.

The corrected confusion matrix is shown below.

Table 1. Confusion Matrix

Confusion Matrix	**Reference:** Non-PCG	**Reference:** PCG	*UA (%)*
Predicted: Non-PCG	20,498	3,243	86.34
Predicted: PCG	2	17,257	99.99
PA (%)	99.99	84.18	
F1-score (%)	-	91.41	
OA (%)			92.08

Note: PA, Producer's Accuracy; UA, User's Accuracy; OA, Overall Accuracy.*

Although there was a mistake in the early version of the confusion matrix, the accuracy metrics presented in Table 2 were correctly calculated from the actual test data. In the revised manuscript, this table has been removed according to Reviewer #1's suggestion.

In addition, to further assess the reliability of the Global-PCG-10 dataset from a quantitative perspective, we constructed an independent test sample set following the stratified random sampling strategy recommended by Olofsson et al. (2014), and consistent with practices from studies such as Wang et al. (2023) and Tian et al. (2025). Samples were drawn in proportion to the actual mapped area of PCG and non-PCG across the classification region.

The newly reconstructed confusion matrix is presented below, where "±" denotes the 95% confidence interval for each metric:

Table 1. Confusion Matrix

Confusion Matrix	**Reference:** Non-PCG Reference	**Reference:** PCG	*UA (%)*
Predicted: Non-PCG	39,991	893	97.82 ± 0.13
Predicted: PCG	9	5,107	99.82 ± 0.11
PA (%)	99.98 ± 0.01	85.12 ± 0.90	
F1-score (%)	-	91.88 ± 2.71	
OA (%)			98.04 ± 0.12

Note: PA, Producer's Accuracy; UA, User's Accuracy; OA, Overall Accuracy.*

In this result, FN = 893 and FP = 9. We have addressed the issue of false positives (FP = 9) in Response #12. The relatively high number of false negatives (FN = 893) can be attributed to the following factors. (1) Omission of small-scale PCG targets. Due to the 10-meter spatial resolution of Sentinel-2 imagery, which is significantly lower than that of high-resolution platforms like PlanetScope, small PCG often occupies only a few to a dozen pixels and are easily affected by mixed pixel issues. This makes it difficult for the model to extract reliable spectral features and leads to missed detections. (2) Limitations in spatiotemporal coverage of imagery. The Sentinel-2 data used in this study were organized by 1° grid tiles. Due to cloud contamination and observation scheduling constraints, it is sometimes challenging to obtain cloud-free imagery for both time periods (spring and summer), which reduces the model's ability to detect PCG in certain regions. (3) Post-classification filtering effects. To reduce false positives, we applied a strict post-processing procedure to the initial classification results when generating the Global-PCG-10 dataset. Specifically, a multi-stage Sieve Filter was used to remove small patches and isolated noise, which effectively suppressed misclassifications and significantly improved the precision (UA) for the PCG class.

Further analysis of the false negative (FN) cases, including field validation and image comparisons, will be presented in Section 4.5.2 Bad Case Analysis of the revised manuscript.

The above contents have been added in Section 4.2 "Reliability of Global-PCG-10" of the manuscript (**see *Lines 445 ~ 469* for details**).

References

Olofsson, P., Foody, G.M., Stehman, S.V. and Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote sensing of environment*, *129*, pp.122-131.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. and Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote sensing of Environment*, *148*, pp.42-57.

Tian, F., Wu, B., Zeng, H., Zhang, M., Zhu, W., Yan, N., Lu, Y. and Li, Y., 2025. GMIE: a global maximum irrigation extent and central pivot irrigation system dataset derived via irrigation performance during drought stress and deep learning methods. *Earth System Science Data*, *17*(3), pp.855-880.

Wang, M., Mao, D., Wang, Y., Xiao, X., Xiang, H., Feng, K., Luo, L., Jia, M., Song, K. and Wang, Z., 2023. Wetland mapping in East Asia by two-stage object-based Random Forest and hierarchical decision tree algorithms on Sentinel-1/2 images. *Remote Sensing of Environment*, *297*, p.113793.

15. In lines 381-382, non-PCG exhibits UA = 99.99% and PA = 86.30%, while PCG shows UA = 84.18% and PA = 99.99%, indicating a strong asymmetry in false-positive and false-negative distributions. Please discuss the origin of this bias.

Response:

In the reconstructed confusion matrix, the classification bias for the Non-PCG class has been substantially mitigated. However, the PCG class still exhibits an asymmetry in accuracy, i.e., its user's accuracy (UA) reaches as high as 99.99%, while the producer's accuracy (PA) remains relatively low at 84.18%. This discrepancy reflects an imbalance between false positives and false negatives in the model's predictions.

We believe this asymmetry is primarily caused by two factors. Firstly, under the 10-meter spatial resolution of Sentinel-2 imagery, many PCG areas consist of only a very limited number of pixels. In particular, small-scale PCG often occupies just a few, or a dozen mixed pixels. This makes it challenging for the model to extract robust spatial-spectral features, leading to frequent omissions and significantly lowering the PA for the PCG class. In Section 4.5.2, we further explore this issue through bad case analysis, where we provide visual examples of small PCGs being omitted. Secondly, to enhance the reliability of the final map product, we applied multiple post-processing steps to the Global-PCG-10 predictions, including a Sieve Filter to remove isolated or marginal misclassified regions. While this procedure effectively reduced false positives, thus improving the UA, it also suppressed small or fragmented PCG predictions, indirectly contributing to the lower PA.

In summary, this asymmetric accuracy reveals a systematic bias in the model's performance when

dealing with extremely rare, spatially small, and spectrally mixed targets such as PCG, rather than a result of random error. We elaborate on these issues in detail in the bad case analysis section (***Lines 459 ~ 484***) of the revised manuscript.

16. In section 4.3.2, authors compare visually against Tong et al. (2024) using 3 m data (Figure 12) but lack objective quantitative metrics. Please add quantitative evaluation and report specific comparative values.

Response:

Thank you for your constructive comment regarding the lack of objective quantitative metrics in Section 4.3.2. We have revised this section accordingly and taken the following measures to address your concern:

1. Accuracy assessment using 46,000 high-confidence test samples for Global-PCG-10

In this study, we conducted a quantitative evaluation of the Global-PCG-10 dataset using 46,000 high-confidence test samples, including 6,000 positive samples (PCG) and 40,000 negative samples (non-PCG). These samples were manually interpreted and verified to ensure high confidence. Specifically, the PCG samples were initially derived from the 2019 global 3-meter PCG map published by Tong et al. (2024), and were further validated using historical high-resolution imagery from Google Earth. However, upon comparing this validation sample set against both our Sentinel-2-based (10 m resolution) product and the 3-meter reference product (Tong et al., 2024), we observed that the 10 m product exhibited certain limitations in detecting small-scale PCG structures. This limitation is particularly pronounced for PCG areas that span only a few pixels and are often affected by mixed-pixel issues. This challenge likely contributes to the relatively lower user's accuracy (UA) for the PCG class in the confusion matrix results. To explain this phenomenon, we have added detailed quantitative analysis and bad case discussions in the revised Section 4.3.2.

2. Independent validation in high-density PCG regions and at the global scale

To provide a more objective and fair comparison, we followed the methodology proposed by Huang et al. (2022) and conducted a quantitative consistency analysis between the two datasets in terms of global PCG spatial distribution. Specifically, we selected four representative 1° × 1° grid regions with varying PCG densities. Each of these grids was further subdivided into multiple 0.01° × 0.01° sub-grid units. Within each sub-grid, we calculated the proportion of PCG pixels relative to the total number of pixels for both datasets (i.e., PCG area ratio, ranging from 0 to 1). Using these continuous ratio-based data, we applied linear regression analysis to calculate the coefficient of determination (R^2), thereby quantifying the spatial distribution consistency between the two datasets across different regions. Unlike methods that rely on discrete classification labels, this approach leverages continuous area proportions, making it more

suitable for evaluating agreement between remote sensing datasets with differing spatial resolutions. As shown in Figure 14a ~ d, the experimental results in four typical study area indicate that, in high-density PCG regions, our 10-meter resolution PCG dataset demonstrates a high degree of spatial consistency with the 3-meter reference dataset.

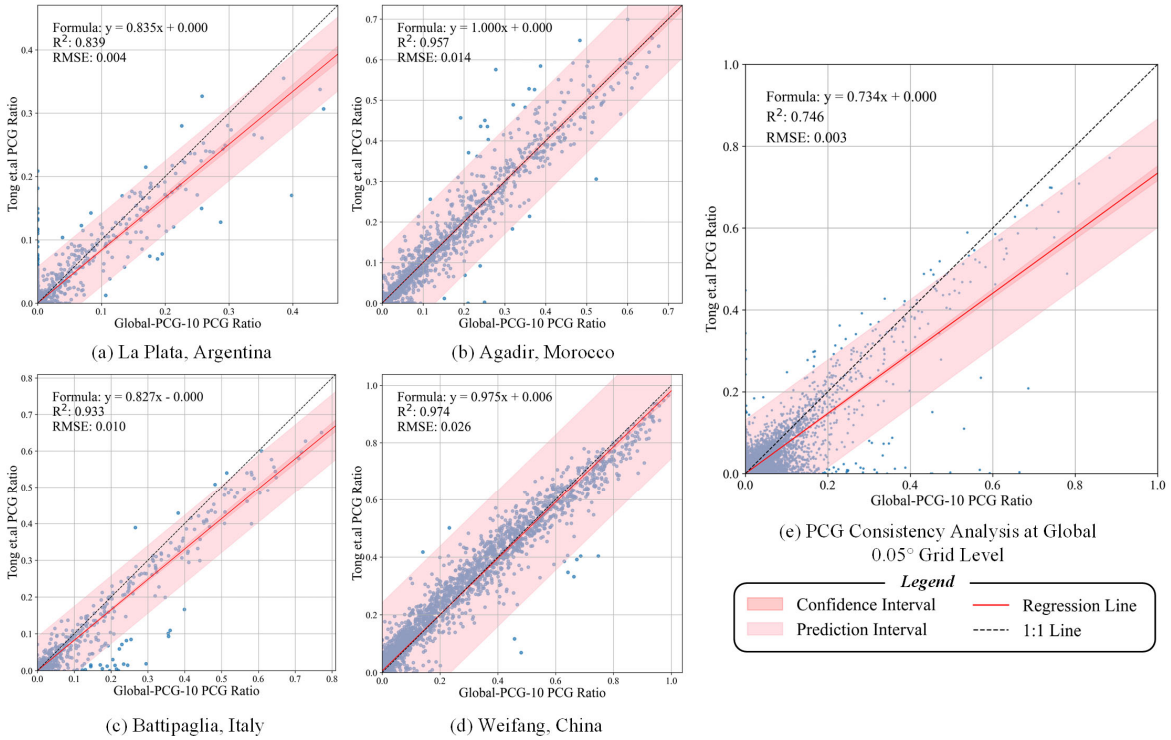

Figure 14. The consistency performance across the four representative regions and between the dataset by Tong et al. and Global-PCG-10 in representative regions.

To further evaluate spatial consistency at the global scale, we applied a standard regression-based consistency analysis across the entire globe, with reference to the analytical approach and spatial resolution (i.e., 0.05° grid) used by Huang et al, (2022). The coefficient of determination (R^2) was again employed as the primary evaluation metric. As shown in Figure 14e, the comparison based on a 0.05° grid reveals strong agreement in the global spatial distribution of PCG between the dataset published by Tong et al. (2024) and the Global-PCG-10 dataset. The regression analysis yields an R^2 of 0.746, a root mean square error (RMSE) of 0.003, and a regression equation of $y = 0.734x + 0.000$. These results indicate a moderate to strong spatial correlation between the two datasets, further validating the effectiveness of the Global-PCG-10 dataset in capturing the global distribution pattern of PCG.

As illustrated in Figure 14, the Global-PCG-10 dataset exhibits strong agreement with the reference data in typical regions (Figure 14a–d), whereas a moderate overestimation trend is observed at the global scale. This discrepancy may be attributed to the spatial resolution limitations of Sentinel-2 imagery. As a medium-resolution satellite (10m), Sentinel-2 is more susceptible to intra-class spectral variability and inter-class spectral confusion. In sparely distributed greenhouse areas, non-PCG features such as bare soil,

inter-greenhouse roads, or adjacent agricultural structures may exhibit spectral signatures similar to plastic-covered greenhouses, leading to misclassification and systematic overestimation of PCG coverage. Moreover, within the same spatial aggregation unit (e.g., a 0.05° grid cell), Sentinel-2 offers fewer pixels compared to PlanetScope (3m), making PCG area statistics more sensitive to per-pixel classification errors. Consequently, in typical regions with more homogeneous greenhouse patterns, clearer boundaries, the classification results are more stable and consistent. In contrast, at the global scale, the combined effects of landscape heterogeneity and resolution-induced error propagation contribute to reduced agreement.

The above contents have been added in Section 4.2 "Reliability of Global-PCG-10" of the manuscript (**see *Lines 551 ~ 572* for details**).

Reference:

Huang, X., Yang, J., Wang, W. and Liu, Z., 2022. Mapping 10-m global impervious surface area (GISA-10m) using multi-source geospatial data. *Earth System Science Data Discussions*, *2022*, pp.1-39.

17. Recommend including a discussion of potential applications and the limitations of the produced dataset.

Response:

We sincerely thank you for the valuable suggestion. We fully agree with your comment that further discussion is needed regarding the application potential and limitations of the proposed dataset to better illustrate its scientific value and boundaries of use.

Accordingly, we have revised and expanded Section 4.4 in Chapter 4 of the manuscript to include the following subsection.

4.4 Application potential and limitations of the dataset

As described above, Global-PCG-10 is a global-scale dataset of PCG derived from open-access Sentinel-2 imagery. By leveraging freely available satellite data, the dataset significantly reduces production costs while providing a standardized and well-structured data format that can be easily integrated with other open-source remote sensing products.

As the first global PCG dataset with 10-meter spatial resolution, Global-PCG-10 has strong application potential in various domains. (1) In agricultural monitoring and statistics, the dataset reveals the spatial distribution pattern of global protected agriculture, offering valuable support for agricultural structure optimization, farmland use monitoring and irrigation estimation. (2) In agro-environmental assessments, it provides high-resolution spatial information on protected agriculture, supporting efforts

by governments and international organizations to conduct agricultural censuses, develop regional agricultural strategies and implement climate-adaptive agricultural policies. (3) In open-source land use/land cover (LULC) applications, PCG are often underrepresented in current global LULC products. This dataset helps fill that gap by explicitly including PCG as a key cropland subtype.

Despite its usefulness, Global-PCG-10 still has several limitations that need to be addressed in future work. Firstly, due to the 10-meter resolution of Sentinel-2 imagery, it remains difficult to detect small-scale or scattered PCG units, especially in regions dominated by smallholder agriculture. This may lead to omission errors. In the future, we plan to integrate higher-resolution remote sensing data to develop regional PCG datasets with finer spatial detail. Secondly, the classification task in this study focused primarily on the overall category of PCG, without further distinguishing among its subtypes. In future research, we plan to explore fine-grained classification methods for agricultural greenhouses (AG), including the differentiation of daylight greenhouses, conventional plastic greenhouses and small arch sheds, in order to further enhance the accuracy and practical applicability of PCG dataset. Thirdly, as the dataset only contains PCG in 2020, it does not capture dynamic PCG changes such as recent expansion or degradation regions. We plan to extend this work to develop a global time-series dataset of greenhouses, enabling long-term monitoring and trend analysis. Besides, the current pipeline for PCG mapping, which combines deep learning and active learning, still relies on a semi-automated weak-label updating strategy and does not yet support full end-to-end automation. In the future, we aim to explore end-to-end weak-label learning frameworks to build a more efficient and automated data processing system.

The above contents have been added in Section 4.2 “Application potential and limitations of the dataset” of the manuscript (**see *Lines 573 ~ 597* for details**).

Thank you again for your comments. They are valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our studies.

Yours sincerely,

Bowen Niu, Quanlong Feng

on behalf of all the co-authors

Responses to CC #1

General Comments:

This article used deep learning methods to produce a refined distribution of global Plastic-covered greenhouse in 2020. The methodology of the paper is sound and the data is novel. Although there are some global PCG data in the current study, the data set has some value in terms of resolution. There is some value for global agricultural macro-management.

Specific Comments:

There are some problems and deficiencies that need to be corrected as follows:

1. Plastic-covered greenhouse can divided into daylight greenhouses and plastic greenhouses. Is there a significant difference in the spectra of the two kinds.

Response:

Thank you for your thoughtful question. To address this, we selected Weifang in Shandong Province as a representative region and analyzed the seasonal spectral reflectance differences between Daylight Agricultural Greenhouses (sunlight greenhouses) and Normal Agricultural Greenhouses (conventional greenhouses). A comparative reflectance plot illustrating these seasonal differences is provided below.

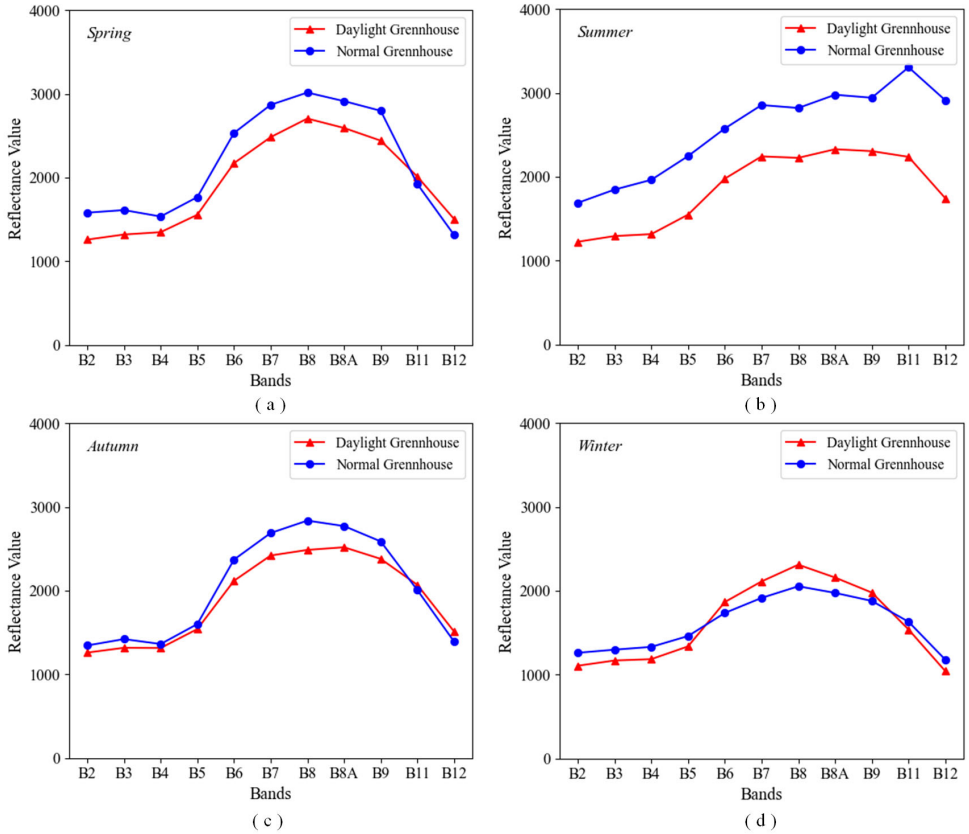

Figure CC1-1. Comparison of Spectral Reflectance Between Two Types of Greenhouses

As shown in the figure, the two types of agricultural greenhouses exhibit noticeable spectral differences, particularly during the summer months (June to September), when the distinction is most pronounced. Therefore, summer-season remote sensing imagery is essential for distinguishing between Daylight Agricultural Greenhouses and Normal Agricultural Greenhouses. In addition, the spectral differences are also relatively evident in spring, making spring imagery the second best alternative when summer data are unavailable. In contrast, the spectral differences during autumn and winter are relatively small, making it difficult to effectively differentiate between the two types of greenhouses during those seasons. These findings indicate that while Daylight and Normal greenhouses can be spectrally distinguished, this separability is season-dependent and varies significantly across different times of the year.

In future studies, we will further explore the fine-scale classification of agricultural greenhouses (AG), including distinguishing between daylight greenhouses, conventional plastic greenhouses, glass greenhouses and small arch sheds to improve the accuracy and application value of PCG dataset.

2. Line 150-155, authors generated initial PCGs labels via RFM and GEE. Did the authors use some manually labeled data to validate the results?

Response:

Thank you for your good question. Actually, we employed multiple strategies to ensure the accuracy of initial PCG labels during the training sample collection process. Firstly, field investigations were conducted in typical greenhouse distribution regions in China, such as Weifang (Shandong), Kunming (Yunnan) and Lishu (Jilin), where we consulted with local farmers to confirm the types and spatial distribution of greenhouses. Considering that agricultural PCG generally has a long lifespan and stable structures (Ou et al., 2020), we conducted systematic manual visual interpretation of high-resolution historical imagery from Google Earth across various regions, both in China and abroad, to obtain high-confidence samples. For international regions, we supplemented the sampling process with literature review, meta-analysis and information from online sources, followed by manual verification using Google Earth. All manually interpreted samples were subsequently cross-validated using Sentinel-2 imagery to confirm their presence in the year 2020.

Based on these manually annotated samples, we constructed the training dataset with the Google Earth Engine (GEE) platform. The samples were split into training and validation subsets using standard practices widely adopted in the LULC (land use/land cover) domain. We then used these datasets to evaluate the initial classification results generated by the Random Forest model in GEE through confusion matrix analysis. Therefore, both the training and validation samples were derived from manual interpretation, ensuring the accuracy and reliability of the labeling process.

3. Line 100-105, this paper mentioned a case which used commercial satellite data to extract the PCGs. What's its resolution? And what's different from that research? The authors point that the scientific problem is how to use open access data to generate PCGs map. I think this was not the true scientific problem.

Response:

Thank you for your insightful comment. Regarding the study you mentioned that uses commercial satellite data to extract PCG, we have addressed this in the manuscript **(see Line 99 ~ 100)**. That study employed PlanetScope imagery with a spatial resolution of 3 meters, which provides stronger texture representation due to its higher spatial detail. However, commercial satellite data are typically expensive, subject to access restrictions, and often lack globally consistent coverage within a single time period. Therefore, we argue that a key scientific challenge in the current context is to develop a low-cost, scalable and replicable methodology for global PCG mapping based on open and accessible data sources.

We also observed from the dataset released by Tong et al. that their 2019 global PCG product incorporates imagery from 2021 to supplement the 2019 data. While this approach may be feasible for greenhouse identification tasks, it inevitably introduces temporal inconsistency. In contrast, our study utilizes multi-temporal Sentinel-2 imagery strictly from the year 2020, which allows for temporally consistent large-scale mapping and reduces potential errors caused by temporal mismatches.

Moreover, compared to manual annotation based on 3-meter resolution imagery, our approach enables automated large-scale PCG extraction using 10-meter resolution data, significantly reducing labor costs. The grid-based data organization strategy adopted in this study also improves data processing efficiency.

In summary, we believe that our research not only proposes a practically applicable method but also addresses the scientific challenges of global-scale PCG mapping under the constraint of open-access remote sensing data. The proposed approach provides a reliable technical framework for future studies with similar objectives.

4. Figure 4 illustrated the multiple-temporal NDVI profile of bareland, PCGs and PMFs in a representative sub-region of Gansu Province, China. There are PMFs in some southern area. Farmer used plastic much earlier than the northern area, which would induce the difference of multiple-temporal NDVI.

Response:

We sincerely thanks for your detailed observation. We agree with your point that in Southern China, the mulching period of PMF (Plastic-Mulched Farmland) typically begins earlier than in northern regions, and this temporal difference is indeed reflected in the fluctuation patterns of the NDVI time series. Based

on our constructed 2020 PMF distribution dataset of China (Niu et al., 2025), we observed that the spatial extent of PMF in Southern China is relatively limited. However, agricultural systems in this region are mainly characterized by a double-cropping pattern, which results in shorter cultivation cycles and more rapid vegetation changes. In contrast, PCG exhibit more stable and long-term coverage characteristics, leading to more consistent and regular NDVI patterns. It is precisely this marked difference in the multi-temporal NDVI curves that facilitates the effective differentiation between PMF and PCG in our classification model.

As illustrated in **Figure CC1-2**, for example, in the case of Shandong Province, Ou et al. (2021) constructed NDVI curves for various land cover types using Landsat-8 imagery. The cropping system in this region follows a double-cropping schedule, which is comparable to that in most parts of Southern China. Their data clearly show significant spectral differences between PCG, cropland and PMF during two key temporal windows, from day 73 to 136, and from day 165 to 248. Consequently, their study also focused on spring and summer imagery when analyzing regions with a double-cropping system.

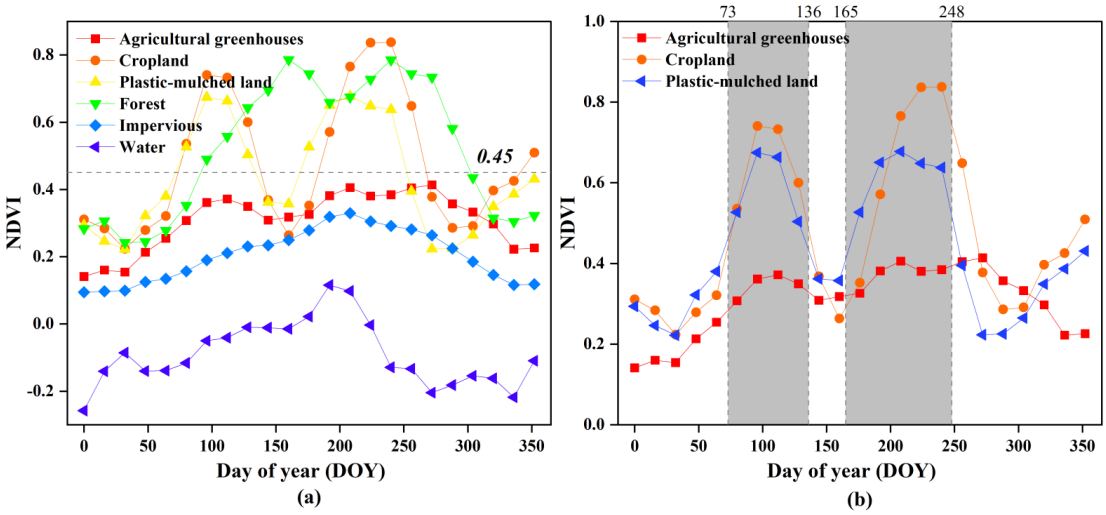

Figure 6. (**a**,**b**) Mapping window selection based on vegetation growth.

Figure CC1-2. Time-Series NDVI of Different Land Cover Types

In addition, we conducted sampling of PCG and cropland in the Yangtze River Delta region of Eastern China, and constructed NDVI time series curves for both land cover types in 2020 based on Sentinel-2 imagery. The results in Figure **CC1-3** indicate that although the period of separability between PCG and cropland may occur slightly earlier in this region compared to Northern China, the two classes still exhibit clear separability during the key seasonal windows of spring (March–May) and summer (June–September). This pattern further confirms the rationality and applicability of using spring and summer imagery in our classification model.

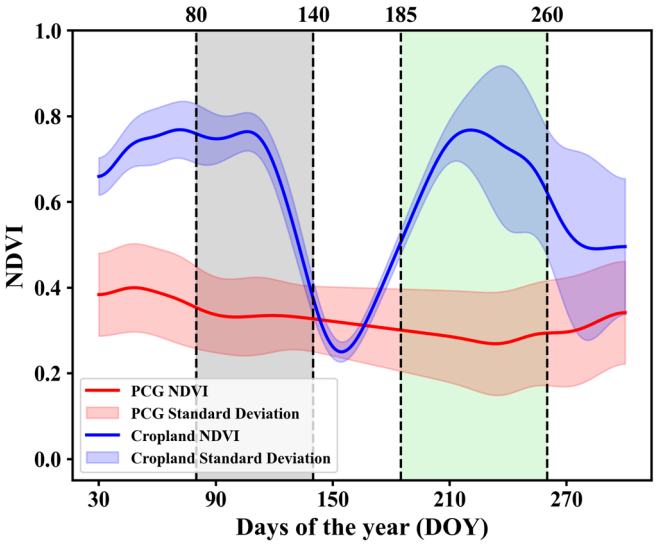

Figure. CC1-3. Time-Series NDVI of PCGs and Cropland

In fact, PMF (Plastic-Mulched Farmland) is essentially a form of cropland, and its overall phenological pattern does not change significantly due to the use of plastic mulch. Therefore, using spring and summer remote sensing imagery is an effective strategy to enhance the separability between PMF and PCG, as well as between general cropland and PCG. This choice is also well supported by our previous research findings and practical experience.

Reference:

Niu, B., Feng, Q., Zhang, X., Qiu, B., Zeng, Y., Su, S., Xu G., Gong J., Yan X., Huang J., Yin G., Liu J., Yang J., Zhu D. China-PMF-10: a 10-m national map of plastic-mulched farmlands in China of 2020 using deep semantic segmentation. 2025. figshare. Dataset.

5. Line 420-425, the PCGs dataset from the university of Copenhagen is the year of 2019. Differences between the two data years may result in discrepancies. Of course, it would be better if the data include a few more years.

Response:

Thank you for your valuable suggestion. We fully acknowledge that the PCG dataset provided by the University of Copenhagen corresponds to the year 2019, while the target year of our study is 2020, and thus there is indeed a temporal discrepancy between the two. To minimize the potential impact of this difference, all manually labeled samples and model validation efforts in our study were strictly based on Sentinel-2 imagery and historical Google Earth images from 2020, ensuring consistency between the data used and the target year of analysis.

Moreover, considering that PCG typically have long usage cycles and relatively stable spatial

distributions over short periods, we believe that the 2019 dataset still holds substantial reference value in terms of spatial distribution and can serve as a useful auxiliary resource. In future studies, we plan to incorporate multi-year remote sensing data to further enhance the temporal adaptability and robustness of our model, thereby supporting long-term agricultural monitoring and change detection.

Thank you again for your comments. They are valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our studies.

Yours sincerely,
Bowen Niu, Quanlong Feng
on behalf of all the co-authors

Responses to CC #2

General Comments:

This manuscript developed a novel approach for producing the global 10-meter Plastic-covered greenhouse (PCGs) dataset for 2020 year. This approach combines the active learning strategy and the deep learning model, so as to let the model to learn more robust PCG features by optimizing weak labels. The results of the PCGs classification has been compared with publicly released land use and land cover (LULC) datasets, and showed the high accuracy. This PCGs dataset can characterize the global spatial distribution of plastic-mulched coverage in 2020. This manuscript is novel in topic selection, innovative in technical approach, and solves the problem of the lack of global-scale plastic-mulched products, which is of high scientific significance and practical value.

Specific Comments:

The following issues still need to be revised.

1. Do the samples include actual ground-collected sample points? The manuscript has examples from Gansu Province in China, are there similar ground samples from other areas?

Response:

Thank you for your good comments. Actually, as for actual ground-collected sample points, we conducted field surveys in typical greenhouse-concentrated regions of China, such as Weifang (Shandong), Kunming (Yunnan) and Lishu (Jilin) during the construction of training samples for the GEE-based Random Forest (RF) classification. These surveys involved on-site investigations and interviews with local farmers to confirm the distribution and types of PCG.

Given that PCG typically has a lifespan of around 10 years or longer and exhibit high structural stability (Ou et al., 2021), we further performed systematic visual interpretation of high-resolution historical imagery from Google Earth to obtain high-confidence samples across multiple regions globally. For non-Chinese regions, the identification process was supported by literature review, meta-analysis and online resources. All collected samples were cross-validated using Sentinel-2 imagery to ensure their actual presence in the year 2020. Additionally, we refined PCG and non-PCG labels within each grid based on classification outputs to ensure labeling accuracy.

As for Figure 4, a region from Gansu was selected as a representative area because it features a large number of both plastic-covered greenhouses (PCG) and plastic-mulched farmland (PMF), which often coexist and are spatially interwoven. The figure presents differences in multi-temporal NDVI curves among PCG, PMF and bare land, illustrating how multi-temporal features can effectively distinguish between these easily confusable classes. This method is also applicable to other regions, particularly in

China, the world's largest user of plastic mulch films, where multi-temporal imagery proves effective in resolving confusion between PCG and PMF.

2. From table 1. we can see that the UA of Non-PCG is as high as 99.99%, and the that of PCG is 84.18%. Do all these sample points for accuracy evaluation also come from GEE automatically selected or manually decoded? Is there a relationship between the high accuracy of Non-PCG and the sample selection?

Response:

Thanks for your advices. Actually, the current description regarding the sampling strategy, sample proportion design and the reliability of the test samples in the confusion matrix was insufficient. We have revised this section based on the reconstructed confusion matrix and now provide a detailed explanation of the test sample collection process. The specific modifications are as follows:

To further quantitatively evaluate the reliability of the Global-PCG-10 dataset, we constructed a dedicated test sample set**.** The spatial distribution of these test samples is shown in Figure 10.

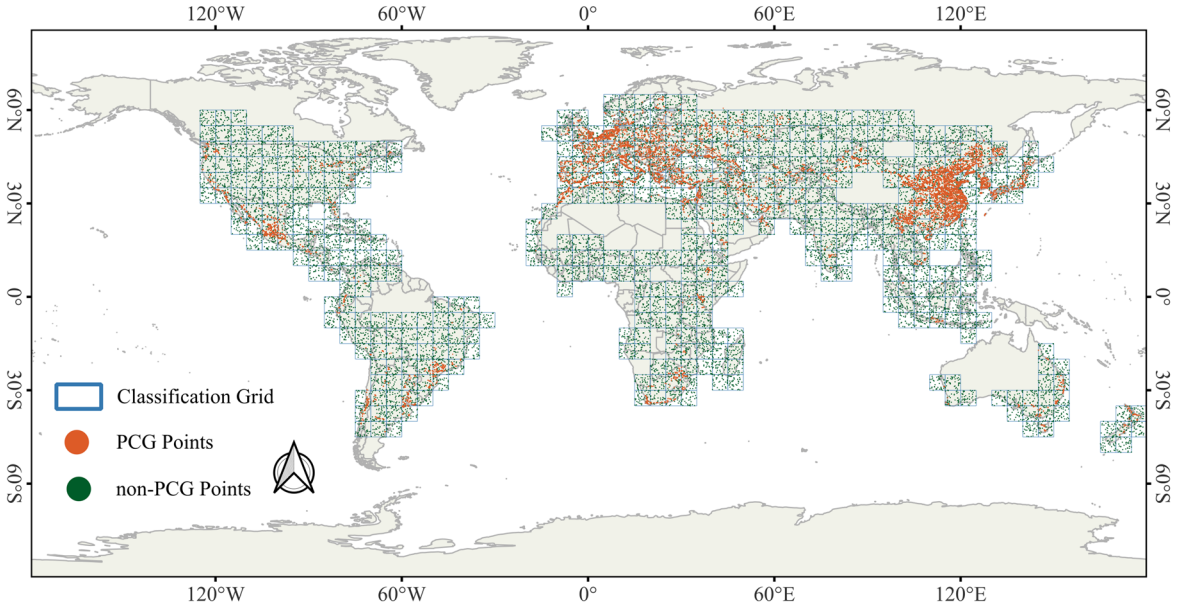

Figure 10. Spatial distribution of global test samples.

The dataset includes two categories of PCG and non-PCG. Based on previous research practices (Olofsson et al., 2013, 2014; Tian et al., 2025; Wang et al., 2023), we followed the stratified random sampling strategy recommended by Olofsson et al. (2014), in which samples are drawn in proportion to the mapped area of each class within the actual mapping region. However, since the global coverage of PCG is less than 1%, strictly proportional sampling would result in very few PCG samples to support a statistically robust accuracy assessment. To address this issue, and consistent with the approaches adopted in the above studies, we moderately increased the proportion of PCG samples in the validation set to

approximately 10%. This adjustment significantly enhances the evaluation capability for this minority class.

Table 1. Confusion matrix.

Confusion Matrix	**Reference**: Non-PCG	**Reference**: PCG	*UA (%)*
Predicted: Non-PCG	39,991	893	97.82 ± 0.13
Predicted: PCG	9	5,107	99.82 ± 0.11
PA (%)	99.98 ± 0.01	85.12 ± 0.90	
F1-score (%)	-	91.88 ± 2.71	
OA (%)			98.04 ± 0.12

Note: PA, Producer's Accuracy; UA, User's Accuracy; OA, Overall Accuracy.*

As shown in Table 1, the total number of test samples is 46,000, with 6,000 PCG samples and 40,000 non-PCG samples. To ensure the validity of both classes, we applied separate sampling strategies for each. As for PCG, test samples were derived from the 2019 global 3-meter PCG dataset developed by Tong et al. (2024), and manually verified through Google Earth visual interpretation. Since the Global-PCG-10 dataset is for the year 2020, and considering that PCG typically have long lifespans and stable structures, this 2019 dataset provides a reliable reference. Additionally, we performed a second round of verification using historical Google Earth imagery in around 2020 to confirm their existence and status, minimizing sampling bias from prior knowledge. And for non-PCG, due to the large quantity required, manual sampling was impractical. We thus randomly sampled from the GLC_FCS30D dataset to ensure independence and randomness. All samples were also verified through visual interpretation of historical Google Earth imagery in around 2020 to ensure label correctness.

Based on this validation dataset, as shown in Table 1, Global-PCG-10 achieved a PA of 85.12% ± 0.90%, UA of 99.82% ± 0.11%, F1-score of 91.88% ± 2.71% and an overall accuracy of 98.04% ± 0.12%. These results indicate that the recall (PA) for PCG is relatively low, which is likely due to omission errors of small-scale PCG instances. This issue is further analyzed in the bad **case study presented in Section 4.3**. In contrast, the model demonstrates a very high precision (UA), primarily attributable to a series of post-processing operations applied to the preliminary predictions of Global-PCG-10. Among these, the most critical step was the use of a Sieve Filter, which was implemented in multiple stages to effectively remove a large number of misclassified areas.

The above contents have been added in Section 4.2 "Reliability of Global-PCG-10" of the manuscript (**see *Lines 423 ~ 455* for details**).

References

Olofsson, P., Foody, G.M., Stehman, S.V. and Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote sensing of environment*, *129*, pp.122-131.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. and Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote sensing of Environment*, *148*, pp.42-57.

Tian, F., Wu, B., Zeng, H., Zhang, M., Zhu, W., Yan, N., Lu, Y. and Li, Y., 2025. GMIE: a global maximum irrigation extent and central pivot irrigation system dataset derived via irrigation performance during drought stress and deep learning methods. *Earth System Science Data*, *17*(3), pp.855-880.

Wang, M., Mao, D., Wang, Y., Xiao, X., Xiang, H., Feng, K., Luo, L., Jia, M., Song, K. and Wang, Z., 2023. Wetland mapping in East Asia by two-stage object-based Random Forest and hierarchical decision tree algorithms on Sentinel-1/2 images. *Remote Sensing of Environment*, *297*, p.113793.

3. In the section of '4.3 Comparison with other studies', the authors performed a comparison of the results of the spatial extraction of PCG in different regions of the globe. In addition, it is advisable to compare the total area of PCG by different continents (or regions). For example, the total areas of PCG in a region acquired by different data products, or compare the proportion of area where PCG overlaps in the same region by different products. This part of the study needs to be deepened in terms of difficulty, and merely comparing spatially with the LULC products or similar PCG products does not seem to be enough to prove the accuracy of this product. It is also recommended to add some statistical yearbooks or public data information for comparison.

Response:

To provide a more objective and fair comparison, we followed the methodology proposed by Huang et al. (2022) and conducted a quantitative consistency analysis between the two datasets in terms of global PCG spatial distribution. Specifically, we selected four representative 1° × 1° grid regions with varying PCG densities. Each of these grids was further subdivided into multiple 0.01° × 0.01° sub-grid units. Within each sub-grid, we calculated the proportion of PCG pixels relative to the total number of pixels for both datasets (i.e., PCG area ratio, ranging from 0 to 1). Using these continuous ratio-based data, we applied linear regression analysis to calculate the coefficient of determination (R^2), thereby quantifying the spatial distribution consistency between the two datasets across different regions. Unlike methods that rely on discrete classification labels, this approach leverages continuous area proportions, making it more suitable for evaluating agreement between remote sensing datasets with differing spatial resolutions. As shown in Figure 14a ~ d, the experimental results in four typical study area indicate that, in high-density PCG regions, our 10-meter resolution PCG dataset demonstrates a high degree of spatial consistency with

the 3-meter reference dataset.

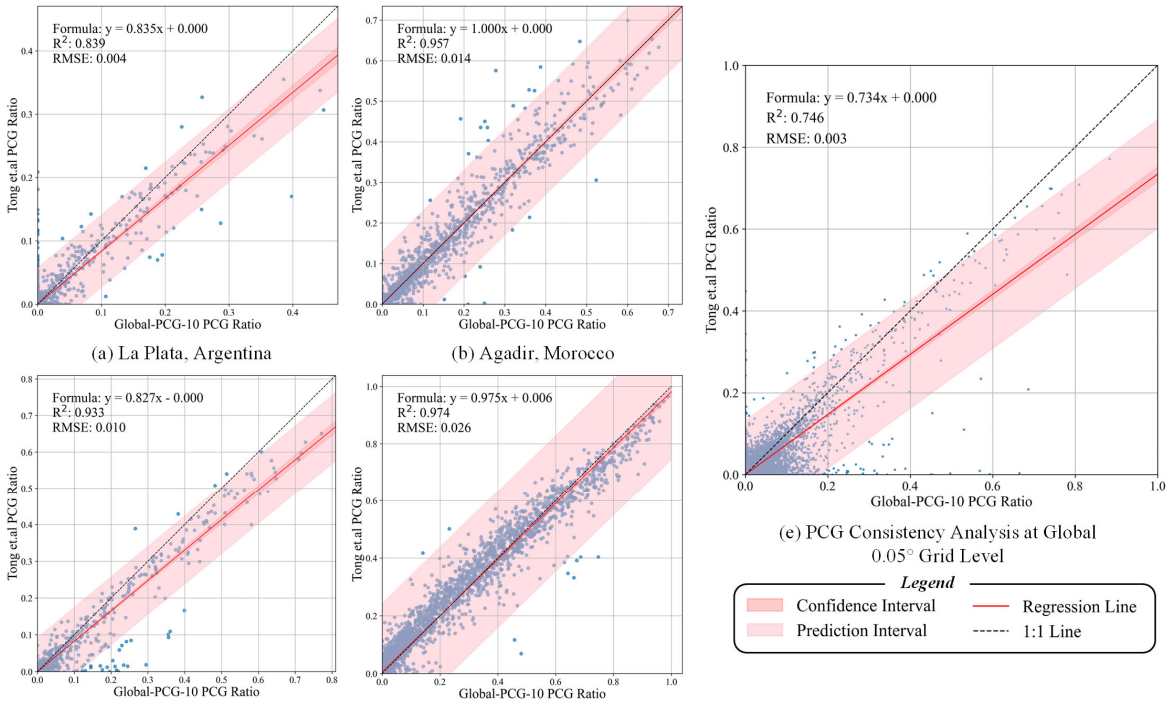

Figure 14. The consistency performance across the four representative regions and between the dataset by Tong et al. and Global-PCG-10 in representative regions.

To further evaluate spatial consistency at the global scale, we applied a standard regression-based consistency analysis across the entire study area, with reference to the analytical approach and spatial resolution (i.e., 0.05° grid) used by Huang et al, (2022). The coefficient of determination (R^2) was again employed as the primary evaluation metric. As shown in Figure 14e, the comparison based on a 0.05° grid reveals strong agreement in the global spatial distribution of PCG between the dataset published by Tong et al. (2024) and the Global-PCG-10 dataset. The regression analysis yields an R^2 of 0.746, a root mean square error (RMSE) of 0.003, and a regression equation of y = 0.734x + 0.000. These results indicate a moderate to strong spatial correlation between the two datasets, further validating the effectiveness of the Global-PCG-10 dataset in capturing the global distribution pattern of PCG.

As illustrated in Figure 14, the Global-PCG-10 dataset exhibits strong agreement with the reference data in typical regions (Figure 14a–d), whereas a moderate overestimation trend is observed at the global scale. This discrepancy may be attributed to the spatial resolution limitations of Sentinel-2 imagery. As a medium-resolution satellite (10m), Sentinel-2 is more susceptible to intra-class spectral variability and inter-class spectral confusion. In sparely distributed greenhouse areas, non-PCG features such as bare soil, inter-greenhouse roads, or adjacent agricultural structures may exhibit spectral signatures similar to plastic-covered greenhouses, leading to misclassification and systematic overestimation of PCG coverage. Moreover, within the same spatial aggregation unit (e.g., a 0.05° grid cell), Sentinel-2 offers fewer pixels

compared to PlanetScope (3m), making PCG area statistics more sensitive to per-pixel classification errors. Consequently, in typical regions with more homogeneous greenhouse patterns, clearer boundaries, the classification results are more stable and consistent. In contrast, at the global scale, the combined effects of landscape heterogeneity and resolution-induced error propagation contribute to reduced agreement.

The above contents have been added in Section 4.2 “Reliability of Global-PCG-10” of the manuscript (**see *Lines 551 ~ 572* for details**).

Reference:

Huang, X., Yang, J., Wang, W. and Liu, Z., 2022. Mapping 10-m global impervious surface area (GISA-10m) using multi-source geospatial data. *Earth System Science Data Discussions*, *2022*, pp.1-39.

Thank you again for your comments. They are valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our studies.

Yours sincerely,

Bowen Niu, Quanlong Feng

on behalf of all the co-authors