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.

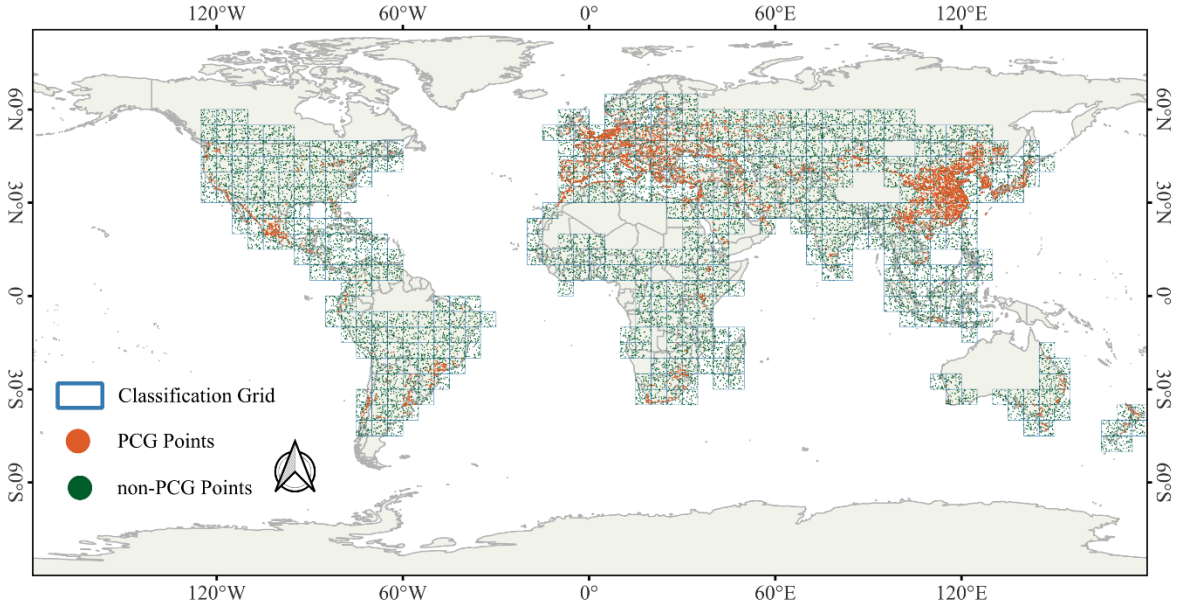

Figure 10. Spatial distribution of global test samples.

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.

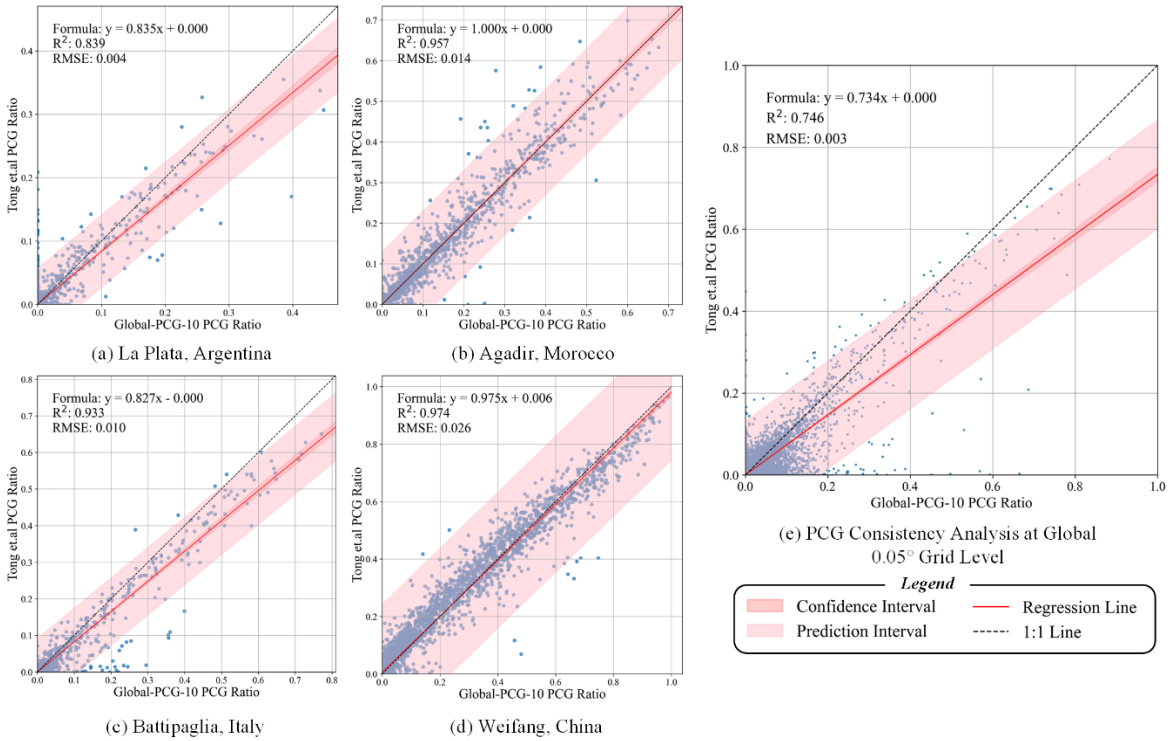

Figure 14. The consistency performance across the four representative regions and between the dataset by Tong et al. and Global-PCG-10 in representative regions.

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