Response to the reviewers (#essd-2025-695)

Dear Editor,

We sincerely appreciate the valuable and constructive comments you have provided. In response, we have conducted a comprehensive and thorough revision of the manuscript to address all the comments and suggestions. We have responded to each point in detail to ensure that all concerns have been fully addressed. Your insightful feedback has significantly improved the overall quality of this manuscript. For ease of review, the original comments are presented in *italics*, our responses are provided in regular font, and the corresponding revisions in the manuscript are highlighted in red:

Reviewer #3 (Remarks to the Author):

Reviewer #3 **Overall comments** *This manuscript collected RS based reflectance, vegetation, water and built-up indices, texture and SAR data and trained a random forest (RF) classifier to extract floating water photovoltaics (WPV) over the downstream area of Yangtze river. Authors validated the final products and analyzed the recent trend of WPV projects. This manuscript is well written, except for some technical details that are missing. In addition, my major concern on this work is that I'm conservative on the application of this WPV product. The authors mentioned some cases of how others can use this product. But as an earth system science researcher, I strongly recommend the author to provide an application scenario under a broader earth system framework, which is also the main scope of ESSD, e.g., how to use this data product to improve the energy sectors and its interaction with others under an integrated assessment perspective.*

[Response] Thank you for this thoughtful comment. We agree that the broader application value of the WPV dataset should be framed more explicitly within an Earth system science perspective. In the revised manuscript, we expanded Section 4.2 to clarify that the dataset is not only an inventory of WPV distribution, but also a spatially explicit basis for analyzing interactions among renewable-energy development, waterbody use, and environmental change. In this way, the dataset can support integrated assessment of low-carbon energy transitions, particularly in regions where water-surface solar deployment is closely linked to aquaculture, reservoir regulation, and ecosystem management.

In addition, because the dataset distinguishes WPV-covered and uncovered areas within the same water body and provides annual information on deployment timing, it enables spatially explicit comparisons and before-and-after analyses. This creates opportunities to combine the dataset with other Earth observation products to investigate potential hydrological, thermal, and ecological responses associated with WPV expansion. For example, it can be integrated with thermal infrared products to examine possible water-surface temperature differences related to partial shading, and with water-color or water-quality indicators (e.g., chlorophyll-a, turbidity, or algal bloom proxies) to explore possible changes in optical properties and algal dynamics.

These applications are relevant not only for identifying possible environmental risks in high-coverage WPV systems, but also for understanding how energy-sector expansion may interact with water-resource functions and ecosystem processes under a broader coupled human-environment framework. The relevant text has been revised in Section 4.2 (Implications and Potential Applications). "The high-precision, decade-long WPV dataset developed in this study has both practical and scientific relevance**...** which require dedicated thermal infrared, water-quality remote sensing, or field observations." (Pages 16–18, Lines 424–459 in the clean version of the manuscript)

***Reviewer #3* Major Comments**

***Reviewer #3* Specific comment 1** *Input data of RF classifier include both direct reflectance of each Sentinel-2 MSI band and the combination of them, i.e., normalized indices, which have strong dependency among these different inputs. Though the classifier is relatively simple (only tell if a grid is WPV or not), which can cause the impact of this issue not reflected in your study, I would still recommend authors to discuss the potential impact from dependency in your inputs, i.e. multi-collinearity, when training your RF classifier.*

[Response] We appreciate this helpful comment. We agree that some spectral indices are mathematically derived from the original optical bands and may therefore be correlated with them, leading to potential multicollinearity among the input features. However, Random Forest (RF) is generally robust to correlated predictors in classification tasks, and such multicollinearity is therefore unlikely to substantially affect model performance, although it may influence the interpretation of variable importance. In our case, the task is a binary classification problem (WPV vs. non-WPV), and we did not observe any evident degradation in classification performance associated with the inclusion of correlated variables. Following the reviewer's suggestion, we added a brief clarification in Section 2.3.1 (WPV Feature Engineering) of the revised manuscript: "Although some spectral indices are derived from the original optical bands and are therefore correlated with them, Random Forest is generally robust to multicollinearity among predictors in classification tasks. As a result, the inclusion of both original bands and derived indices is not expected to adversely affect model performance, although it may influence the interpretation of variable importance." (Page 8, Lines 202–206 in the clean version of the manuscript)

***Reviewer #3* Specific comment 2** *For RF classifier training, how did you gain a robust model? Did you consider 10-fold cross-validation? How are parameters determined when training the model? It will also help if authors can think about using the SHAP value to reflect the importance of each feature on FPV detection.*

[Response] This is an important suggestion. We agree that additional details on model validation, parameter settings, and feature interpretation improve methodological transparency. In the revised manuscript, we clarified that the sample dataset was divided into 80 % training and 20 % testing subsets, and that 10-fold stratified cross-validation was conducted on the training subset to assess model stability. The cross-validation yielded a mean accuracy of 0.9729 ± 0.0030 across the ten folds, indicating stable classification performance.

We also added the Random Forest parameter settings. The model was implemented in Google Earth Engine using the smileRandomForest algorithm, with 80 trees and 7 variables randomly selected at each split. The remaining hyperparameters were kept at the default settings of the GEE implementation, including minimum leaf population = 1, bag fraction = 0.5, maximum nodes = null, and seed = 0. In addition, following the reviewer's suggestion, we calculated SHAP values to further interpret feature contributions. The SHAP-based ranking (Table R6) is broadly consistent with the RF variable-importance results, with NDBI, B2, NDPI, B11, and MNDWI emerging as the most influential predictors for WPV detection.

Table R6. Global Feature Importance for WPV Detection Based on SHAP Values.

Feature	Mean Absolute SHAP
NDBI	0.1802
B2	0.0630
NDPI	0.0529
B11	0.0523
MNDWI	0.0457
B12	0.0300
B8	0.0235
SAVI	0.0227

Reviewer #3 **Technical comments**

Reviewer #3 **Specific comment 3** *Line 32: What does "eliminated errors" mean?*

[Response] We intended to indicate that misclassified WPV areas were identified and removed through visual interpretation of Google Earth time-series imagery. Accordingly, "eliminated errors" has been revised to "removed misclassified areas" in the revised manuscript. (Page 2, Line 32 in the clean version of the manuscript)

Reviewer #3 **Specific comment 4** *Line 90: "unprecedented accuracy" can be changed to "high".*

[Response] Agree. We have replaced "unprecedented accuracy" with "high accuracy" to avoid overstatement in the revised manuscript (Page 5, Line 103; Page 16, Line 415 in the clean version of the manuscript).

Reviewer #3 **Specific comment 5** *Line 131: "To reduce cloud interference, a cloud-masking algorithm was applied, and annual median composites were generated from all available images." Repeated sentence. Please delete.*

[Response] We thank the reviewer for pointing this out. The repeated sentence has been removed from the revised manuscript.

Reviewer #3 **Specific comment 6** *Line 134: "These composites ensure radiometric consistency* *and provide a stable spatial baseline for dynamic WPV detection and temporal analysis." Repeated sentence. Please delete.*

[Response] We thank the reviewer for pointing this out. The repeated sentence has been removed from the revised manuscript.

***Reviewer #3* Specific comment 7** *Line 138: It can be helpful to have a flow chart describing how you merged different products into a unified water mask.*

[Response] We thank the reviewer for this valuable suggestion. We agree that a flow chart helps clarify how the unified water mask was constructed, and we have therefore added a new flow chart (Fig. R5) in the revised manuscript. Specifically, HydroLAKES provides lake boundaries and associated attributes, while GRanD supplies reservoir polygons and related information; these were first integrated to form the preliminary waterbody dataset. GOODD and GeoDar were then used as complementary dam datasets to support reservoir identification and improve the consistency of reservoir-related records where polygon or attribute information was incomplete or uncertain. The resulting waterbodies were subsequently classified into two categories: reservoirs, for those identified as dam-controlled systems, and lakes, for the remaining waterbodies. Finally, all datasets were harmonized into a unified maximum historical water-extent layer for the study area, which served as a stable water mask to constrain the analysis to long-term potential water bodies.

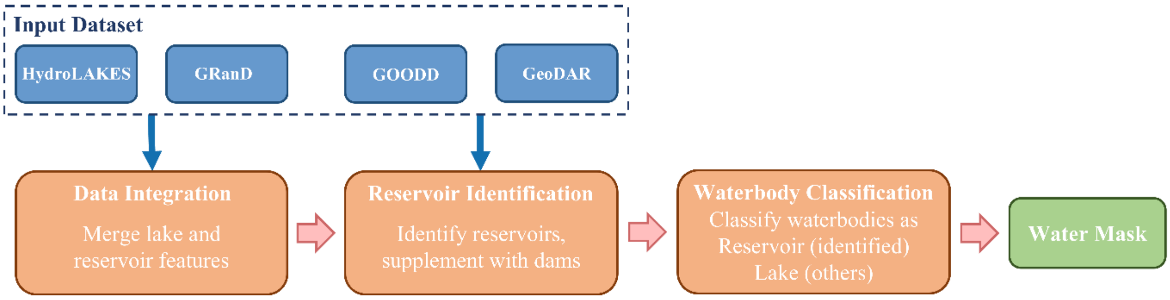

Figure R5. Flow chart of water mask construction. HydroLAKES provides lake boundaries, and GRanD supplies reservoir polygons; these are integrated to form the preliminary waterbody dataset. GOODD and GeoDar are used to improve reservoir identification. Waterbodies are then classified into two categories: identified reservoirs and the remaining lakes. All datasets are finally harmonized into a unified maximum historical water-extent layer, which serves as a stable water mask.

***Reviewer #3* Specific comment 8** *Figure 2: "a" shall be "d", "b" shall be "e", "c-e" shall be "a-c".*

[Response] Thank you for pointing out this error. The panel labels and caption of Figure 2 have been revised accordingly in the revised manuscript.

***Reviewer #3* Specific comment 9** *Line 193: Did you calculate texture features for all 6 bands?*

[Response] Thank you for this question. Texture features were calculated only for the B8 (NIR) band, as it provides a strong contrast between water surfaces and WPV. We have clarified this in Section 2.3.1 ("Texture features") of the revised manuscript: "Calculated from the B8 (NIR) band using the Gray Level Co-occurrence Matrix (Haralick et al., 1973). These features capture the characteristic spatial patterns of WPV arrays, which typically exhibit clear, regular boundaries that contrast with natural water bodies." (Page 9, Lines 212–214 in the clean version of the manuscript)

***Reviewer #3* Specific comment 10** *Line 233: "Each potential region was then subjected to rigorous manual interpretation and correction using high-resolution satellite imagery from Google Earth (Fig. 3c)." Please clarify what "manual interpretation" method you used. Based on expert judgement?*

[Response] We agree that the term "manual interpretation" should be defined more clearly. In this study, the interpretation was based on expert visual judgement using high-resolution Google Earth imagery, but it followed explicit visual criteria and temporal consistency checks rather than purely subjective assessment. Specifically, each potential WPV region was evaluated based on (1) its location within the water body, (2) the presence of regular and repetitive photovoltaic array patterns, and (3) clear separation from non-WPV objects such as shoreline buildings, roads, embankments, or floating vegetation. In addition, temporal consistency in Google Earth historical imagery was examined to confirm plausible installation timing and continued presence after deployment. Based on these criteria, false positives were removed and installation timing was further verified.

We have revised the manuscript (section 2.4.2 Manual Refinement and Final Dataset

Creation) to clarify this manual interpretation and correction procedure. "Each potential region was then interpreted and corrected using high-resolution satellite imagery from Google Earth (Fig. 3c) to accurately identify and remove misclassified non-WPV areas, thereby substantially improving the reliability of the final dataset. Specifically, each potential WPV region was checked for (1) its location within the water body, (2) the presence of regular and repetitive photovoltaic array patterns, and (3) separation from non-WPV objects such as shoreline buildings, roads, embankments, or floating vegetation. Since WPV installations are typically long-lasting, their installation year was determined by identifying the first year each site visibly appeared in high-resolution Google Earth imagery sequences. To maintain temporal consistency across the annual series, previously confirmed WPV areas were retained in subsequent years, and only newly detected regions were added. Through this temporal consistency rule, the annual WPV series followed a cumulative, non-decreasing pattern, effectively reducing spurious year-to-year disappearance caused by classification noise, short-term hydrological variations, or image-quality differences, thereby enhancing the reliability of decadal trend estimation." (Pages 10–11, Lines 252–265 in the clean version of the manuscript)

***Reviewer #3* Specific comment 11** *Line 250: It is not clear how you build a model to account for multi-year data. Did you combine multi-year reflectance, indices and texture and SAR data together and use them as input to train your model? Please explain the methodology.*

[Response] We agree that the multi-year classification workflow should be described more explicitly. In this study, the Random Forest classifier was trained using sample points interpreted from the 2024 annual median composite. We chose 2024 as the reference year because all training samples could be matched to confirmed WPV installations, whereas using earlier-year imagery could introduce samples from locations where WPV had not yet been deployed.

Multi-year reflectance, spectral indices, texture, and SAR features were not combined into a single multi-temporal input stack for model training. Instead, for each year from 2015 to 2024, we separately generated annual Sentinel-1/2 composites and derived the corresponding spectral, index, texture, and SAR features using the same feature-construction procedure. The RF model trained on the 2024 samples was then applied to each year's annual feature set to produce year-specific WPV classification results.

To improve temporal consistency in the annual series, the initial yearly classification outputs were further refined using a cumulative temporal-consistency rule: once a WPV area was confirmed, it was retained in subsequent years, and only newly detected areas were added. Google Earth time-series imagery was additionally used to verify installation timing and to correct potential temporal inconsistencies. We have clarified this methodology in the revised manuscript.

***Reviewer #3* Specific comment 12** *Figures 7, 8: Please change FPV to WPV.*

[Response] Thank you. The captions of Figures 7 and 8 have been updated accordingly.

References

Haralick, R. M., Shanmugam, K., and Dinstein, I.: Textural features for image classification, IEEE Trans. Syst. Man Cybern., SMC-3, 610 – 621, https://doi.org/10.1109/TSMC.1973.4309314, 1973.