Second Round Revision Notes for Manuscript ESSD-2025-137

Topic editor comments: *The revised manuscript presents a well-structured and valuable contribution, with solid methodology and meaningful results. The reviewers raised several minor concerns for the current version primarily related to clarity, figure readability, and transparency of parameter assumptions. These issues are readily addressable and do not affect the overall validity of the work.*

Response:

Thank you for the encouraging comment. We sincerely appreciate the editors and reviewers for their constructive feedback. We have thoroughly revised the manuscript in response to all comments. Detailed point-to-point responses are provided below.

Response to Reviewer#1's Comments

Please refer to the "changes_tracked" version of our manuscript, which is attached by the end of the Notes, to see our detailed edits and revisions.

<u>Replies to the General Comments:</u>

The paper presents an updated ET algorithm and dataset that shows improved performance at flux site, CONUS-basin, and global basin levels. It is a valuable contribution to the literature. I have the following minor comments.

<u>Response:</u>

We sincerely thank the reviewer for the positive evaluation and encouraging feedback. Our point-to-point responses to your comments are listed below.

<u>Replies to the Specific Comments:</u>

1. *p11 lines 283-286: NCEP2 is no doubt a well performing dataset in 2015, but the resolution and methodology of global reanalysis has significantly improved since then, and the use of NCEP2 needs additional justification. How may have NCEP2 biases contributed to the biases observed in Figs.11 and 13? Also, the interpolation of NCEP2 from 1.9-degree to 1/12 degree resolution probably resulted in underestimation of the spatial heterogeneity in ET within the original ~1.9degree cells?*

<u>Response:</u>

Thank you for your insightful and constructive comments.

Justification for using NCEP2: The use of NCEP2 in this study is primarily motivated by its consistency with the dataset used in the original version of our algorithm (Zhang et al., 2015), which facilitates comparability and helps isolate the effects of model improvements from changes in input data. Furthermore, NCEP2 offers a long-term, globally consistent dataset with relatively stable assimilation methods, which suits the multi-decadal scope of this study. We now clarify this rationale in the revised manuscript (see lines 548-551).

Potential contribution of NCEP2 biases to ET simulation: We agree that biases in NCEP2 may have contributed to some of the discrepancies shown in the ET spatial patterns. This is likely due to known cold and dry biases in NCEP2, including the underestimation of vapor pressure in certain regions, which could influence energy availability and evaporative demand. We have added a note discussing this in the revised Discussion section (see lines 551-553, 556-560).

Interpolation and underestimation of spatial heterogeneity: As you pointed out, the coarse resolution of NCEP2 (1.9°) could indeed limit the representation of fine-scale spatial heterogeneity when interpolated. However, since we mainly used NCEP2 for air temperature, wind speed, and vapor pressure—variables that tend to have relatively smoother spatial gradients—the impact of downscaling on spatial heterogeneity might be less severe compared to variables like soil moisture. Nonetheless, we acknowledge this as a limitation in the Discussion (see lines 553-556).

To address these limitations, we have added a paragraph in the Discussion section acknowledging the potential influence of NCEP2 biases on ET simulation and spatial patterns, and expressing our plan to incorporate higher-resolution and more modern reanalysis datasets (e.g., ERA5 or MERRA-2) in future work.

2. *The Feng et al. (2022) P-LSH algorithm has a radiation multiplier but no daylight temperature multiplier. Could the author explain why the swap in this paper? Also, I am not sure what is the meaning of daylight temperature - is it daytime temperature?*

Response:

Thank you for your valuable comment. Regarding the use of stress factors, both our previous work (Feng et al., 2022) and this study include daylight temperature (T_{day}) as stress factor—by which we indeed mean daytime temperature. In Feng et al. (2022), we incorporated five stress factors in the Jarvis-Stewart-based stomatal conductance model: T_{day}, VPD, CO_2, solar radiation (R_s), and soil moisture (SM). In this study, we excluded the radiation scalar and retained T_{day}, VPD, CO_2, and SM.

This change was primarily motivated by following considerations. First, our preliminary analysis showed that T_{day}, VPD and SM exert the strongest influence on stomatal conductance across most vegetation types and climatic zones, while radiation exhibited limited marginal influence in Jarvis-Stewart model. Excluding radiation improved model parsimony without significantly compromising performance.

Second, at the global scale, incorporating fewer but more robust input variables enhances model stability and reduces the propagation of uncertainties from remote sensing data. In many cases, radiation is strongly correlated with other meteorological variables (e.g., temperature, VPD), and its effects can be partially captured indirectly.

Finally, it is worth noting that radiation is already explicitly accounted for in the Penman-Monteith framework used for ET computation. Therefore, its exclusion from the Jarvis-Stewart stomatal conductance formulation does not mean it is neglected in the overall ET estimation framework.

Based on the above analysis, we eventually eliminated the radiation scalar in this study.

3. *The flux tower dots are overlapping and hard to read in Fig. 7. Also the caption does not explain what the "classified statistical values" are in the right panels. It is suggested to re-draw all the data as boxplots.*

Response:

Thank you for the valuable suggestion. We initially chose dot plots to display the individual statistics of each flux towers, aiming for a more intuitive understanding of data distribution. However, we acknowledge that overlapping markers reduce readability. To address this, we have reduced the marker size and applied transparent face colours in the revised figure to enhance clarity (see line 387).

Regarding the use of boxplots, we note that some subcategories contain fewer than five data points, which would make boxplots statistically unreliable and potentially misleading. In such cases, boxplots may not meaningfully represent quartiles or potential outliers, which could lead to misinterpretation.

Regarding the term “classified statistical values” in the right panels, it refers to statistical metrics summarized across all towers, dry towers, and wet towers, respectively, allowing for a clearer comparison of model performance across different zones. We have revised the caption to clarify this definition (see lines 388-391).

4. *Fig. 8: It is virtually impossible to see the P-LSHv2 line due to color choice. Please use a more visible color.*

Response:

Thank you for your suggestion. We have revised Figure 8 to improve the visibility of the P-LSHv2 line by using a more distinguishable and clearer colour (see line 406). We believe this change enhances the figure’s readability and better highlights the performance of our algorithm.

5. *Fig. 9: all the symbols are overlapping. Perhaps shrinking the marker size and setting the face color to transparent will make the reader better distinguish the blue, pink, and dark brown symbols.*

Response:

Thank you for the helpful suggestion. In the revised version of Figure 9, we have reduced the marker size and set the marker face colour to transparent as recommended. These adjustments improve the visual clarity and make it easier to distinguish between the blue, pink, and dark brown symbols. Please refer to the revised Figure 9 (see line 422).

Response to Reviewer#2's Comments

Please refer to the "changes_tracked" version of our manuscript, which is attached by the end of the Notes, to see our detailed edits and revisions.

<u>Replies to the General Comments:</u>

This paper provides an improved process-based land surface ET fluxes algorithm integrated with a soil moisture constraint scheme. The parameters are calibrated based on global flux tower observations over various climates and biomes, with uncertainty and sensitivity well quantified. The calibrated model is then used to generate a global daily evapotranspiration dataset, which outperforms its previous version and aligns well with other benchmark products. This new ET dataset will be a valuable reference for global energy budget and water balance studies.

The resubmitted manuscript is well written with high content clarity and comprehensive analysis. The comments from the two reviewers are well addressed, and the modifications in the updated manuscript are proper. Some comments addressed in the discussion, such as the dominant effects between energy vs. moisture constraints in high latitude regions, and the impact of deeper rootzone soil moisture, are interesting and worthy directions for further studies in the future. I recommend that the updated manuscript be accepted for publication.

<u>Response:</u>

We sincerely thank the reviewer for the positive and encouraging comments on our revised manuscript. We are grateful for the recognition of the improvements made in the algorithm, data calibration, and analysis. Detailed point-to-point responses are provided below.

<u>Replies to the Specific Comments:</u>

1. *"The type of prior distributions used for the parameters should be mentioned somewhere in the text. It would also be helpful to indicate the prior intervals in Figure 4 to make the comparison between prior and posterior more obvious.*

<u>Response:</u>

Thank you for the valuable comment. Since no prior information was available for the parameters, we assumed uniform distribution as prior distribution, which was subsequently updated to posterior distributions through the Bayesian-based methods. This clarification has been explicitly included in the revised manuscript (see lines 212-213).

Additionally, we revised Figure 4 to include dashed black lines indicating the lower and upper bounds of the prior intervals (see lines 334-337), to facilitate a clearer comparison between the prior and posterior distributions.