

This is a preliminary point-by-point response prepared for the discussion stage to summarize our current position and planned revisions. A complete response, including the corresponding manuscript changes and finalized additional analyses, will be provided together with the revised manuscript.

The manuscript developed a differentiable ecohydrological land model and used it to explore the parameter regionalization and identifiability problems in GPP and ET simulation. The development and findings are valuable and the manuscript is very well-written. I only have minor comments

Response: Thank you very much for your positive assessment of the manuscript. We also appreciate your careful reading and constructive suggestions, which will help us improve the evaluation and presentation of the study.

(1) lines 335-337: Could the author detail the range of the grid search and the method to create the representative training subset?

Response: Thank you for pointing this out. In the revised manuscript, we will provide the ranges tested for the hyperparameters and the representative training subset used in the Supplement.

(2) Figure 6 and Figure 7 plot model performance by sub-regions, and the current metrics only include spatially aggregated absolute values and grid-wise relative temporal correlations. Could the authors add whole domain aggregate metrics and some metric on relative spatial pattern (e.g. Pearson correlation on grid-level annual means between model and gridded observations)? The additions will provide complementary information to the existing metrics and more clearly demonstrate the model skill.

Response: We agree. In the revision, we will add pan-European summary statistics for GPP and ET and explicitly evaluate the spatial pattern of multi-year mean fluxes in the main text.

(3) One of the compared regional product, FLUXCOM X-BASE, has partial overlapping training sources as ADLEM - e.g. eddy covariance sites, ERA5 meteorological forcing (which ERA5-Land forcing was downscaled from). This should be acknowledged.

Interestingly, FLUXCOM is more similar to the other strongly observation-constrained products (GOSIF, GLEAM) than ADLEM, especially the the baseline case. This suggests the original structural constraint is quite strong in ADLEM. As the spatial freedom in ADLEM parameters increase, the simulation results become more similar to the gridded observational products. Could the authors discuss whether such convergence is general to differentiable land surface models with distributed parameters, and how to value the trade-off between structural regularization and observational fit?

Response: Thank you for this important point. We will revise the manuscript to explicitly acknowledge the partial dependence between ADELM and FLUXCOM X-BASE through the overlapping use of eddy covariance observations and related meteorological forcing products.

We also appreciate your interpretation of the trade-off between structural constraints and observational fit. We have long been interested in the question of how much process representation is needed and how much parameterization flexibility should be allowed in differentiable land surface models, and we think your interpretation closely coincides with this broader question. In the revision, we will add a discussion paragraph to consider this perspective, including how increasing parameterization flexibility, through learning more parameters as spatial functions of environmental attributes, may affect agreement with observation-constrained products and how this relates to the role of structural constraints in the model.

(3) Figure 10 shows the relationship between g_s and environmental covariates become weaker when more degrees of freedom is allowed. Can the authors add supplementary graphs to check whether this happened to the other vegetation and hydro parameters?

Response: We agree that the comparison should not be limited to g_s . In the revision, we will add supplementary diagnostics for the remaining learned parameters. We will examine whether opening the additional parameter group changes the strength or form of the inferred parameter-attribute relationships and their variability across cross-validation folds.

(4) The lines in Figures 5-7 are thick and overlap each other, making it hard to see some lines. Thinner lines will help.

Response: Thank you for the suggestion. We will revise Figs. 5-7 to improve visual separation among the experiments and reference products.