

General Assessment

I thank authors for revising the manuscript. I went through responses to reviewer#2 comments only. Several technical additions are indeed implemented, including the extension of the parameter recovery experiment to all 21 parameters, the addition of Monte Carlo dropout for uncertainty estimation, and the new computational efficiency and memory benchmarks.

However, some key concerns are only partially addressed. In particular,

- while the manuscript now clearly defines and justifies the prior ranges and the role of the sigmoid mapping and regularisation terms, the evaluation still relies on the same bounded priors. No quantitative sensitivity analysis is provided to assess **how dependent the results are on the choice of prior bounds**. Thus, **my original concern regarding robustness to modelling choices (activation function, priors, and loss formulation) is addressed mainly at the level of justification rather than empirical testing**.
- Similarly, **although uncertainty maps and normalised bias visualisations have been added, the biome-level traceability analysis remains unsupported by independent validation and therefore still risks being over-interpreted as mechanistic insight rather than model-derived structure**.
- The claims in the revised manuscript are softened, but several interpretations still exceed what is strictly supported by the validation presented.

Overall, the revision represents a substantial improvement and resolves several of the more technical criticisms. Nevertheless, important aspects of robustness and interpretability remain insufficiently addressed. I therefore consider the manuscript improved but not yet fully responsive to all methodological concerns raised in my previous review.

Comments remained from previous response letter

Comments 2, 8, 10, and 12, below need further implementation. Comment 1, 4, 11 can be addressed by acknowledging limitations and balancing claims. Comment 3 needs either further implementation or change in claims.

Comment 1 – Validation scope

While the authors have added uncertainty diagnostics and softened some language, the validation remains restricted to random cross-validation within the same data domain. The manuscript still lacks any independent or out-of-domain validation to support broader claims of ecological insight or general applicability.

Comment 2 – Model choice and robustness

Although the authors now document and justify their choices of activation function, priors, and loss formulation, the evaluation does not quantify sensitivity to the prior ranges themselves. Robustness to modelling choices is therefore argued conceptually rather than demonstrated empirically.

Comment 3 – Generalisation

The added memory benchmarks clarify computational trade-offs, but the manuscript still does not assess spatial or biome-level generalisation (e.g., leave-one-biome-out tests). As a result, the scope of applicability remains limited to the training domain.

Comment 4- parameter identifiability & interpretation

Extending the recovery experiment to all 21 parameters reveals heterogeneous identifiability, but interactions among parameters and their implications for interpretability remain unexplored. Weakly identifiable parameters are still used in downstream analyses without clear justification.

Comment 8 – Sensitivity analysis depth

The revised sensitivity analysis is now clearly described and corrected, but it remains a first-order, local analysis. Potential nonlinear interactions among parameters and higher-order sensitivities are not assessed.

Comment 10 – Bias explanation

The explanation of systematic bias between BINN and PRODA is plausible, but remains hypothetical. No targeted experiment is provided to demonstrate that the observed bias is indeed caused by prior-centering of weakly identifiable parameters.

Comment 11 – Spatial bias interpretation

Although normalization and uncertainty maps improve visual interpretability, the ecological meaning of the spatial bias patterns is still not analysed. The manuscript does not explain why specific regions exhibit persistent over- or underestimation.

Comment 12 – Biome-level trade-offs

The decomposition of SOC into carbon inputs versus residence time across biomes remains unsupported by independent or analytical validation. Without such validation, the biome-level trade-off analysis risks being over-interpreted as mechanistic insight rather than a model-derived diagnostic.

Line-by-Line Comment (focused on comments remained)

I hereby provide examples of places that needs to be modified to address the comments above. Please, check the remaining of the manuscript for similar caveats.

Comment 1 – Validation scope & claims

- Lines 20–21 (Abstract): Please qualify the claim 'retrieving mechanistic knowledge from big data'. Validation is limited to random cross-validation within the same dataset.
- Lines 31–33 (Abstract): Agreement with PRODA does not constitute independent validation. Please soften or restrict the claim.
- Lines 34–36 (Abstract): The phrase 'transformative tool' should be tempered or restricted to the current dataset.
- Lines 107–115 (Introduction): Clarify that inferred processes are model-derived and not independently observed.
- Lines 608–609 (Discussion): Acknowledge that process importance is conditional on the model structure and training domain.

Comment 3 – Generalisation

- Lines 386–388 (Methods): Random k-fold CV does not test biome-level transferability. Please clarify.
- Lines 544–545 (Results): Clarify that comparable accuracy does not imply out-of-domain generalisation.
- Lines 652–655 (Discussion): Restrict claims about uncovering mechanisms to the training domain.
- Suggestion: Either add leave-one-biome-out testing or explicitly limit scope to in-domain performance.

Comment 4 – Parameter identifiability & interpretation

- Lines 493–496 (Results): Clarify how heterogeneous identifiability affects interpretation of weak parameters.
- Lines 589–592 (Discussion): Explain whether weakly identifiable parameters are interpreted or excluded.

- Lines 646–649 (Discussion): Restrict mechanistic interpretation to parameters with demonstrated sensitivity and recovery skill.

Comment 11 – Spatial bias interpretation

- Lines 541–543 (Results): Provide interpretation or explicitly state that spatial bias is diagnostic only.
- Lines 546–548 (Results): Clarify what data or model limitations may explain spatial bias.
- Lines 734–735 (Figure 6 caption): State that no causal ecological interpretation is implied.
- Lines 552–559 (Results, Figure 7 description): The text interprets biome-level differences in SOC residence time and carbon inputs as reflecting distinct underlying mechanisms. Given that these patterns are derived entirely from model-based traceability and not independently validated, please clarify that Figure 7 provides a diagnostic decomposition of model behaviour rather than evidence of true ecological mechanisms.
- Figure 7 caption (Lines ~739–741): Please explicitly state that the biome-level trade-offs shown in Figure 7 are conditional on the BINN–CLM5 structure and should not be interpreted as empirical validation of process differences among biomes.
- Lines 630–646 (Discussion): Avoid implying ecological causality from bias patterns without independent support.