

Review of Manuscript titled “**Biogeochemistry-Informed Neural Network (BINN v1.0) for Improving Accuracy of Model 1 Prediction and Scientific Understanding of Soil Organic Carbon Storage**”

The revised manuscript shows clear and substantial improvement over the previous version, with several technical additions including the extension of the parameter recovery experiment, the incorporation of Monte Carlo dropout for uncertainty quantification, and improved clarity in the methodological description. The authors have also moderated some of the stronger claims and better positioned the work within the existing literature.

However, some key concerns remain only partially addressed. In particular, robustness to modelling assumptions (e.g., prior ranges and loss formulations) is discussed more clearly, but still relies primarily on conceptual justification rather than comprehensive empirical evaluation. While the sensitivity tests of prior ranges provide useful insight into interpretability, they do not fully assess the robustness of *inferred parameters* and resulting conclusions. In addition, the validation framework remains limited to in-domain cross-validation without independent or out-of-domain evaluation.

Furthermore, some interpretations—especially regarding mechanistic insights and biome-level patterns—still extend beyond what is strictly supported by the presented validation. While these optimized parameters provide a mechanically consistent representation of processes within the process-based model, the resulting mechanistic interpretations should be more explicitly framed as model-dependent and would benefit from clearer acknowledgement of the need for independent validation using complementary observational constraints (e.g., satellite observations, in situ sensor networks, and field or laboratory experiments). Incorporating relevant references in this context would strengthen the discussion (some are suggested attached).

Overall, the manuscript is significantly improved and technically sound, but would benefit from further strengthening of robustness analysis, clearer limitation of claims, and modest expansion of the discussion on validation. I leave it up to co-authors to polish the work further considering these aspects. I therefore recommend **minor revision**.

References

1. Baldocchi, D., et al. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological Society. [https://doi.org/10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2)
2. Fisher, J. B., et al. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks. Agricultural and Forest Meteorology. <https://doi.org/10.1002/2016WR020175>

3. Qiu, J., Cai, M., Keenan, T., Zhang, Y., Gentile, P., Luo, X., Cattray, M., Zhou, S., & Piao, S. (2025). Large contribution of antecedent climate to ecosystem productivity anomalies. *Nature Geoscience*. <https://doi.org/10.1038/s41561-025-01856-4>