In this article, the authors integrated a well-known process-based model CLM5 for simulating soil organic carbon states into a neural network framework which they call BiNN. They use a neural network to recover CLM5 parameters and compare this approach to Bayesian parameter estimation framework called PRODA. The strength of BiNNs is a significant speed up towards parameter calibration with PRODA and is supported by a preliminary experiment on synthetic data to showcase the approach. While this is an interesting and specific application for embedding neural networks and process-based models, as well as that the article builds on some interesting previous works, I'd recommend changes and extensions to the analysis, as well as restructuring the article before considering it for publication.
In general, I hesitate to consider this as a new model or method, but rather see it as an application of an already established method to an existing model. Further, to my knowledge, articles in GMD require a model version in the title, which in this case should contain at least CLM5, which is calibrated.

As the key idea of this article is that we should use BiNN because it speeds up parameter estimation at a similar level of accuracy as the second approach that estimates space-varying parameters, that is PRODA. However, the PRODA framework isn't properly introduced in the method section and evaluation of both frameworks in the results section is incomplete, so I suggest revisions of this part. As readers we don't know how good PRODA is performing. In section 4.2, only the correlations of PRODA and BiNN predictions are reported, but not the ones with the observations, they remain qualitatively described. Please add the according evaluation metrics here as well (is this what we see in figure 6? Then please refer to this figure).
Accordingly, when comparing the two approaches against each other (Figure 5), it would be good to see both of them evaluated against a common ground truth, e.g. as is displayed in figure 4 for just the BiNNs. A good choice of experimental design would be to do this within the synthetic experiment first, where the ground truth of the parameters is known. This means the results could compare 1) the correlation of recovery parameters from both approaches with the synthetic parameters and simulated SOC, and then 2) in applying this method after parameterisation compare the correlation of both predicted SOC (or other soil states) with the observed SOC. In the discussion, finally, PRODA should be contrasted with the parameter learning approach on a methodological level.

Another larger issue I see with this article is that equifinality isn't addressed in a critical way, while the synthetic results from figure 4 indicate that it is already present when fitting just four parameters (see section 3, why else are correlations not stronger here?). For the application later on, the authors fit all 21 parameters of the model, and it is to be expected that this will not diminish their unidentifiability. If this work aims for methodological demonstration, it could be better to just remain with a smaller set of parameters, but discuss their estimation in more detail, or, if going for many parameters, also do the simulation on many. Generally, the strength of a traditional Bayesian calibration approach is that we get uncertainty with the posterior distributions, information which unfortunately, neither PRoda nor the BiNNs currently provide but which could greatly support interpretation of these findings. This is a limitation of PRODA, that is introduced as a Bayesian approach but doesn't leverage its capabilities. I think the authors should openly discuss the limitations of PRODA in the methods section on PROda and in the discussion.

As a last major point, I have doubts about the naming that may be misleading. While the terminology is certainly used broadly, more recently, PiNNs have been re-defined more specifically as incorporating the learned ode in a combined loss function as teacher forcing (https://doi.org/10.1007/s44379-025-00015-1). The approach introduced in this study falls into the general realm of physics-informed machine learning but differs in the methods significantly from this definition of PiNNs, hence I believe the article would profit from re-specifying their method more precisely in title and abstract to, e.g. an end-to-end local parameter learning/recovery approach (see e.g. https://doi.org/10.1038/s41467-021-26107-z). Further, this specific approach to model calibration should be reviewed in the introduction and the physics-informed ML approach distinguished from other process-informed ML approaches. This will help structuring the article and methods section (for an overview of process-informed ML approaches see e.g. https://doi.org/10.1111/ele.70012).

Please find below a selection of minor points.

Structure: Section 5: These algorithmic details should come earlier and in a section on the fitting the network to CLM5 parameters (e.g. 2.1) . PRODA should totally not be introduced here but in an own subsection of section 2 after the NN and process model. Also, section 2 should be general methods. Why is there an own section on observational performance and computational efficiency? I suggest summarising results in a results section of their own. And move the data preparation methods section. And distribute the contents of section 4.2 to methods and results.

Section 1 (Introduction): 1) The beginning of the introduction lacks biogeochemical examples of parameter learning beyond soil, while mentioning a wide range of fields where hybrid models are applied. Preferably, mention other biogeochemical applications. 2) More importantly, hybrid approaches are introduced without any differentiation of how physical constraints are integrated with machine learning. See general comment on BiNNs, this would greatly help the reader at the beginning to locate the introduced approach (for an overview for e.g. for carbon flux with difference equations see https://doi.org/10.1111/ele.70012).
3) Further, if the goal is enabling interpretability of biogeochemical dynamics, as stated at the end of the introduction, I would like to know why and how BiNNs can lead to improvement here towards, e.g. traditional Bayesian approaches.
4) In contrast to hybrid and mechanistic approaches, there are also established statistical models that estimate spatially-varying parameters while maintaining

direct interpretability of their coefficients, such as SVCMs or Geographically-weighted regression. See for example: (10.1186/s12862-024-02260-z). This should at least be mentioned.

p. 5, Section 2: General introduction – state more precisely if this end-to-end or a two-step procedure – from Figure 1 and description in 2.2 I expect a fully integrated hybrid model, but here it sounds as if its two fragmented steps – please clarify.

Equation (2): Please check for mistakes, y_i is not defined, could be z_i.

Equation (5): How do you choose tau, also in the HP search? And why p_j – 0.5, i.e. could you elaborate on why you chose 0.5 - from the description I would expect a parameter-specific value for each p_j, i.e. the center of the prior distribution, unless you scaled them. If so, please mention.

Section 2.4: Please give the training details you have in section 5 here.

Section 3: If the goal of the paper is clearly stated in title and abstract, this section could simply be called simulation experiment.

Please state the type of CLM5 sensitivity analysis.

Given the equifinality and wide distributions in Figure 3, as this experiment was repeated in a CV, would it be possible to also report the standard deviation on the correlations?

Figure 3: See above. What sensitivity index was used and how was this done? Looks to me like feature importances.

Section 7: General. Quickly introduce BiNNs and the Bayesian approach at the beginning.

> “High correlations between BINN-retrieved and prescribed biogeochemical parameter values in a controlled parameter recovery experiment demonstrate BINN’s ability to recover causal relationships between covariates and SOC dynamics. Faithful retrieval of biogeochemical parameters from data substantially reduces uncertainty in SOC model predictions[26,31]
> .

These two sentences come quite lonely as they are. Please link back to your findings: Where do we see this?

Section 7.3: I agree that this approach may provide a new tool to model unresolved processes, it is not very clear on how it can help towards better mechanistic understanding with e.g. tracebility analysis mentioned. Could you explain this better? Also, there’s a lot of repetition in this paragraph.

The positional encoder that was used in the NN that informs the networks about the location. This design decision may blur the biogeochemical interpretation of parameter estimates if that is the goal. Here, a post-hoc check of sensitivity to location could be useful and if sensitive, run the analysis without the positional encoder.