Dear editor,

In the following, I try to evaluate the revisions done by the authors upon suggestions by the two reviewers of the first peer-review round. There were a lot of revisions successfully done by the authors and especially the implementation of uncertainty is a major achievement.

Reviewer #1:

Comment 1: The reviewer questioned the novelty of the approach. I share this concern, and the response of the authors could not fully resolve that. The possibility of including other models is also inherent to other code bases and not demonstrated within the manuscript, furthermore do the following two articles describe similar end-to-end setups (and further articles could be found):

1) Inferring dynamic parameters from dynamic covariates: ElGhawi et. al, Hybrid modeling of evapotranspiration: inferring stomatal and aerodynamic resistances using combined physics-based and machine learning, Environmental Research Letters, 2023.
2) Inferring static parameters from static covariates (in dynamic model): Baghirov et. al, H2MV (v1.0): Global Physically-Constrained Deep Learning Water Cycle Model with Vegetation, Geoscientific Model Development, 2024.

The novelty of doing similar but in a purely static or steady-state dynamic setup, using the matrix approach, uncertainty estimation or the application in a soil context are indeed given, but current phrasing overstates the innovation.

Comment 2: The reviewer asked for further introduction of PRODA (as the article focuses on comparison to it) and clarification on the PRODAs abilities. The extended introduction in response is sufficient to inform the readers. Also the added evaluation is completing the comparison of BINNs to PRODA. Nevertheless, are some conclusions, drawn from the correlation to PRODA, overstated.

1) 31-32: R of 0.77 with PRODA confirms correlation with PRODA, and the here made conclusion of that is not proven.
2) 46: Modeling assumptions are combined with a large dataset. As predictions of processes were not compared to independent data sources, identifying is too strong wording. This claim of identifying processes is made multiple times in the manuscript (e.g. 586-588).

Generally is PRODA presented as methodologically flawless. The fact that wrong models can also fit to data is not mentioned. The requested comparison on a methodological level is also not yet present.

Comment 3: The reviewer requested more analysis and discussion on equifinality. In response this issue is named in the text, but not prominently in abstract or conclusion. MC dropout is a interesting approach to identify equifinality, which could also be stated more prominently. But methodological I have following the following concern: 265-266: Implausible values should be prevented by the sigmoid already. Is this done to limit uncertainty of retrieved parameters? Please explain which effects the use of the hyperbolic cosine might have on the dropout-inferred uncertainty.

Comment 4: The reviewer identified possible confusion given the naming of the framework. The authors extended the introduction in response to position the framework as not being within the scope of PINNs. As conclusion I see the potential confusion even more obvious. Unfortunately is the almost identical naming to a (within the hybrid modeling world) very different method only resolvable by different naming. Furthermore do I see following issues in the positioning of the method (apart from points stated to Comment 1):

1) 75: Hard-PINNs would do that. (Le et. al, Physics-Informed Neural Networks with Hard Constraints for Inverse Design, Computational Methods in Science and Engineering, 2021)
2) 83-85: What is challenging about it? The impact of the paper depends on the challenges which were overcome here.
3) 95: Does reference 23 really say, that GWR has this limit?

In my understanding are PINNs emulating Differential equation (systems), while the here presented work learns parameters from covariates and thereby extends rather the given system to include unknown variables (covariates):

1) Raissi, Perdicaris, Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics, 2019.
2) Karpatne, Jia, Kumar, Knowledge-guided Machine Learning: Current Trends and Future Prospects, 2024

Comment 5 & 6: Apart from concerns mentioned earlier, I consider these as resolved.

Comment 7: While the mentioned issues were resolved, I had following additional concerns while reading this section.

1) 325-326: Why separation of depth here? Especially as results are later summarized again.
2) 400: Only positive inputs means, that LeakyReLU uses only positive linear part. No activation happens and effect of nodes collapses to a linear combination equally to Linear regression. Further should the sigmoid function only get positive inputs and only the upper half of the parameter range is effectively used. Problem might be fixed by the embeddings. But please clarify.
3) 434: Why is NSE used for accuracy and r for reporting parameter identification?
4) 443: This approach doesn't distinguish between train and test samples. More informative would be to take only the test samples per fold.
5) 467-468: Realistic time would be with early stopping (evaluation is provided everywhere) for each method. One converging faster than the other is an essential part of efficiency.
6) 168: This is called a parameterized sigmoid.
7) 181: CPU number is a hardware setting rather than a hyperparameter
8) 183: What are the loss function hyperparameters?
9) 188: Dropout rate was not mentioned earlier. Generally should the hyperparameter optimization be described after the framework, so terms are clear to the reader.

10) Eq. 11: Why is a capital p used here? Additional S indices for site and depth could clarify further what was done. Generally could paragraph 2.2.1 profit from rephrasing. Name it “first-order Sobol index”.
11) 380-381: What was validation used for?
12) 388: By averaging, potential model instability is rather hidden.
13) 456-458: I am not sure to which results this sentence hints, please clarify!

Comment 8 & 9: Apart from concerns mentioned earlier, I consider these as resolved.

Comment 10: Vide notes to Comment 2 & 3.

Comment 11: Why was spatial encoding used if it bares potential of data leakage and did not improve predictions?

Reviewer #2:

Comment 1: The reviewer requests more modest wording. Although this was taken care of mostly, following points still show problematic wording:

1) 18: Big data is a, not defined, buzzword and should be avoided.
2) 34: “transformative tool that harnesses the power of both AI and process-based modeling” was specifically mentioned by the reviewer and still remains in the text.
3) 53-55: Starting from AI is too broad here. Machine learning is the discipline, deep learning is one method of machine learning.
4) 65-67: Redundant.

Comment 2:

1) Resolved.
2) Vide notes to Reviewer #1, Comment 3.
3) Vide notes to Reviewer #1, Comment 2.
4) Vide notes to Reviewer #1, Comment 3.

Comment 3:

1) Resolved.
2) This could partly support the claims on causality. It is fine to leave that, but sticking to modest claims it necessary then.

Comment 4: Apart from concerns mentioned earlier, I consider these as resolved.

Comment 5, 6, 7, 8, 9: Resolved.

Comment 10 & 11: Fig. 5: I recommend to use a different color bar (partly white color code on white background).

Comment 12: This could partly support the claims on causality. It is fine to leave that, but sticking to modest claims it necessary then.

Additional remarks:

1) 498: approximate posterior distribution
2) 502: Which interval is used here?
3) 503: Can be, could be or should be? What is the general take home message of this paragraph? That equifinality is controlled well? If not, what would be the implications for results?
4) 506-507: Maybe I oversaw a reference on how this is done. Please call it 'Traceability analysis' throughout the text to help the reader.
5) 540: I think that is not Figure S6.
6) 553: This would be the definition of equifinality.
7) 570: S7 should be S8.
8) 4.1 and 4.3 could be one paragraph.
9) 635-637: Also well understood processes are parameterized.
10) 640: Fig 6a shows that there is no improved model fit.
11) 652-666: This is an outlook and should be formulated throughout in the conjunctive.