Response to RC1 https://doi.org/10.5194/egusphere-2025-6015-RC1

Thank you for taking the time to carefully review our paper. We have revised our manuscript as a result of your comments and we respond below to each of your points.

Comment 1: Page 3 states the land surface model “does not incorporate water in the sub-surface”, in which case I would have expected latent heat flux to be zero. However much of the paper is given to discussing the partitioning of latent/sensible heat fluxes, and the inverse model is asked to find the Bowen Ratio. Can authors explain from where latent heat fluxes in this scenario are originating or be explicit that latent heat fluxes are expected to be zero. If zero, can authors simplify the parameter inputs to exclude Bowen Ratio (i.e. all turbulent fluxes are partitioned into sensible heat in this scenario)?

Authors’ reply: To clarify, we do not mean that there is no water in the system. Rather, we did not incorporate a subsurface moisture model to represent processes such as heat transport by water in the subsurface. We have revised the line to make this clearer. We have amended the final paragraph in the abstract, the final paragraph of the introduction and the penultimate paragraph of the conclusions.

Comment 2: On Page 8 the solar zenith angle is parameterised. A short explanation of the form of the parametrisation would be useful for readers.

Authors’ reply: In line with your suggestions, we have revised the paragraph following Equation (15) to provide more details on the parameterisation.

Comment 3: The provided code (https://github.com/RuiyueH/SEB-model) does not include any instructions or README. In its current state I would not say this study is easily reproducible.

Authors’ reply: We have revised the code repository and added instructions as well as a README to make it easier to follow. The latest version can be found at https://doi.org/10.5281/zenodo.19344692.

Comment 4: If authors wish to reach a larger audience, more consideration could be given to different reader backgrounds. I believe those with machine learning interests are well served, but authors could also consider users of traditional land-surface

models and teams that observe land-atmosphere fluxes at flux tower sites. Both may find this machine learning technique interesting and potentially useful (although the current no-moisture LSM employed is a barrier, as authors have noted in the conclusion). Still, it could be beneficial for the reach of this study if authors further consider these users in the text and then provide an easier-to-follow code example. The current codebase includes no instructions for these users. **Authors' reply:** We have included some sample codes for Neural Physics using a simpler example - solving the second-order derivative of sin(x). For the code for the land surface model, we have split them into (i) forward model (synthetic data), (ii) inverse model (synthetic data) and (iii) inverse model (West Phoenix data), so that users can reproduce the results step-by-step. **Comment 5:** Authors could also take the opportunity to reconsider the abstract with a wider audience in mind (as above), for example indicating the potential of this approach for determining specific soil characteristics but mentioning current barriers (e.g. the current assumption of no soil water). **Authors' reply:** We have rewritten parts of the abstract and introduction to try to make them more generally accessible and included limitations in the abstract.
Response to CEC1 https://doi.org/10.5194/egusphere-2025-6015-CEC1 Dear Juan, Thank you for your comment. We sincerely apologise for not fully complying with the relevant policies in our initial submission. We have since revised the code repository to (i) provide clearer instructions and explanations of both the model and the Neural Physics approach, (ii) specify the required versions of all third party libraries, and (iii) include the necessary input and output files. The revised code has been published under https://doi.org/10.5281/zenodo.19344692. The manuscript has also been revised to cite the DOI in the 'Code and Data Availability' section, and the corresponding reference has been added to the bibliography. We hope the revised repository is now clear and complete. We apologise again for the oversight in our initial submission.

Response to RC2 https://doi.org/10.5194/egusphere-2025-6015-RC2

Thank you for taking the time to carefully review our paper. We have revised our manuscript as a result of your comments and we respond below to each of your points.

Comment 1: I am not convinced the article is significant enough for GMD and there are also flaws in the approach of the authors. First, about the significance: I am not a domain expert in land surface models but I've seen a wealth of research in differentiable models and hybrid ML e.g. for global hydrological models that are least somewhat related and these approaches go significantly beyond the model complexity shown here. The model the authors use isn't very complex and they just do a simple parameter estimation via gradient-descent from what I understand.

Authors' reply: Thank you for this comment. We present a fully differentiable, physics-based formulation of a land-surface model that enables end-to-end parameter estimation using gradient-based optimisation. While the forward model is intentionally simple, this allows us to isolate and systematically analyse the optimisation problem, including issues of identifiability and parameter dependence. The key contribution is not increased model complexity, but the demonstration that such models can be calibrated efficiently without adjoint derivation, and that the framework provides direct insight into when parameter estimation is well-posed and how this depends on the observational configuration.

Our approach is therefore distinct from recent hybrid physics–machine learning models, which introduce trainable components to learn unresolved processes. Instead, we retain a fully physics-based model and use machine-learning functionality solely to enable differentiability and optimisation. This provides a practical pathway for applying data assimilation and sensitivity analysis to land-surface models, without requiring bespoke adjoint implementations or specialised optimisation infrastructure.

We have revised the abstract and introduction accordingly. In particular, see paragraphs beginning "A number of researchers … " and "A related class…" in the introduction.

Comment 2: Especially for the Phoenix dataset: It is absolute standard practice to do a training - validation - test dataset split. The authors seemingly haven't done so. They even remark themselves: "This is not surprising, as these are the time series that were used in the optimisation". Without a data split you can't evaluate your results properly.

Authors' reply: Note that we do not train any neural networks for the results shown in our paper. However, we do agree with the reviewer that we should test our model on data not used in the parameter estimation process. We now use the first 150 hours of soil temperature measurements for the parameter estimation and use a subsequent 50 hours of data for evaluation of the model. Section 3.3 has been updated to reflect this, as well as a sentence in the final paragraph of the introduction and a sentence in the conclusion. We note that due to temporal correlation in the forcing and state variables, this split does not represent fully independent samples, but provides a first test of out-of-sample predictive capability.

Comment 3: I know GMD is not an ML or statistics journal but I still really don't think you need a flow chart to explain gradient decent optimisation. It's such a basic algorithm.

Authors' reply: Thank you for highlighting this. We agree that gradient descent is a well-established and widely understood optimisation method. We have therefore removed this diagram and replaced it with one that illustrates our end-to-end differentiable physics-based approach to parameter estimation, thereby clarifying the key features of the proposed methodology. It appears as Figure 2 in the revised manuscript.

Comment 4: Why write down the formula for a vanilla gradient descent in Eq 15 when you don't actually use it, but ADAM. ADAM includes momentum and doesn't follow Eq15.

Authors' reply: Thank you for the suggestion. We have removed this equation.

Comment 5: ADAM uses adaptive learning rates, yet you write of set values of the learning rate for each variable. How can I understand this?

Authors' reply: The Adam algorithm starts with an initial step length for each parameter which is modified as the iterations progress. All we do is to set a different initial step length for each parameter because the parameters have very different magnitudes. The step length is modified as usual by Adam. According to Kingma and Ba (2017), the setting of the initial learning rate (or stepsize) is still important for Adam, because the effective magnitude of the steps taken in parameter space in each iteration are approximately bounded by the step size that was set. We also learned this from our own tests: C changes very little each iteration if we use the same small step size as other parameters. We have modified the text in Section 2.3.

Comment 6: In the reverse pass (so the computation of the gradients) did you pass through all iterations of the Jacobi iteration or did you use an adjoint of the operation?

Authors' reply: Yes, we pass through all iterations of the Jacobi method, which is the discrete adjoint.

Comment 7: I've never seen the learning rate η being called "step length". It's very unusual.

Authors' reply: In classical optimisation, η is very often referred to as step length or step size, see Fletcher (1987) or Nocedal and Wright (2006), for example. Even in machine-learning texts it can be referred to as stepsize, see Kingma and Ba (2017), although in the machine-learning context the term learning rate is much more common. We wish to convey to the reader that we are not training a neural network, instead we are using steepest descent to solve an optimisation problem and so chose the terminology "step length" in line with classical texts on optimisation. We have added references to the text in Section 2.3.

Comment 8: Why is the data in the Phoenix dataset bias-corrected and what impact has this on the results?

Authors' reply: According to Lipson et al (2022, 2024), the ERA5 reanalysis data is bias-corrected using site observations, to account for diurnal and seasonal biases in the reanalysis method. The bias-corrected reanalysis data is then used to fill gaps in the original data collected from flux tower.

References:

Fletcher (1987) Practical Methods of Optimization. Second edition, Wiley & Sons, doi:10.1002/9781118723203.

Nocedal and Wright (2006) Numerical Optimization. Second Edition, Springer, doi:10.1007/978-0-387-40065-5.

Kingma and Ba (2017) Adam: A Method for Stochastic Optimization, ICLR 2015, arXiv preprint 1412.6980, doi:10.48550/arXiv.1412.6980.

Lipson, M., Grimmond, S., Best, M., Chow, W. T. L., Christen, A., Chrysoulakis, N., Coutts, A., Crawford, B., Earl, S., Evans, J., Fortuniak,K., Heusinkveld, B. G., Hong, J.-W., Hong, J., Järvi, L., Jo, S., Kim, Y.-H., Kotthaus, S., Lee, K., Masson, V., McFadden, J. P., Michels, O., Pawlak, W., Roth, M., Sugawara, H., Tapper, N., Velasco, E., and Ward, H. C.: Harmonized gap-filled datasets from 20 urban flux tower sites, Earth System Science Data, 14, 5157–5178, https://doi.org/10.5194/essd-14-5157-2022, 2022

Lipson, M. J., Grimmond, S., Best, M., Abramowitz, G., Coutts, A., Tapper, N., Baik, J.-J., Beyers, M., Blunn, L., Boussetta, S., Bou-Zeid, E., De Kauwe, M. G., de Munck, C., Demuzere, M., Fatichi, S., Fortuniak, K., Han, B.-S., Hendry, M. A., Kikegawa, Y., Kondo, H., Lee, D.-I., Lee, S.-H., Lemonsu, A., Machado, T., Manoli, G., Martilli, A., Masson, V., McNorton, J., Meili, N., Meyer, D., Nice, K. A., Oleson, K. W., Park, S.-B., Roth, M., Schoetter, R., Simón-Moral, A., Steeneveld, G.-J., Sun, T., Takane, Y., Thatcher, M., Tsiringakis, A., Varentsov, M., Wang, C., Wang, Z.-H., and Pitman, A. J.: Evaluation of 30 urban land surface models in the Urban-PLUMBER project: Phase 1 results, Quarterly Journal of the Royal Meteorological Society, 150, 126–169, https://doi.org/https://doi.org/10.1002/qj.4589, 2024.

Response to EC1 https://doi.org/10.5194/egusphere-2025-6015-EC1

Thank you for taking the time to carefully review our paper. We have revised our manuscript as a result of your comments and we respond below to each of your points.

Comment 1: First of all, there are missing key literature reviews on differentiable land surface model (LSM). Differentiable LSM is not new anymore. Several studies in the past few years developed different differentiable LSMs by refactoring the physical models into a framework that supports automatic differentiation (e.g., JAX, PyTorch, and Julia), including but not limited to, Fang and Gentine (2024; doi: 10.1029/2024MS004308) and Jiang et al. (2025; doi: 10.1029/2024WR038116). They are neither referred or discussed in the manuscript. It is also unclear what's the benefit of the proposed differentiable LSM compared to the existing one, instead of having a new model available in the market.

Authors' reply: Thank you for bringing our attention to those papers. We have included them and some others in our literature review, and revised the text to make clear how our method differs from other approaches. In a nutshell, our method reimagines numerical discretisations as convolutional neural networks where the weights are set analytically according to the choice of discretisation scheme. Our approach is purely physics-based and we do not train the neural network or develop a hybrid approach as such.

The key contribution of this work is the formulation of a land-surface model as a fully differentiable, physics-based system that enables efficient, end-to-end parameter estimation and analysis. By expressing the governing equations within a machine-learning framework, we avoid the need for deriving and maintaining adjoint models and can directly apply gradient-based optimisation to time-dependent problems. In addition, this formulation provides insight into parameter identifiability and confounding (e.g. identifiable parameter combinations and the role of observational configuration), which is typically difficult to assess in traditional calibration workflows. We have revised the manuscript to clarify that the contribution lies in this differentiable formulation and its implications for parameter estimation and analysis, rather than in the development of a new operational LSM.

We have modified the abstract (opening paragraph), introduction (in particular see paragraphs beginning “A number of researchers …”, “A related class …” and “The key contribution of this paper …”) and the conclusions (opening paragraph).

Comment 2: Second, the so-called “Neural physics” might be over-stated. While the model is reimplemented in PyTorch, no trainable deep learning layer is used except the adoption of a convolutional layer with fixed parameters to solve the heat equation. In that sense, the whole model is still pure physics-based and no DL is involved to learn additional dynamics that are not captured by the model. It is oftentimes expected to see the adoption of hybrid modeling in these differentiable modeling, which in fact has been studied in Fang and Gentine (2024) and Jiang et al. (2025).

Authors’ reply: Neural Physics is a term coined in Chen et al. (2026) that refers to using the convolutional layers of a neural network to implement numerical discretisations. In this paper we showcase its ability as a land-surface model and demonstrate how a model like this can be used straightforwardly to calibrate the parameters – a process that is typically tedious (see Lipson et al. 2024) . As such the concept is not over-stated, rather it is the core of what we present here. We recognise that the term may be interpreted differently across communities and have clarified in the manuscript that no trainable neural network components are used in this study. Please see the introduction (in particular see paragraphs beginning “A number of researchers…” and “A related class…”)

Whilst it is certainly possible to augment our model with a deep learning layer, the parameter calibration shows that the model as posed is able to reproduce the observations reasonably well. For longer time-series, more sophisticated modelling is required and it might be needed to add a deep learning layer. We have added a sentence in the conclusions on this (penultimate paragraph).

Note that there is an appetite for differentiable physics-based solvers. As we were uploading our paper to GMD, another paper was under review with a similar scope to ours (Tian, 2026), which investigates the physics-based modelling of permafrost using Recurrent Neural Networks. Although similar, our use of Convolutional Neural Networks gives greater scope for implementing multigrid solvers, advantageous in improving convergence.

We have modified the abstract, introduction and conclusion to better explain the approach that we propose in this paper.

Comment 3: Essentially, the adopted optimization technique is nothing new but a gradient-based optimization. While supported by AD, this is fundamentally no different from other gradient-based approach. Trying different initial values is also nothing new in optimization field which is always struggling in getting the global optimal solution.

Authors' reply: Exactly. In this paper we use gradient-based optimisation to assimilate data and estimate parameters of LSMs. The novelty of the method lies in using convolutional neural networks to represent the discretised physics as explained in our response to this reviewer's first and second points. The key advance is not the use of gradient-based optimisation per se, but that the full parameter estimation problem for a time-dependent LSM can be solved end-to-end in a differentiable framework without adjoint derivation or bespoke optimisation infrastructure. While gradient-based optimisation itself is not new, the differentiable implementation removes the need for deriving and maintaining adjoint models, which is a significant practical barrier in traditional LSM calibration workflows.

Comment 4: Lastly, the challenge of estimating dependent or confounding parameters is not addressed except a restatement of the existing difficulty in this particularly setting. In fact, such equifinality issue is a long-standing challenge in calibrating earth system model or any physical models. I would love to see any new angle to tackle the problem. But again, I didn't find new results out of here.

Authors' reply: Thank you for this comment. We agree that equifinality and parameter dependence are long-standing challenges; however, our work goes beyond restating this difficulty by quantitatively diagnosing when and how it arises, and under what conditions it can be reduced. In particular, using synthetic experiments with known ground truth, we show that parameter estimation based on temperature observations at a single depth is fundamentally ill-posed: multiple parameter combinations (e.g. h and β) collapse onto an identifiable product $h(1+\beta^{-1})$, and soil parameters (λ, C) exhibit strong correlations with a large spread in their individual estimates. Crucially, we then demonstrate that this non-identifiability is not intrinsic to the model alone but depends on the observational configuration: adding a second temperature depth resolves this degeneracy and enables reliable recovery of thermal conductivity and heat capacity, independent of initial conditions . At the same time, we show that some confounding cannot be resolved without additional physics or data—for example, sensible and latent heat fluxes remain inseparable without direct flux measurements. This provides a concrete, quantitative characterisation of equifinality in this setting and identifies specific pathways (multi-depth observations, parameter reparameterisation, or additional measurements) through which it can be mitigated. This moves beyond diagnosing equifinality by explicitly linking identifiability to observational design, thereby providing guidance on what measurements are required to constrain specific parameter combinations. We report our findings in Sections 3.1 and 3.2, and have expanded the discussion of this in the first paragraph of the conclusions.