Response to Review Comment #1

The state-of-the-art land surface models (LSMs) have been reported to perform poorly in representing permafrost processes. To address this gap, the authors present NoahPy—a fully differentiable LSM developed by reconstructing the Noah LSM's governing partial differential equations into a process-encapsulated recurrent neural network. NoahPy was compared with both the original and an improved version of the Noah LSM, and evaluated at a permafrost site. I find the model to be skillful and the results reasonable.

Response：We sincerely thank the reviewer for the constructive and thoughtful comments, which helped us improve the quality and clarity of the manuscript. We have prepared revisions to address all of these comments.

1. Introduction**:** It would be beneficial to restructure the introduction to better highlight the significance of permafrost, particularly as the authors aim to introduce the model to the permafrost research community. The section could begin by underscoring the importance of permafrost, followed by a critical review of how current LSMs represent permafrost processes, clearly outlining existing limitations. Addressing this gap, the authors should then introduce deep learning methods and explain how such approaches can provide an effective solution to improve permafrost modeling.

Response: Thank you.

The reviewer suggested a restructuring for the Introduction section and introducing deep learning as a solution to improve permafrost modeling. While we agree on the importance of these topics, our paper's central contribution is slightly different. Our primary goal is not to introduce new physics to solve the well-documented LSM deficiencies, but rather to solve a critical technical gap: the non-differentiable nature of existing permafrost-capable LSMs.

This technical gap is what currently prevents the permafrost community from leveraging the power of hybrid AI modeling, AI-driven calibration, and end-to-end differentiable workflows. Therefore, our introduction is structured to first establish the promise of hybrid AI, then identify the differentiable gap, and

then frame our work as the solution to this specific technical gap for the permafrost community.

In this revision, to better address the reviewer's points, we enhanced our permafrost-centric focus in the introduction section as well as other sections.

Please refer to the revised manuscript for the detailed changes.

2. Discussion: The advantages and limitations are currently intermingled in this section. Please consider: (a) adding a brief outlook on future model development; and (b) using subsections to enhance the readability of the manuscript.

Response:

We thank the reviewer for the careful reading and constructive comments on the "Discussion" section of our manuscript. We fully understand the two key issues raised: (a) the mixing of model strengths and limitations within the section, and (b) the lack of outlook on future model development. In response, while retaining the "Discussion" as a single cohesive section, we have reorganized its structure and logic as follows:

1) Structural adjustments:

Within the "Discussion" section, we have reordered the paragraphs to improve clarity and logical flow. We first highlight the main advantages of NoahPy, including the significant improvements brought by its differentiable framework in model transparency, parameter optimization efficiency, and error diagnostics, as well as its higher stability and scalability compared with traditional optimization algorithms such as SCE-UA. We then discuss the current limitations of the model, such as inheriting known physical deficiencies from the Noah LSM and its validation being primarily limited to the TGL sites on the Tibetan Plateau.

Finally, we have added a paragraph outlining future model development. We emphasize that NoahPy is not a finished product but an open and extensible framework intended to provide the permafrost modeling community with a platform for continuous improvement. This framework supports deep coupling with external machine learning models and can learn complex mappings

between environmental covariates (e.g., topography, vegetation, soil type) and physical parameters (e.g., hydraulic conductivity, thermal conductivity), thereby enhancing regional transferability of model parameters and reducing reliance on expensive pointwise calibration, effectively mitigating parameter uncertainty in permafrost simulations. Leveraging automatic differentiation, NoahPy also enables “modular” updates of specific physical processes—for example, embedding neural networks to replace empirical hydraulic parameterizations—while preserving energy and mass conservation constraints and learning more accurate physical relationships. This work helps bridge the gap between process-based modeling and AI, establishing a path toward the next generation of hybrid Earth System Models capable of reducing uncertainty and providing more reliable projections for the future of the cryosphere.

2) On sectioning:

We carefully considered the suggestion to add subsection headings. However, given the relatively concise length of the “Discussion,” splitting it into multiple subsections would result in overly short segments, potentially disrupting overall coherence and reading flow. Therefore, we opted to maintain a single-section structure, using natural transitions and logical connections to differentiate between model strengths, limitations, and future perspectives.

We believe that these revisions significantly improve the logical flow and readability of the discussion while fully addressing the reviewer’s comments regarding structural clarity and outlook on future model developments.

3. The language should be improved throughout for clarity and academic tone.

Response: Thanks, we have improved the language thoroughly in this revision.

4. L18: Avoid using the word “perfectly,” as no model can be considered perfect. Please revise this throughout the manuscript (e.g., L224 and others). As noted on the GMD homepage:“Essentially, all models are wrong, but some are useful.” (George E. P. Box, 1979)

Response：We thank the reviewer for this valuable suggestion. We agree that the word “perfectly” may imply unrealistic precision and have revised it throughout the manuscript. All “perfect” and similar words are revised. Specifically:

- In the Abstract (L18), we replaced "perfectly replicates" with "very closely replicates."
- In the Conclusions (L392), we replaced "perfectly reproduces" with "faithfully reproduces."
- In the Conclusions (L393), we changed "near-perfect match" to "very close match."

All similar expressions have been revised accordingly to avoid overstatement.

5. L22: “SCE-UA” is not defined.

Response: We thank the reviewer for the careful reading and valuable suggestion. We agree with this comment. The full name “Shuffled Complex Evolution - developed at the University of Arizona (SCE-UA)” has been added when it first appears in the abstract.

6. L105: Eq.6: Clarify what “β^F” and “β^G” refer to.

Response：β^F and β^G represent parameter sets involved in the control equations and output equations, respectively. Revised in manuscript.

7. L204: While a reference is provided, please explain the principle of the Shuffled Complex Evolution (SCE-UA) algorithm. For example, what does the strength of the SCE-UA algorithm stem from?

Response：We thank the reviewer for this constructive comment. In the revised manuscript, we have added a detailed explanation of the underlying principle and strengths of the Shuffled Complex Evolution (SCE-UA) algorithm in Section 2.3.3. The newly added text reads as follows:
“The Shuffled Complex Evolution (SCE-UA) algorithm (Duan et al., 1994) is a global optimization method that combines probabilistic sampling with competitive evolution. It starts by generating multiple ‘complexes’, each

representing a subgroup of candidate parameter sets. Within each complex, solutions evolve independently through processes analogous to selection, crossover, and mutation to produce new trial members. Periodic ‘shuffling’ of complexes allows information exchange among subpopulations, helping the search escape local minima and preserve population diversity (Rahnamay et al., 2019). This shuffled and competitive framework enables SCE-UA to efficiently balance global exploration and local exploitation, offering strong robustness and reliability for calibrating complex, nonlinear hydrological and land surface models.”

References:

Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, Journal of Hydrology, 158, 265-284, https://doi.org/10.1016/0022-1694(94)90057-4, 1994.

Rahnamay Naeini, M., Analui, B., Gupta, H. V., Duan, Q., and Sorooshian, S.: Three decades of the Shuffled Complex Evolution (SCE-UA) optimization algorithm: Review and applications, Scientia Iranica, 26, 2015-2031, https://doi.org/10.24200/sci.2019.21500, 2019.

8. L255: This could be easily verified by additionally evaluating snow water equivalent or snow depth against observations. I assume snow depth data are available at the TGL site (e.g., Xiao et al., 2013).

Response：

As suggested, we obtained the daily snow depth observations from the TGL site (with gratitude to the Cryosphere Research Station on the Qinghai–Tibet Plateau, CAS) and compared them with our NoahPy simulations.

This comparison (the following figure) provides definitive proof of our statement. Our model significantly underestimates the peak snow depth during the 2008-2009 winter. This is precisely the period where our soil temperature simulation exhibited its most pronounced cold bias (as seen in our original Figure 4a). The model’s simulated snowpack is far too shallow and melts too quickly, which confirms that the lack of insulation from this

underestimated snowpack is the potential cause of the simulated soil temperature bias.

We have revised the manuscript text in Section 3.2 (around L255) to reflect this new, stronger evidence.

“However, the model exhibits a cold bias during the winter of 2008–2009, with simulated temperatures falling below observations (Figure 4a). This period was characterized by heavy snowfall at the site. The cold bias is confirmed to be a direct result of the relatively simplistic snow scheme in the Noah LSM. A direct comparison with observed snow depth data from the TGL site, shows the model significantly underestimates the peak snow accumulation during this exact 2008-2009 winter and melts the snowpack too rapidly. The resulting shallower simulated snowpack provides less insulation, allowing excessive heat loss from the soil to the cold atmosphere.”

We are not adding this new figure to the main manuscript, as we wish to keep the paper's focus squarely on the soil thermo-hydrology and the novel differentiable workflow. However, we are very grateful for this suggestion

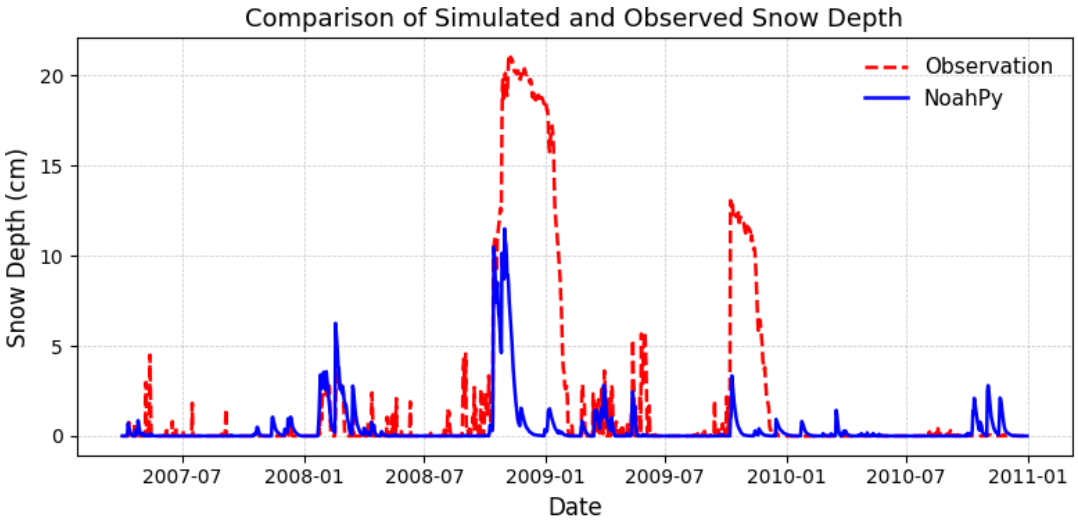

Figure S1 Comparison of Simulated(NoahPy) and Observed Snow Depth

9. L296: There are two opening parentheses here. Similar typos occur elsewhere in the manuscript; please revise carefully.

Response: The typos are corrected in revised manuscript.

10. L407: What does the underline signify here?

Response: The underline is removed in the revised manuscript.