NoahPy: A differentiable Noah land surface model for simulating permafrost thermo-hydrology

Wenbiao Tian[1,2], Hu Yu[3], Shuping Zhao[2*], Yuhe Cao[4], Wenjun Yi[2], Jiwei Xu[2], Zhuotong Nan[1,2,5*]

[1]State Key Laboratory of Climate System Prediction and Risk Management, Nanjing Normal University, Nanjing 210023, China
[2]Key Laboratory of Ministry of Education on Virtual Geographic Environment, Nanjing Normal University, Nanjing 210023, China
[3]North Information Control Research Academy Group Co., Ltd., Nanjing 211153, China
[4]College of Forestry, Northeast Forestry University, Harbin 150040, China
[5]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China

Correspondence: Shuping Zhao (shuping@njnu.edu.cn) and Zhuotong Nan (nanzt@njnu.edu.cn)

Abstract. Accurately representing permafrost in Earth System Models is a grand challenge that creates major uncertainty. A promising path forward is to create hybrid models that synergize process-based physics with deep learning, but this is fundamentally hindered by the non-differentiable nature of traditional land surface models (LSMs), which are incompatible with modern AI workflows. To overcome this limitation, we present NoahPy, a differentiable LSM developed by reconstructing the Noah LSM's governing partial differential equations into a process-encapsulated Recurrent Neural Network (RNN), with the heat–moisture solver forming the computational core. We first demonstrate that NoahPy very closely replicates the numerical behaviour of the modified Noah LSM, achieving Nash-Sutcliffe Efficiency (NSE) coefficients above 0.99 for both soil temperature and liquid water. We then show that at a permafrost site, the calibrated NoahPy achieves robust simulation performance for soil temperature (NSE > 0.9) and liquid water (NSE > 0.8). Critically, the differentiable workflow, when combined with the Adam optimizer, is significantly faster, more stable, and yields simulations with lower uncertainty compared to traditional Shuffled Complex Evolution (SCE-UA) calibration algorithm. NoahPy thus provides a foundational, "glass-box" framework that closes a key technical gap, enabling the development of the next generation of hybrid AI-physics models needed to more reliably predict the future of the cryosphere.

1 Introduction

The advent of deep learning has catalyzed a paradigm shift in Earth system science. Large-scale, data-driven models like Google DeepMind's GraphCast (Lam et al., 2023) and Huawei's Pangu-Weather (Bi et al., 2023) demonstrate remarkable skill in Earth system forecasting. However, their predictive power is often shadowed by a critical limitation: as "black-box" systems, they offer no guarantee of physical consistency or interpretability (Nearing et al., 2021; Wi and Steinschneider, 2022). While techniques from eXplainable AI (XAI) can provide post-hoc insights (Rudin, 2019; O'loughlin et al., 2025), they cannot

enforce physical laws, creating the risk of learning statistically powerful but mechanistically flawed relationships. This challenge is especially pronounced in complex, data-scare environments like the cryosphere. This "physics gap" has spurred a movement towards hybrid modeling that synergize the predictive prowess of machine learning with the mechanistic rigor of process-based physical models (Irrgang et al., 2021; Reichstein et al., 2019; Chen et al., 2023).

A powerful approach in this domain is the physics-informed neural network, which embeds the governing equations of a physical system directly into the model's architecture (Reichstein et al., 2019; Chen et al., 2023). Unlike "loosely-coupled" hybrids that use physics as a soft penalty in the loss function (Wang et al., 2020; Xie et al., 2022) or use machine learning to correct a physical model's output (Bonavita and Laloyaux, 2020), this deeply-integrated approach imposes hard constraints, rendering the model structurally incapable of violating fundamental laws. The primary obstacle to this integration for the land surface and permafrost modeling community has been technical: most established geophysical models, including well-known land surface models (LSMs), are implemented as non-differentiable numerical solvers, making them incompatible with the gradient-based optimization central to deep learning (Rumelhart et al., 1986). A transformative solution is differentiable programming, which involves rewriting a physical model's logic using differentiable operations within a machine learning framework like PyTorch or TensorFlow. This recasts the physical model into a "glass-box" system that is both physically interpretable and trainable end-to-end via backpropagation (Shen et al., 2023). Recent successes in hydrology have demonstrated the potential of this approach, yielding models with higher accuracy and improved generalization (Feng et al., 2022; Wang et al., 2024).

This approach is particularly critical for modeling permafrost. Improving the representation of freeze-thaw processes in Earth system models is a grand challenge (Schädel et al., 2024). Covering nearly 15% of the Northern Hemisphere's exposed land area, permafrost is a crucial regulator of global water, energy, and carbon cycles (Obu, 2021). Despite its vast scale, state-of-the-art LSMs), as the foundational components of climate models, have well-documented deficiencies in representing freeze-thaw processes in these regions (Matthes et al., 2025; Abdelhamed et al., 2023). They often simplify or omit key thermo-hydrological dynamics, such as abrupt thaw (thermokarst), the formation of excess ground ice, the insulation from thick organic soil layers, and complex water transport at the freeze-thaw front (cryosuction). These simplifications lead to significant biases in simulating active layer dynamics and the rate of permafrost thaw, and low confidence in the timing and magnitude of the permafrost carbon feedback, undermining the reliability of climate projections and estimates of the remaining carbon budget. While significant effort has gone into improving the physics of permafrost specific models (Ji et al., 2022; Wu et al., 2018; Xiao et al., 2013; Zhao et al., 2023), these improved models remain non-differentiable, preventing their integration into model AI-driven calibration and hybrid modeling workflows.

A differentiable LSM, by itself, does not inherently fix these physical deficiencies. Its true power is unlocked when applied to an already improved physical core, enabling it to serve as a foundational component for more sophisticated hybrid artificial intelligence (AI) systems. A differentiable, permafrost-focused LSM enables AI-driven parameterization, where the differentiable LSM is coupled with a neural network that learns to predict its internal parameters (e.g., hydraulic conductivity,

thermal properties) from external data, thus addressing the long-standing challenge of parameter uncertainty (Tsai et al., 2021; Wang et al., 2024; Sun et al., 2024). More importantly, it can be embedded as a physics core within a larger, end-to-end trainable AI-based Earth system model. This forces the larger model to follow the laws of land surface physics, providing essential bounds for its predictions in data-scarce permafrost regions.

Therefore, creating a differentiable permafrost-focused LSM is not an incremental step but a necessary foundation for the next generation of hybrid Earth system models. To address this gap, we introduce NoahPy: a differentiable LSM built upon a version of the Noah LSM already modified and validated for simulating permafrost thermos-hydrology on the Qinghai-Tibet Plateau (QTP). We have rewritten this permafrost-centric, Fortran-based model into a differentiable Python framework by encapsulating its governing partial differential equations within a Recurrent Neural Network (RNN) structure. This novel implementation preserves the complete mechanistic integrity of the physically-improved model while unlocking the full power of gradient-based optimization.

2 Material and methods

2.1 The modified Noah LSM

The Noah LSM (v3.4.1) (Chen et al., 1997) is a widely used model that simulates one-dimensional thermo-hydrological transport within the atmosphere-vegetation-soil continuum. It serves as the land-surface module in prominent systems like the Weather Research and Forecasting (WRF) model (Ek et al., 2003) and the Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004). In the Noah LSM, the governing equation for soil heat transfer is the one-dimensional heat conduction equation:

$$C_s \frac{\partial T_s}{\partial t} = \frac{\partial}{\partial z}\left(\lambda \frac{\partial T_s}{\partial z} \right) + Q \tag{1}$$

where T_s is the soil temperature (K), t is time (s), z is soil depth (m), C_s is the volumetric soil heat capacity ($J \cdot m^{-3} \cdot K^{-1}$), λ is the soil thermal conductivity ($W \cdot m^{-1} \cdot K^{-1}$), and Q represents the source/sink term ($W \cdot m^{-3}$), such as the latent heat of fusion during ice-water phase change. The soil heat capacity, C_s , is calculated as a weighted sum of its constituents:

$$C_s\left(\theta,\ \theta_{ice}\right) = \left(\theta - \theta_{ice}\right) C_w + \theta_{ice} C_{ice} + \left(1 - \theta_s\right) C_{soil} + \left(\theta_s - \theta\right) C_{air} \tag{2}$$

where θ is the volumetric water content ($m^3 \cdot m^{-3}$), θ_{ice} is the volumetric ice content ($m^3 \cdot m^{-3}$), θ_s is the saturated volumetric water content ($m^3 \cdot m^{-3}$), and C_w, C_{ice}, C_{soil}, and C_{air} are the heat capacities of water, ice, soil solids, and air, respectively.

Liquid water movement in the soil is simulated by the Richards' equation (Chen et al., 1996):

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[D(\theta) \frac{\partial \theta}{\partial z} \right] + \frac{\partial K(\theta)}{\partial z} + S(\theta) \tag{3}$$

where $D = K(\theta)\frac{\partial \Psi}{\partial \theta}$, known as the soil-water diffusivity ($m^2 \cdot s^{-1}$), K is the hydraulic conductivity ($m \cdot s^{-1}$), Ψ is the soil matric potential (m). S represents water sources and sinks (s^{-1}) (e.g., infiltration and evapotranspiration). The empirical soil hydraulic scheme proposed by Campbell (1974) is utilized to parameterize Ψ–θ and K–θ, relationships :

$$K(\theta) = K_s \left(\frac{\theta}{\theta_s} \right)^{2b+3} \tag{4}$$

$$\Psi(\theta) = \Psi_s \left(\frac{\theta}{\theta_s} \right)^{-b} \tag{5}$$

where K_s represent the saturated hydraulic conductivity ($m \cdot s^{-1}$), Ψ_s is the soil water potential at air entry (m), and b is an empirical parameter (dimensionless) related to the pore size distribution of the soil matrix.

For this study, we used a version of the Noah LSM specifically modified for permafrost applications (Chen et al., 2015; Wu et al., 2018), which improves upon the original model (Noah LSM v3.4.1) in several key ways. These modifications include an improved thermodynamic roughness length parameterization for sparse vegetation (Rodell et al., 2004) to correct the underestimation of ground heat flux, a new thermal conductivity scheme (Côté and Konrad, 2005) better suited for the coarse-grained, high-porosity soils common on the QTP, and an impedance factor related to ground ice content, which constrains the soil hydraulic conductivity to account for the impedance of water flow by ice (Zhang et al., 2007). The model's soil column was extended to a depth beyond the zero annual amplitude (~10 m for typical permafrost on the QTP (Zhao et al., 2010) and discretized into multiple, vertically heterogeneous soil layers. This modified Noah LSM has been successfully validated at the Tanggula (TGL) site and applied in previous studies of permafrost degradation on the QTP (Ji et al., 2022; Zhang et al., 2022a), confirming its robust simulation capabilities in permafrost environment.

2.2 Implementation of NoahPy

The implementation of NoahPy involves recasting the numerical solution of the modified Noah LSM's governing equations into a differentiable computational structure. We use the following partial differential equations (PDEs) set to describe the dynamic system of the modified Noah LSM:

$$\begin{cases} \frac{\partial}{\partial t} s(t,z) = F(s(t,z), u(t,z), \beta^F) \\ y(t,z) = G(s(t,z), u(t,z), \beta^G) \end{cases} \tag{6}$$

where, $s(t,z)$ represents the state vectors that vary in time t and space z (e.g. soil temperature profile), $u(t,z)$ is the input vector of external forcings (e.g., meteorological data), $y(t,z)$ is the output vector (e.g., simulated variables for validation), and β^F and β^G are the parameters associated with the state update function F and the output function G.

In the Noah LSM, the heat conduction (Equation 1) and Richards' (Equation 3) equations are solved using a finite-difference numerical approach. Following the spatial discretization scheme of Pan and Mahrt (1987) and the temporal scheme

of Kalnay and Kanamitsu (1988), the PDEs are expressed in terms of explicit coefficients and implicit states. After discretization, the PDEs can be converted into a system of algebraic equations, which is then efficiently solved using the tridiagonal matrix algorithm. To ensure numerical stability, this calculation is applied twice for each time step when infiltration fluxes are large (Zheng et al., 2015).

The discretized form of Richards' equation, for example, for each soil layer k and time step t is:

$$\frac{\theta_k^{t+1}-\theta_k^t}{\Delta t}=\frac{1}{\Delta z_k}\left[D(\theta_{k-1})\frac{\theta_{k-1}^{t+1}-\theta_k^{t+1}}{\Delta\widetilde{z_{k-1}}}-D(\theta_k)\frac{\theta_k^{t+1}-\theta_{k+1}^{t+1}}{\Delta\widetilde{z_k}}+K_{k-1}-K_k+S\right] \tag{7}$$

By letting $A=-\frac{D(\theta_{k-1})\Delta t}{\Delta z_k\Delta\widetilde{z_{k-1}}}$, $C=-\frac{D(\theta_k)\Delta t}{\Delta z_k\Delta\widetilde{z_k}}$, Equation 7 can be rearranged to:

$$A(\theta_{k-1}^{t+1}-\theta_{k-1}^t)+B(\theta_k^{t+1}-\theta_k^t)+C(\theta_{k+1}^{t+1}-\theta_{k+1}^t)=RHS \tag{8}$$

$$RHS=\frac{S+K_{k-1}-K_k}{\Delta z_k}\cdot\Delta t+A(\theta_k^t-\theta_{k-1}^t)+C(\theta_k^t-\theta_{k+1}^t), B=1-(A+C) \tag{9}$$

where Δz_k is the thickness of the k-th soil layer; and $\Delta\widetilde{z_k}$ is the distance between the centers of layer k and layer k+1. This equation can be rearranged into a tridiagonal system of linear equations, which is solved at each time step to update the soil moisture profile, θ^{t+1}:

$$\begin{bmatrix} B_1 & C_1 & 0 & 0 & 0 & \cdots & 0 \\ A_2 & B_2 & C_2 & 0 & 0 & \cdots & 0 \\ 0 & A_3 & B_3 & C_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & \cdots & 0 & A_{k-2} & B_{k-2} & C_{k-2} & 0 \\ 0 & \cdots & 0 & 0 & A_{k-1} & B_{k-1} & C_{k-1} \\ 0 & \cdots & 0 & 0 & 0 & A_k & B_k \end{bmatrix} \begin{bmatrix} \theta_1^{t+1}-\theta_1^t \\ \theta_2^{t+1}-\theta_2^t \\ \theta_3^{t+1}-\theta_3^t \\ \vdots \\ \theta_{k-2}^{t+1}-\theta_{k-2}^t \\ \theta_{k-1}^{t+1}-\theta_{k-1}^t \\ \theta_k^{t+1}-\theta_k^t \end{bmatrix} = \begin{bmatrix} RHS_1 \\ RHS_2 \\ RHS_3 \\ \vdots \\ RHS_{k-2} \\ RHS_{k-1} \\ RHS_k \end{bmatrix} \tag{10}$$

To make this process differentiable, we implemented the model within a RNN framework. A standard RNN updates an abstract hidden state, h_t, using a learned function (Figure 1a):

$$h_t=\sigma\left(W_h h_{(t-1)}+W_x x_t+b_h\right) \tag{11}$$

where σ is the nonlinear activation function; h_{t-1} and h_t are the hidden states at the previous and current time steps, respectively; W_h and W_x are the weight matrices applied to the previous hidden state and the current input vector x_t, respectively; and b_h is the bias vector.

In NoahPy, we replace this learned function with the entire physical time-step solution described above. The state of the system is a vector of physically meaningful variables, s_t (e.g., soil temperature, moisture), which is updated according to the model's deterministic physics (Figure 1b):

$$s_t = F_{Noah\ LSM}\left(s_{t-1}, x_t, \overrightarrow{\beta}\right) \tag{12}$$

where $F_{Noah\ LSM}$ represents the complete numerical solution for one time step, including the differentiable solver for the tridiagonal system derived from Equation 6; x_t is the meteorological forcing, and $\overrightarrow{\beta}$ is the set of model parameters. This is made possible by implementing every step of the numerical solution using the differentiable operations native to the PyTorch deep learning library (Paszke et al., 2019).

It is important to note that this reformulation includes all physical parameterizations, such as those for vegetation and snow processes shown in Figure 1c. While some of these processes are mathematically not strictly differentiable, re-implementing them within PyTorch ensures that a valid gradient can be computed for every operation via the automatic differentiation engine. This makes the model differentiable in the context of gradient-based optimization. A specific example of this is the handling of phase-dependent processes. The Noah LSM handles the latent heat of fusion using a source term method, as represented by the Q term in the heat conduction equation (Equation 1). This term explicitly calculates and applies the latent heat required to be released or absorbed to keep the soil temperature at the freezing point during a phase change. While this represents an abrupt physical transition, numerically, this is not a true discontinuity but is implemented as a conditional logic check. In NoahPy, this entire conditional logic is re-expressed using a chain of native, computationally differentiable PyTorch operations, primarily torch.where, torch.min, and torch.max. PyTorch's automatic differentiation engine is designed to backpropagate through these subgradients, which is the same fundamental principle that enables the training of neural networks with ReLU activations (Glorot et al.). This numerical implementation avoids a mathematical discontinuity. Therefore, PyTorch's autograd engine can compute a valid gradient through this logic.

By constructing the model in this way, the entire time-stepping simulation allows the gradient of any model output with respect to any parameter (β) to be calculated efficiently using the backpropagation through time (BPTT) (Werbos, 1990), powered by PyTorch's automatic differentiation engine. Furthermore, all operations in NoahPy are vectorized to maximize the parallel computing power of modern hardware.

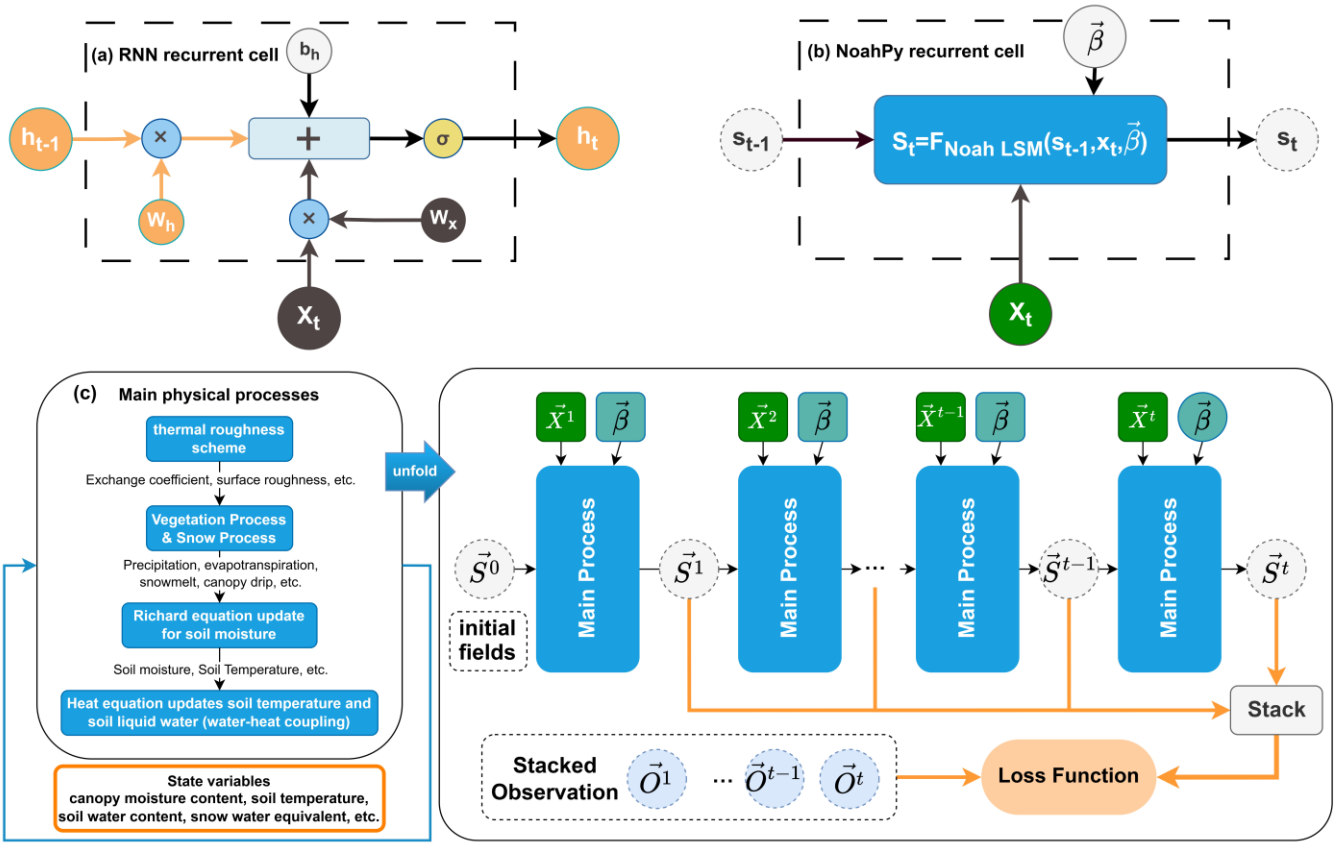

Figure 1. NoahPy architecture as a physics-based Recurrent Neural Network (RNN). (a) A standard RNN recurrent cell; (b) The NoahPy recurrent cell, which replaces the learned transformation with the physical model ($F_{\text{Noah LSM}}$); (c) The unfolded representation of the NoahPy simulation, where the model state (S) is updated at each time step. $\vec{\boldsymbol{X}}$, $\vec{\boldsymbol{S}}$, and $\vec{\boldsymbol{O}}$ represent the meteorological forcing, state, and observation vectors, respectively, and $\vec{\boldsymbol{\beta}}$ is the vector of model parameters.

2.3 Validations

2.3.1 Validation of numerical equivalence

The first validation step was to confirm that NoahPy, written in Python, accurately reproduces the numerical output of the original Fortran-based modified Noah LSM. This benchmark test ensures that the model reformulation process did not introduce numerical artifacts. The experiment was conducted at three randomly selected grid cells on the QTP: Grid1 (28.75°N, 93.85°E), Grid2 (34.75°N, 98.25°E) and Grid3 (37.55°N, 100.55°E). Both models were driven by the China Meteorological Forcing Dataset (ITP-forcing) (He et al., 2020) for the period of 2000-2010. The year 1999 was used as a spin-up period (repeating for 500 years) to allow the model to reach equilibrium, and the model states at the end of this period

were used as the initial conditions for the formal simulation. For both models, soil types were defined using the MSTD dataset (Wu and Nan, 2016), and vegetation types were based on the 1:1,000,000 China Vegetation Type Map (Zhang, 2007). Since the goal was a direct numerical comparison, model parameters were assigned using the default lookup table values corresponding to the soil and vegetation types. The soil column was configured with 18 layers extending to a depth of 15.2 m.

To quantify the agreement between the two models, we used three statistical metrics: Bias, Pearson correlation coefficient (Corr), and the Nash-Sutcliffe Efficiency coefficient (NSE):

$$Bias = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - y_i^*\right) \tag{13}$$

$$Corr = \frac{\sum_{i=1}^{N}\left[\left(y_i - \bar{y}\right)\left(y_i^* - \overline{y^*}\right)\right]}{\sqrt{\sum_{i=1}^{N}\left[\left(y_i - \bar{y}\right)^2\right]}\sqrt{\sum_{i=1}^{N}\left[\left(y_i^* - \overline{y^*}\right)^2\right]}} \tag{14}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left[\left(y_i - \bar{y}\right)^2\right]}{\sum_{i=1}^{N}\left[\left(y_i^* - \overline{y^*}\right)^2\right]} \tag{15}$$

where, y_i is a value from the NoahPy simulation time series, y_i^* is the corresponding value from the modified Noah LSM simulation,$\bar{y}$ and $\overline{y^*}$are the mean values of their respective time series, and N is the total number of samples.

2.3.2 Validation of backpropagation capability

To validate NoahPy's capability for backpropagation-driven parameter optimization, we conducted an experiment using observational data from the TGL permafrost site on the QTP. The model was driven by daily meteorological observations from the TGL station from April 1, 2007 to December 31, 2010. These data included air temperature, wind speed, relative humidity, incoming shortwave and longwave radiation, and precipitation. In-situ observations of active layer soil temperature and liquid water content from the site were used to constrain the model during optimization. The dataset was split into a training period (April 1, 2007 to December 31, 2009) and a validation period (January 1, 2010 to December 31, 2010). The NoahPy soil column was discretized into 20 layers to match the observation depths at the site. This included ten shallow, higher-resolution layers (at 0.045, 0.091, 0.166, 0.289, 0.493, 0.829, 1.2, 1.6, 2.0, and 2.4 m) to capture rapid variations near the surface, and

ten deeper layers (2.8, 3.8, 4.8, 5.8, 6.8, 7.8, 8.8, 10.8, 12.8, and 14.8 m) extending to 14.8 m. The lower boundary of the simulation domain was set to a depth of 40 m, with the boundary temperature condition prescribed according to previous studies (Chen et al., 2015).

We selected four key soil hydraulic parameters, known to be highly sensitive to liquid water content (Brandhorst and Neuweiler, 2023; Szabó et al., 2024; Teuling et al., 2009), as the target for optimization: saturated hydraulic conductivity (K_s), saturated water content (θ_s), soil matric potential at air entry (ψ_s), and the pore-size distribution index (b). The allowable ranges for these parameters, drawn from previous studies (Rosero et al., 2009; Stuurop et al., 2021; Li et al., 2019; Wang et al., 2021), are provided in Table 1. Initial values were chosen randomly within these bounds. To ensure physical realism, we imposed a constraint that parameter values for the same soil type could not vary by more than 10% across different depths (Zhao et al., 2023).

The observational data for this study extend to a maximum depth of 2.45 m, corresponding to the model's 10th soil layer. Therefore, simulated liquid water content from the top ten model layers was interpolated to the measurement depths. The NSE between the interpolated simulations and the observations was used as the loss function to be maximized. We used the widely adopted Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.0005 and default decay rates of 0.9 and 0.999.To improve convergence, a ReduceLROnPlateau learning rate scheduler was implemented. This scheduler monitored the NSE on the validation set and automatically reduced the learning rate by a factor of 0.1 if no improvement was observed for ten consecutive epochs. The training was run for a maximum of 300 epochs, with a minimum learning rate of 1×10^{-6} to prevent stagnation. The agreement between the optimized model simulations and the observations was quantified using the NSE, correlation coefficient, and Root Mean Square Error (RMSE).

Table 1. Target parameters to be optimized by backpropagation and their value ranges

Parameter	Symbol (Unit)	Value Range
saturated hydraulic conductivity	K_s ($m \cdot s^{-1}$)	10^{-7}-6×10^{-3}
saturated water content	θ_s ($m^3 \cdot m^{-3}$)	0.3-0.65
soil matric potential at air entry	ψ_s (m)	0.01-0.65
pore-size distribution	b (Dimensionless)	2.5-12

2.3.3 Performance comparison with traditional optimization

To demonstrate the advantages of a differentiable modeling approach, we compared the performance of NoahPy against both the original and modified Noah LSMs when calibrated with a traditional, widely used optimization algorithm. We evaluated three distinct model-optimizer combinations: NoahPy optimized with the gradient-based Adam optimizer; the modified Noah LSM calibrated with the Shuffled Complex Evolution (SCE-UA) algorithm (Duan et al., 1994); and the original

Noah LSM (v3.4.1) calibrated with the SCE-UA algorithm. The SCE-UA is a widely used global optimization algorithm that does not require gradient information. Its primary strength is its robust global search capability, which helps it avoid getting trapped in local optima by systematically exploring the parameter space (Rahnamay Naeini et al., 2019). The model configurations, forcing data, and target parameters for all three setups were identical to those described in Section 2.3.2. A key difference is that the original Noah LSM does not account for vertical soil heterogeneity; therefore, its soil profile was configured uniformly using the properties of the surface layer. For a robust comparison, each optimization algorithm was run ten times with a maximum of 500 iterations.

In addition to the NSE, we used the Kling-Gupta Efficiency (KGE) as a more comprehensive performance metric. KGE provides a multi-faceted assessment by decomposing performance into three distinct components:

$$KGE = 1 - \sqrt{(Corr - 1)^2 + (\alpha - 1)^2 + (\gamma - 1)^2} \tag{16}$$

where, *Corr* is the Pearson correlation coefficient between simulated and observed values, α is the bias ratio (mean of simulated values / mean of observed values), and γ is the variability ratio (coefficient of variation of simulated values / coefficient of variation of observed values). To determine if the performance differences among the three model setups were statistically significant, we employed a two-step non-parametric testing procedure on the KGE values from all soil depths. First, the Friedman test was used to assess whether any significant differences existed within the group of three models. If the Friedman test returned a p-value < 0.05, we then performed the Dunn's post-hoc test for pairwise comparisons to identify which specific model pairs differed significantly from one another. A p-value < 0.05 in the Dunn's test was considered a statistically significant difference in performance.

3 Results

3.1 Numerical equivalence with the modified Noah LSM

The validation confirms that NoahPy successfully replicates the numerical behaviour of the Fortran-based modified Noah LSM. As shown in the scatter plots in Figure 2, the simulated daily soil temperature and liquid water content from NoahPy exhibit a near-perfect 1:1 relationship with the outputs from the modified LSM across all tested depths (0.1, 0.5, 0.8, 1.3, and 2.5 m) aggregated from three randomly chosen grid cells on the QTP. The performance is exceptionally strong, with NSE coefficients greater than 0.999 and near-zero bias (<0.01) for both variables at every depth.

A minor degree of scatter is visible in the soil moisture comparisons (Figure 2b, d, f, h, j), which is not present in the soil temperature results. These small deviations are likely attributable to minor differences in floating-point arithmetic and numerical precision between the Python/PyTorch environment and the original Fortran compiler. Importantly, NoahPy maintains this high accuracy in deeper soil layers, with no amplification of numerical errors with depth. This demonstrates the

high numerical stability of the NoahPy implementation and confirms that it serves as a faithful and reliable replacement for the original model.

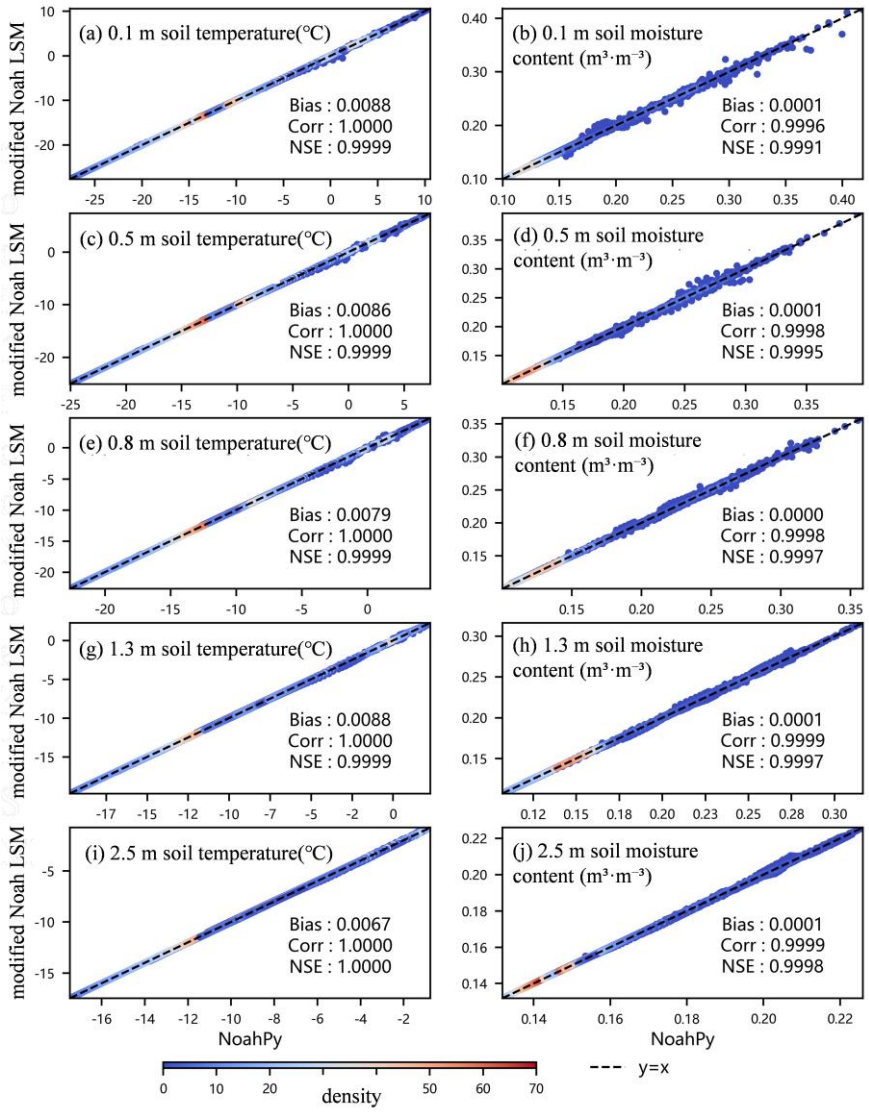

Figure 2. Comparison of NoahPy and modified Noah LSM outputs for soil temperature and moisture. The density scatter plots compare daily model outputs at five different soil depths (0.1, 0.5, 0.8, 1.3, and 2.5 m), aggregated from three randomly chosen grid cells (28.75°N, 93.85°E; 34.75°N, 98.25°E; 37.55°N, 100.55°E) on the Tibetan Plateau (QTP). The dashed line represents perfect agreement (y=x). Inset values show the Bias, correlation (Corr), and NSE.

3.2 Performance of the calibrated NoahPy at the Tanggula site

The gradient-based optimization process effectively calibrated the NoahPy model parameters. The training process demonstrates rapid convergence, with the NSE for soil liquid water increasing from an initial value of -0.2 to an optimal value of 0.84 (Figure 3). Correspondingly, the RMSE steadily decreases. This result successfully validates that NoahPy's differentiable framework allows for the effective use of backpropagation to optimize model parameters against observational data.

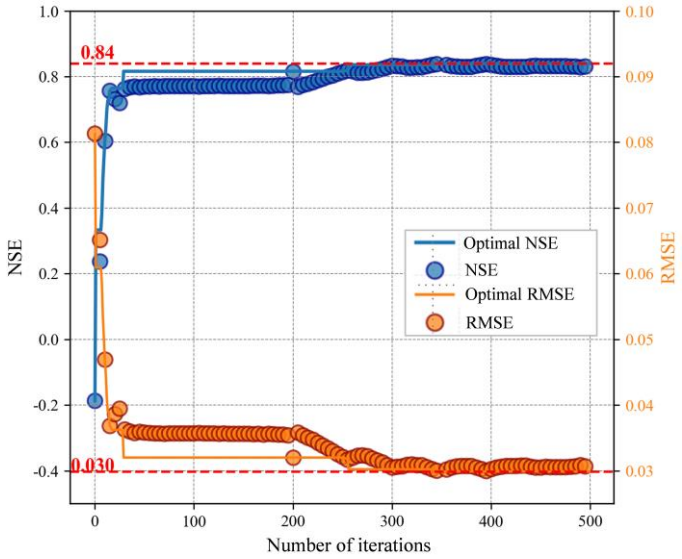

Figure 3. Training convergence for the soil liquid water simulation at the Tanggula site. The plot shows the improvement in the Nash-Sutcliffe Efficiency (NSE, blue line) and the corresponding reduction in the Root Mean Square Error (RMSE, orange line) over 500 optimization iterations. The dashed red lines mark the best performance achieved.

After calibration, NoahPy's simulations showed excellent agreement with the observed data at the TGL site during both the calibration (2007-2009) and validation (2010) periods (Figure 4). The model accurately reproduced the seasonal cycle of soil temperature at all depths. For most layers, the NSE values exceeded 0.9, and the RMSE decreased with depth, reflecting the reduced temperature variability in deeper soil. 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. Additionally, anomalous fluctuations were observed in the measured deep soil temperatures (1.05 m and 2.45 m) during the summer of 2009 (Figure

4d, e). Given that deep soil temperatures should respond slowly to short-term atmospheric changes, these fluctuations are likely attributable to instrumental error.

While more complex than temperature, the dynamics of soil liquid water were also well-captured, with NSE values exceeding 0.7 and RMSE values below 0.05 $m^3 m^{-3}$ for most layers. The model successfully simulated soil moisture responses to freeze-thaw cycles and summer precipitation events, particularly in the shallow soil layers (Figure 4f, g). However, several discrepancies were noted, particularly in deeper soil. Simulations at depths of 1.05 m and 2.45 m deviate more pronouncedly from the measured data (Figure 4i, j). The model tended to overestimate liquid water content during the winter freezing period at some depths (Figure 4h, i). This can be attributed to the model's hydraulic parameterization scheme, which is based on the Campbell formulation; this approach neglects the effects of ice suction and effective porosity. Omitting these mechanisms, which influence soil water redistribution at the freezing front, can lead to an overestimation of liquid water content during winter (Zhao et al., 2023). Additionally, some observations appear anomalous. For example, the measured unfrozen water content in winter drops to exactly zero at 0.4 m and 1.05 m, which is physically unlikely and suggests potential instrument error at low moisture levels. Similarly, sharp, isolated increases in measured water content at deeper layers during the summer of 2009 (Figure 4h, i, j) without corresponding signals in the layers above suggest these are likely not caused by surface infiltration and may also be data artifacts.

Despite the well-diagnosed limitations of specific model parameterizations and potential artifacts in the observational data, the results for all soil depths demonstrate that the calibrated NoahPy model reliably reproduces the key seasonal dynamics of soil temperature and liquid water during complex freeze-thaw cycles at the TGL site.

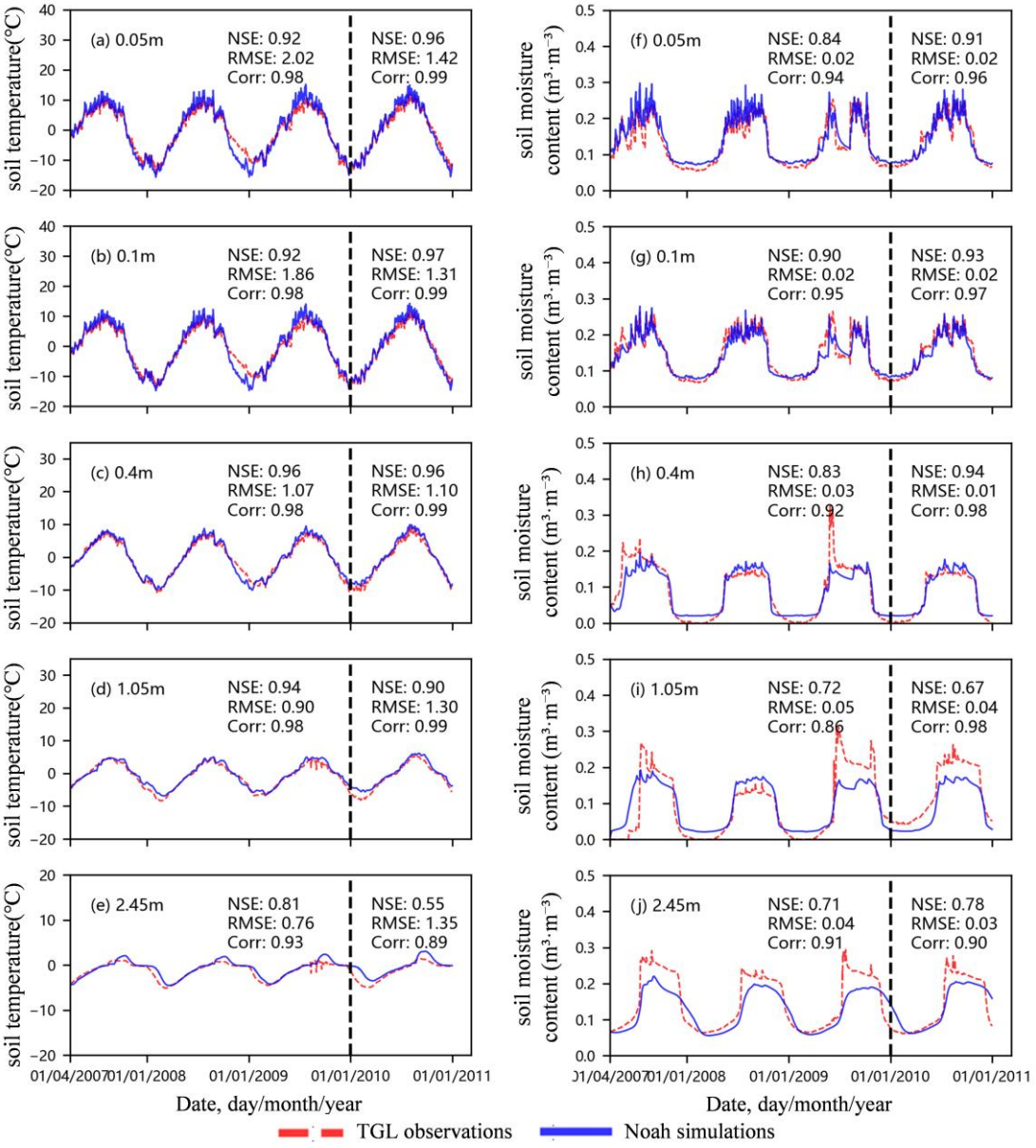

Figure 4. Simulated and observed daily soil temperature and liquid soil water content at various depths (daily; 0.05, 0.1, 0.4, 1.05, and 2.45m) for the Tanggula (TGL) site. The vertical black dashed line separates the calibration period (April 1, 2007–December 31, 2009) from the validation period (January 1, 2010–December 31, 2010). Inset text in each panel provides the NSE, RMSE, and correlation coefficient (Corr) for both periods.

3.3 Comparative performance evaluation

The primary advantage of the differentiable approach is evident in the parameter optimization process. NoahPy paired with the Adam optimizer converges extremely rapidly, reaching a high level of accuracy within only 100 iterations (Figure 5). This is due to the Adam optimizer's use of gradient information and an adaptive learning rate. In contrast, the traditional SCE-UA algorithm applied to the Noah and modified Noah LSMs converges much more slowly, requiring significantly more iterations to approach an optimal solution (Figure 5). While the SCE-UA algorithm's strength is its global search capability, which helps it avoid getting trapped in local optima, its convergence becomes prohibitively slow in high-dimensional parameter spaces, requiring significantly more iterations to find a solution. Furthermore, the gradient-based approach demonstrates greater stability. The shaded 95% uncertainty band around the convergence trajectory for NoahPy is visibly narrower than for the SCE-UA method (Figure 5), indicating that the Adam optimizer finds a robust solution more consistently across repeated runs.

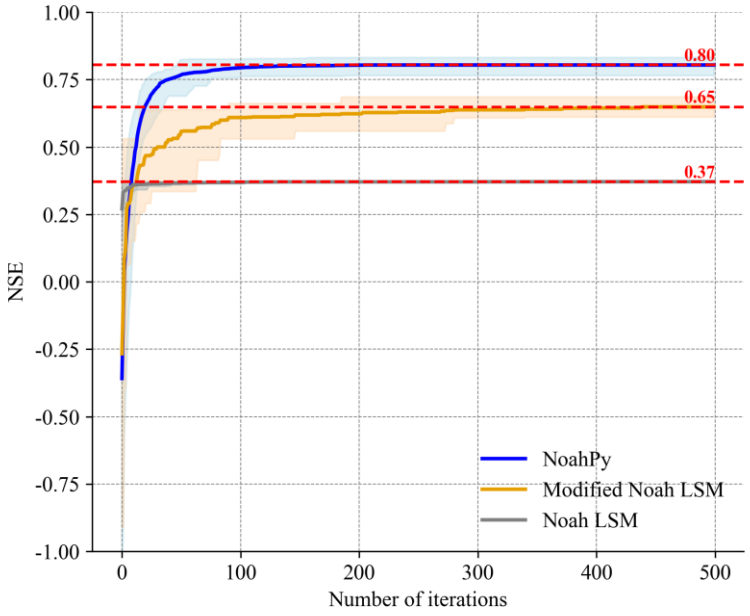

Figure 5. Convergence of NoahPy, the modified Noah LSM, and the original Noah LSM in terms of NSE. Each line represents the mean NSE from 10 optimization runs, with the shaded area indicating the 95% uncertainty band. NoahPy was optimized with the Adam optimizer, while the other two models were calibrated with the SCE-UA algorithm.

When comparing the calibrated models' ability to simulate soil temperature (Figure 6), all three setups perform well in the shallow soil layers (0.05 m and 0.4 m), with NSE values exceeding 0.9. However, a major performance gap appears in the deep soil (2.45 m). The original Noah LSM, which neglects vertical soil heterogeneity, exhibits a pronounced cold bias, with

an RMSE of 1.68°C (Figure 6i). NoahPy and the modified Noah LSM, which both account for varying soil layers, perform significantly better, with RMSE values of 0.51°C (Figure 6c) and 0.85°C (Figure 6f), respectively. In essence, the error is magnified with depth because the impact of incorrect thermal properties is compounded over the longer time and distance it takes for heat to travel to the deep soil.

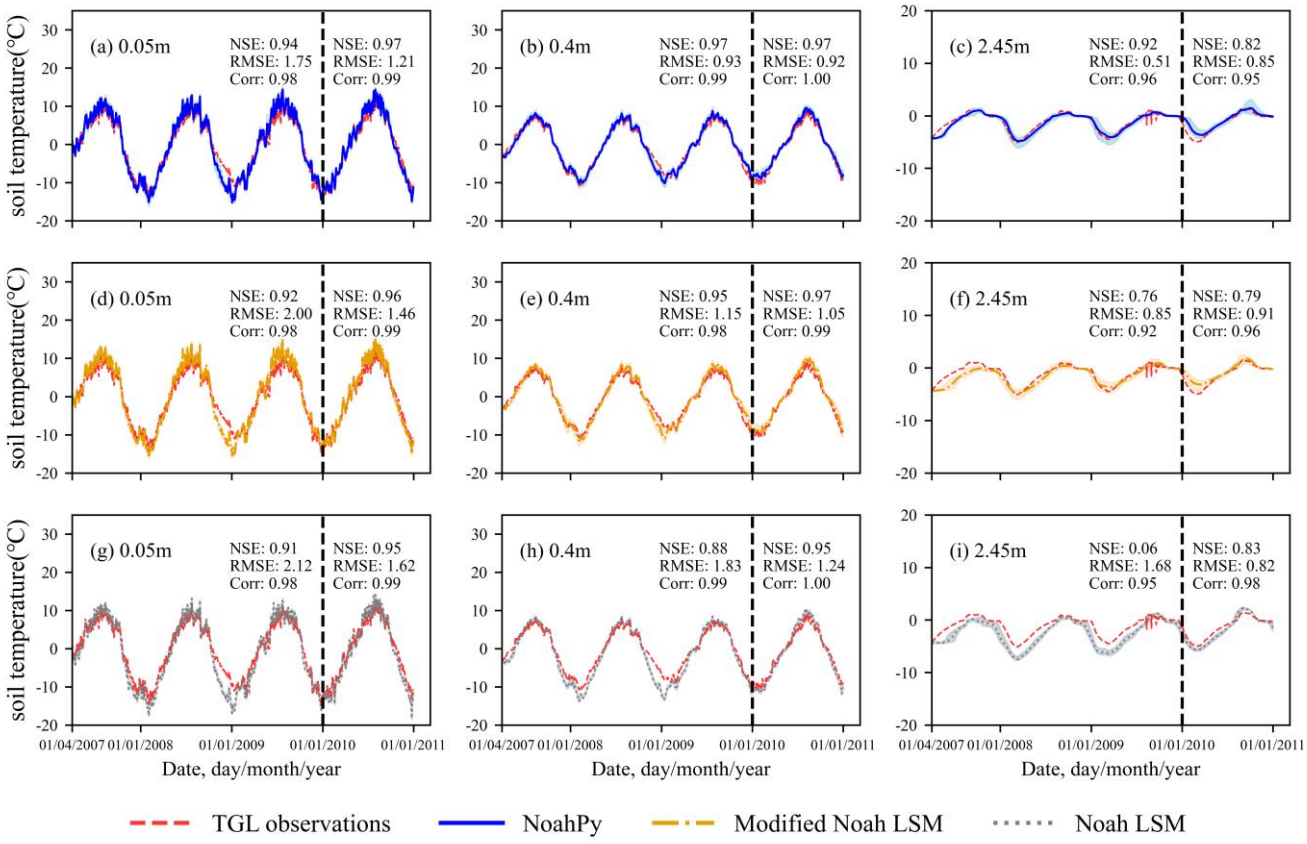

Figure 6. Comparison of calibrated model performance for daily soil temperature at the TGL site. Each panel compares in-situ observations (red dashed line) against the simulations from the three calibrated models at a specific depth. The models are NoahPy (optimized with Adam; blue line and shading), the modified Noah LSM (calibrated with SCE-UA; orange line and shading), and the original Noah LSM (calibrated with SCE-UA; gray dashed line and shading). The shaded areas represent the 95% uncertainty band from 10 repeated optimization runs. The vertical dashed line separates the calibration and validation periods.

The results for soil liquid water simulation show an even starker contrast (Figure 7). Both NoahPy and the modified Noah LSM produce satisfactory results, with RMSE below 0.05 $m^3 m^{-3}$ across all three layers. These models accurately capture the key seasonal dynamics, including soil moisture fluctuations driven by summer precipitation and the rapid changes associated with freeze-thaw phase transitions, which align well with observations. The original Noah LSM, however, performs poorly. It

fails to capture moisture fluctuations from summer rainfall and shows significant biases in winter. Its performance deteriorates sharply with depth, with the NSE value dropping to -0.09 in the deepest layer (2.45 m) (Figure 7i). This negative NSE reflects a substantial underestimation of the liquid water increase during the spring thaw. This finding is consistent with previous research (Wu et al., 2018).

A Friedman test performed on the KGE values for all models (Table 2) confirmed a statistically significant difference in their overall performance ($p \approx 0$). A subsequent Dunn's post-hoc test revealed that both NoahPy and the modified Noah LSM performed significantly better than the original Noah LSM. Interestingly, the statistical test showed no significant difference between NoahPy and the modified Noah LSM ($p = 0.1659$). This is expected, as they share identical physics. However, NoahPy consistently demonstrated practical advantages in performance. As shown in Figure 5, NoahPy converges markedly faster with the Adam optimizer, approaching its optimal solution in roughly 100 iterations, whereas the modified Noah LSM requires substantially more iterations to converge with the SCE-UA algorithm. Furthermore, NoahPy's final calibrated simulations have noticeably lower uncertainty (i.e., smaller shaded bands in Figures 6 and 7) compared to the modified Noah LSM, particularly for winter liquid water content (Figure 7a,c vs. (Figure 7d,f). This lower uncertainty is a direct result of the more stable and efficient optimization provided by the gradient-based Adam algorithm, highlighting a key practical advantage of the differentiable modeling approach.

Table 2. Mean Kling-Gupta Efficiency (KGE) values for the three calibrated models. Values represent the mean KGE from 10 repeated optimization runs for NoahPy, the modified Noah LSM, and the original Noah LSM.

Variable	Depth(m)	NoahPy	Modified Noah LSM	original Noah LSM
soil temperature	0.05	0.83	0.79	0.79
	0.1	0.86	0.81	0.8
	0.4	0.93	0.91	0.74
	1.05	0.89	0.85	0.57
	2.45	0.93	0.83	0.28
soil liquid water content	0.05	0.91	0.82	0.51
	0.1	0.95	0.88	0.69
	0.4	0.64	0.51	0.23
	1.05	0.64	0.52	0.31
	2.45	0.8	0.56	0.27

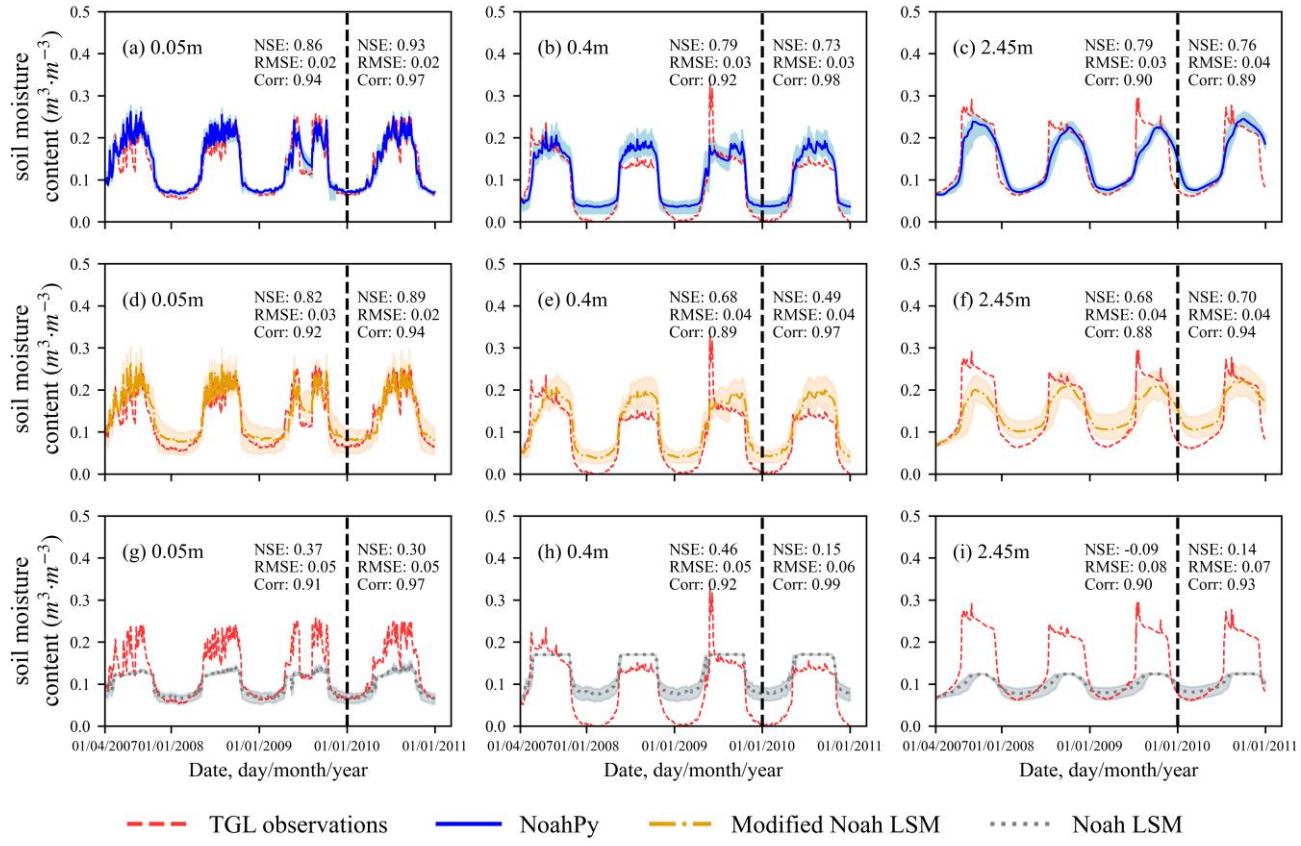

Figure 7. Comparison of calibrated model performance for daily liquid soil water at the TGL site. Each panel compares in-situ observations (red dashed line) against the simulations from the three calibrated models at a specific depth. The models are NoahPy (optimized with Adam; blue line and shading), the modified Noah LSM (calibrated with SCE-UA; orange line and shading), and the original Noah LSM (calibrated with SCE-UA; gray dashed line and shading). The shaded areas represent the 95% uncertainty band from 10 repeated optimization runs. The vertical dashed line separates the calibration and validation periods.

4 Discussion

This study successfully demonstrated the development and application of NoahPy, a differentiable land surface model for permafrost. Our results confirm that this re-implementation not only preserves this enhanced physical integrity of the modified Noah LSM but also unlocks a parameter optimization workflow that is significantly faster and more robust than traditional methods. The successful calibration and diagnostic analysis in this study highlight the theoretical merits of our "glass-box" approach. A common alternative for making a physical model compatible with machine learning workflows is to develop a surrogate model: a neural network trained to mimic the input-output behavior of the original, non-differentiable code (Razavi

et al., 2012). While easier to implement, this approach treats the model as a "black box" and suffers from the curse of dimensionality (Asher et al., 2015). As the number of parameters grows, the required simulations increase exponentially, making surrogates infeasible for complex LSMs. While such a surrogate could potentially replicate the final simulation results, it obscures the internal model dynamics. In contrast, the full interpretability of NoahPy allowed us to diagnose specific physical process errors, such as the cold bias from the simplified snow scheme and the overestimation of winter liquid water due to missing cryosuction physics. This ability to directly attribute simulation errors to specific physical parameterizations is a fundamental advantage of the differentiable physics-based approach and is essential for targeted scientific model improvement.

Gradient-based optimization is particularly advantageous when coupling NoahPy with neural networks for hybrid modeling. It allows the simultaneous calibration of a large number of model parameters, which would be prohibitively difficult using traditional gradient-free methods such as SCE-UA. While SCE-UA can perform a global search and avoid local minima, its performance degrades substantially in high-dimensional parameter spaces. By contrast, optimizers like Adam exploit precise gradients to iteratively improve parameter values, facilitating effective end-to-end training of hybrid systems. It should be noted that we do not provide absolute comparisons of computational speed, as differences in model implementation (Fortran vs Python) and numerical schemes limit direct benchmarking. Instead, the focus here is on the iterative optimization capability of gradient-based methods, which underpins the scalability and feasibility of hybrid training strategies.

This study has two primary limitations. First, while successfully validated at the Tanggula site on the Qinghai-Tibet Plateau, the performance and applicability of NoahPy in other permafrost regions with different characteristics (e.g., the ice-rich Yedoma of Siberia or the boreal forests of North America) have yet to be confirmed. Second, NoahPy inherits the known physical deficiencies of its parent Noah LSM, including a simplistic snow scheme and the omission of processes critical to permafrost carbon cycling, such as the effects of soil organic matter, convective heat transfer, and abrupt thaw dynamics.

However, these limitations highlight its future potential and intended purpose. The framework presented here is not intended as a final product, but as a flexible and extensible foundation for the community to address these very issues. This framework holds promise for addressing challenges in permafrost domain, where parameterization for key soil properties in permafrost environment such as Qinghai-Tibet Plateau (QTP) like thermal conductivity (Ji et al., 2024), hydraulic conductivity (Hu et al., 2023), and matric potential (Zhao et al., 2023) may be incomplete. While NoahPy, in its current form, inherits the physical limitations of its parent model, its true power lies in its potential as a foundational framework for a new generation of hybrid models.The NoahPy framework allows for coupling with external machine learning models that can learn the complex mapping between environmental covariates (e.g., topography, vegetation, soil type) and the model's physical parameters (such as hydraulic and thermal parameters) from direct observations (e.g., soil temperature, soil moisture content). This could dramatically improve the spatial transferability of parameters across diverse regions, reducing the reliance on costly site-specific calibration and mitigating parameter uncertainty, a key challenge in permafrost modelling (Harp et al., 2016; Dai et al., 2019). The hybrid, seamless physics-machine learning models coupling enabled by automatic differentiation also allows for targeted replacement of model components. For instance, empirical parameterizations where physical knowledge is weak,

such as the Campbell-based hydraulic scheme, can be replaced by an embedded neural network. In such a hybrid mode, the neural network can learn more complex and accurate relationships from data, while the surrounding physical equations ensure its predictions remain constrained by fundamental laws like the conservation of mass and energy. By recasting a permafrost-capable LSM into the deep learning ecosystem, we have created a tool that can leverage the rapid advancements in computational hardware (e.g., GPUs, TPUs) and software (Sevilla et al., 2022; Kochkov et al., 2024). 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 of the future of the cryosphere.

5 Conclusions

In this study, we developed NoahPy, a differentiable land surface model specifically improved for permafrost thermo-hydrology. We successfully recast the widely-used, Fortran-based Noah LSM into a "glass-box" Python framework that is both physically interpretable and implements differentiable operations for gradient-based optimization. Based on our results, we draw the following key conclusions:

(1) NoahPy faithfully reproduces the numerical behaviour of the permafrost-specific modified Noah LSM. Validations show a very close match, with NSE values exceeding 0.99 for both soil temperature and liquid water across all soil layers, confirming the fidelity of the model's re-implementation.

(2) The differentiable framework enables robust, gradient-based parameter optimization. Validation at a permafrost site on the QTP demonstrates that NoahPy can effectively use backpropagation to learn from observational data. The resulting calibrated model shows strong performance, achieving NSE values above 0.9 for soil temperature and 0.8 for liquid water.

(3) The NoahPy-Adam workflow is superior to traditional calibration methods. The combination of the differentiable model with a gradient-based optimizer (Adam) results in a parameter optimization that is significantly faster, more stable, and yields final simulations with lower uncertainty compared to the traditional SCE-UA algorithm.

This work delivers a foundational tool that was previously missing for the permafrost community. It closes the technical gap that has hindered the development of deeply-integrated hybrid models for the cryosphere. This study thus lays the necessary groundwork for future AI-based models that aim to lower uncertainty and deliver more credible predictions of permafrost's response to a changing climate.

Code and data availability. The NoahPy model code used in this study is available at https://github.com/nanzt/NoahPy, and the exact version used to generate the results presented here is archived on Zenodo (Tian and Nan, 2025a, https://doi.org/10.5281/zenodo.16530326). The original Noah LSM (v3.4.1) code used in this study is available at https://ral.ucar.edu/model/unified-noah-lsm (last access: August 28, 2025). The modified version of Noah LSM code is available at https://doi.org/10.17605/osf.io/g7jqr (Zhang et al., 2022b). The simulation data generated in this study are available on Figshare (Tian and Nan, 2025b, https://doi.org/10.6084/m9.figshare.29988163).

Author contributions. **W.T.:** Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **H.Y.**: Formal analysis, Validation. **S.Z.:** Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. **Y.C.**: Formal analysis. **W.Y.**: Investigation, Writing – original draft, Writing – review & editing. **J.X.**: Writing – original draft, Writing – review & editing. **Z.N.:** Conceptualization, Funding acquisition, Methodology, Resources, Writing – original draft, Writing – review & editing.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements: This work is supported by the National Key Research and Development Program of China (No. 2022YFF0711703) and National Natural Science Foundation of China (No. 42171125, 42571149). We sincerely thank the Cryosphere Research Station on the Qinghai–Tibet Plateau, CAS (http://www.crs.ac.cn/) for providing the TGL site observation data used in this study.

References

Asher, M. J., Croke, B. F. W., Jakeman, A. J., and Peeters, L. J. M.: A review of surrogate models and their application to groundwater modeling, Water Resources Research, 51, 5957-5973, https://doi.org/10.1002/2015WR016967, 2015.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533-538, https://doi.org/10.1038/s41586-023-06185-3, 2023.

Bonavita, M. and Laloyaux, P.: Machine Learning for Model Error Inference and Correction, Journal of Advances in Modeling Earth Systems, 12, e2020MS002232, https://doi.org/10.1029/2020MS002232, 2020.

Brandhorst, N. and Neuweiler, I.: Impact of parameter updates on soil moisture assimilation in a 3D heterogeneous hillslope model, Hydrology and Earth System Sciences, 27, 1301-1323, https://doi.org/10.5194/hess-27-1301-2023, 2023.

Campbell, G. S.: A simple method for determining unsaturated conductivity from moisture retention data, Soil Science, 117, https://doi.org/10.1097/00010694-197406000-00001, 1974.

Chen, F., Janjić, Z., and Mitchell, K.: Impact of Atmospheric Surface-layer Parameterizations in the new Land-surface Scheme of the NCEP Mesoscale Eta Model, Boundary-Layer Meteorology, 85, 391-421, https://doi.org/10.1023/A:1000531001463, 1997.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q. Y., Ek, M., and Betts, A.: Modeling of land surface evaporation by four schemes and comparison with FIFE observations, Journal of Geophysical Research: Atmospheres, 101, 7251-7268, https://doi.org/10.1029/95JD02165, 1996.

Chen, H., Nan, Z., Zhao, L., Ding, Y., Chen, J., and Pang, Q.: Noah Modelling of the Permafrost Distribution and Characteristics in the West Kunlun Area, Qinghai-Tibet Plateau, China, Permafrost and Periglacial Processes, 26, 160-174, https://doi.org/10.1002/ppp.1841, 2015.

Chen, M., Qian, Z., Boers, N., Jakeman, A. J., Kettner, A. J., Brandt, M., Kwan, M.-P., Batty, M., Li, W., Zhu, R., Luo, W., Ames, D. P., Barton, C. M., Cuddy, S. M., Koirala, S., Zhang, F., Ratti, C., Liu, J., Zhong, T., Liu, J., Wen, Y., Yue, S., Zhu, Z., Zhang, Z., Sun, Z., Lin, J., Ma, Z., He, Y., Xu, K., Zhang, C., Lin, H., and Lü, G.: Iterative integration of deep learning in hybrid Earth surface system modelling, Nature Reviews Earth & Environment, 4, 568-581, https://doi.org/10.1038/s43017-023-00452-7, 2023.

Côté, J. and Konrad, J.-M.: A generalized thermal conductivity model for soils and construction materials, Canadian Geotechnical Journal, 42, 443-458, https://doi.org/10.1139/t04-106, 2005.

Dai, Y., Xin, Q., Wei, N., Zhang, Y., Shangguan, W., Yuan, H., Zhang, S., Liu, S., and Lu, X.: A Global High-Resolution Data Set of Soil Hydraulic and Thermal Properties for Land Surface Modeling, Journal of Advances in Modeling Earth Systems, 11, 2996-3023, https://doi.org/10.1029/2019MS001784, 2019.

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.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD003296, 2003.

Feng, D., Liu, J., Lawson, K., and Shen, C.: Differentiable, Learnable, Regionalized Process-Based Models With Multiphysical Outputs can Approach State-Of-The-Art Hydrologic Prediction Accuracy, Water Resources Research, 58, e2022WR032404, https://doi.org/10.1029/2022WR032404, 2022.

Glorot, X., Bordes, A., and Bengio, Y.: Deep Sparse Rectifier Neural Networks, In Proceedings of the fourteenth international conference on artificial intelligence and statistics, 315-323, 2011.

Harp, D. R., Atchley, A. L., Painter, S. L., Coon, E. T., Wilson, C. J., Romanovsky, V. E., and Rowland, J. C.: Effect of soil property uncertainties on permafrost thaw projections: a calibration-constrained analysis, The Cryosphere, 10, 341-358, https://doi.org/10.5194/tc-10-341-2016, 2016.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., and Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China, Scientific Data, 7, 25, https://doi.org/10.1038/s41597-020-0369-y, 2020.

Hu, G., Zhao, L., Li, R., Park, H., Wu, X., Su, Y., Guggenberger, G., Wu, T., Zou, D., Zhu, X., Zhang, W., Wu, Y., and Hao, J.: Water and heat coupling processes and its simulation in frozen soils: Current status and future research directions, Catena, 222, https://doi.org/10.1016/j.catena.2022.106844, 2023.

Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., and Saynisch-Wagner, J.: Towards neural Earth system modelling by integrating artificial intelligence in Earth system science, Nature Machine Intelligence, 3, 667-674, https://doi.org/10.1038/s42256-021-00374-3, 2021.

Ji, H., Fu, X., Nan, Z., and Zhao, S.: An effective medium theory-based unified model for estimating thermal conductivity of unfrozen and frozen soils, CATENA, 239, 107942, https://doi.org/10.1016/j.catena.2024.107942, 2024.

Ji, H., Nan, Z., Hu, J., Zhao, Y., and Zhang, Y.: On the Spin-Up Strategy for Spatial Modeling of Permafrost Dynamics: A Case Study on the Qinghai-Tibet Plateau, Journal of Advances in Modeling Earth Systems, 14, e2021MS002750, https://doi.org/10.1029/2021MS002750, 2022.

Kalnay, E. and Kanamitsu, M.: Time Schemes for Strongly Nonlinear Damping Equations, Monthly Weather Review, 116, 1945-1958, https://doi.org/10.1175/1520-0493(1988)116<1945:TSFSND>2.0.CO;2, 1988.

Kingma, D. and Ba, J.: Adam: A Method for Stochastic Optimization, Computer Science, https://doi.org/10.48550/arXiv.1412.6980, 2014.

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., and Hoyer, S.: Neural general circulation models for weather and climate, Nature, 632, 1060-1066, https://doi.org/10.1038/s41586-024-07744-y, 2024.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning skillful medium-range global weather forecasting, Science, 382, 1416-1421, https://doi.org/10.1126/science.adi2336, 2023.

Li, P., Wang, D., Ding, C., and Reng, Y.: Distribution characteristics of soil saturated hydraulic conductivity and soil bulk density in a small watershed in the alpine zone of the Loess Plateau, Science of Soil and Water Conservation, 17, 9-17, https://doi.org/10.16843/j.sswc.2019.04.002, 2019.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, Water Resources Research, 57, e2020WR028091, https://doi.org/10.1029/2020WR028091, 2021.

O'Loughlin, R. J., Li, D., Neale, R., and O'Brien, T. A.: Moving beyond post hoc explainable artificial intelligence: a perspective paper on lessons learned from dynamical climate modeling, Geoscientific Model Development, 18, 787-802, https://doi.org/10.5194/gmd-18-787-2025, 2025.

Obu, J.: How much of the earth's surface is underlain by permafrost?, J. Geophys. Res. Earth Surf., 126, e2021JF006123, https://doi.org/10.1029/2021JF006123, 2021.

Pan, H. L. and Mahrt, L.: Interaction between soil hydrology and boundary-layer development, Boundary-Layer Meteorology, 38, 185-202, https://doi.org/10.1007/BF00121563, 1987.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and

Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Proceedings of the 33rd International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA, 2019.

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.

Razavi, S., Tolson, B. A., and Burn, D. H.: Review of surrogate modeling in water resources, Water Resources Research, 48, https://doi.org/10.1029/2011WR011527, 2012.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381-394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Rosero, E., Yang, Z.-L., Gulden, L. E., Niu, G.-Y., and Gochis, D. J.: Evaluating Enhanced Hydrological Representations in Noah LSM over Transition Zones: Implications for Model Development, Journal of Hydrometeorology, 10, 600-622, https://doi.org/10.1175/2009JHM1029.1, 2009.

Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence, 1, 206-215, https://doi.org/10.1038/s42256-019-0048-x10.1038/s42256-019-0048-x, 2019.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, Nature, 323, 533-536, https://doi.org/10.1038/323533a0, 1986.

Schädel, C., Rogers, B. M., Lawrence, D. M., Koven, C. D., Brovkin, V., Burke, E. J., Genet, H., Huntzinger, D. N., Jafarov, E., McGuire, A. D., Riley, W. J., and Natali, S. M.: Earth system models must include permafrost carbon processes, Nature Climate Change, 14, 114-116, https://doi.org/10.1038/s41558-023-01909-9, 2024.

Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., and Villalobos, P.: Compute Trends Across Three Eras of Machine Learning, 2022 International Joint Conference on Neural Networks (IJCNN), 18-23 July 2022, 1-8, https://doi.org/10.1109/IJCNN55064.2022.9891914,

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., Rahmani, F., Song, Y.,

Beck, H. E., Bindas, T., Dwivedi, D., Fang, K., Höge, M., Rackauckas, C., Mohanty, B., Roy, T., Xu, C., and Lawson, K.: Differentiable modelling to unify machine learning and physical models for geosciences, Nature Reviews Earth & Environment, 4, 552-567, https://doi.org/10.1038/s43017-023-00450-9, 2023.

Stuurop, J. C., van der Zee, S. E. A. T. M., Voss, C. I., and French, H. K.: Simulating water and heat transport with freezing and cryosuction in unsaturated soil: Comparing an empirical, semi-empirical and physically-based approach, Advances in Water Resources, 149, 103846, https://doi.org/10.1016/j.advwatres.2021.103846, 2021.

Sun, R., Pan, B., and Duan, Q.: Learning Distributed Parameters of Land Surface Hydrologic Models Using a Generative Adversarial Network, Water Resources Research, 60, e2024WR037380, https://doi.org/10.1029/2024WR037380, 2024.

Szabó, B., Kassai, P., Plunge, S., Nemes, A., Braun, P., Strauch, M., Witing, F., Mészáros, J., and Čerkasova, N.: Addressing soil data needs and data gaps in catchment-scale environmental modelling: the European perspective, Soil, 10, 587-617, https://doi.org/10.5194/soil-10-587-2024, 2024.

Teuling, A. J., Uijlenhoet, R., van den Hurk, B., and Seneviratne, S. I.: Parameter Sensitivity in LSMs: An Analysis Using Stochastic Soil Moisture Models and ELDAS Soil Parameters, Journal of Hydrometeorology, 10, 751-765, https://doi.org/10.1175/2008JHM1033.1, 2009.

Tian, W. and Nan, Z.: wbtian/NoahPy: NoahPy (v1.0.1), Zenodo [code], https://doi.org/10.5281/zenodo.16530326, 2025a.

Tian, W. and Nan, Z.: Data associated with the paper "NoahPy: A differentiable Noah land surface model for simulating permafrost thermo-hydrology" figshare [dataset], https://doi.org/10.6084/m9.figshare.29988163, 2025b.

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., and Shen, C.: From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling, Nature Communications, 12, 5988, https://doi.org/10.1038/s41467-021-26107-z, 2021.

Wang, C., Jiang, S., Zheng, Y., Han, F., Kumar, R., Rakovec, O., and Li, S.: Distributed Hydrological Modeling With Physics-Encoded Deep Learning: A General Framework and Its Application in the Amazon, Water Resources Research, 60, e2023WR036170, https://doi.org/10.1029/2023WR036170, 2024.

Wang, N., Zhang, D., Chang, H., and Li, H.: Deep learning of subsurface flow via theory-guided neural network, Journal of Hydrology, 584, 124700, https://doi.org/10.1016/j.jhydrol.2020.124700, 2020.
Wang, Z., Shao, M., Huang, L., Pei, Y., and Li, R.: Distribution and Influencing Factors of Soil Saturated Hydraulic Conductivity Under Different Land Use Patterns in Eastern Qinghai Province, Journal of Soil and Water Conservation, 35, 150-155, https://doi.org/10.13870/j.cnki.stbcxb.2021.03.021, 2021.
Werbos, P. J.: Backpropagation through time: what it does and how to do it, Proceedings of the IEEE, 78, 1550-1560, https://doi.org/10.1109/5.58337, 1990.
Wi, S. and Steinschneider, S.: Assessing the Physical Realism of Deep Learning Hydrologic Model Projections Under Climate Change, Water Resources Research, 58, e2022WR032123, https://doi.org/10.1029/2022WR032123, 2022.
Wu, X. and Nan, Z.: A multilayer soil texture dataset for permafrost modeling over Qinghai-Tibetan Plateau, 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 10-15 July 2016, 4917-4920, https://doi.org/10.1109/IGARSS.2016.7730283,
Wu, X., Nan, Z., Zhao, S., Zhao, L., and Cheng, G.: Spatial modeling of permafrost distribution and properties on the Qinghai-Tibet Plateau, Permafrost and Periglacial Processes, 29, 86-99, https://doi.org/10.1002/ppp.1971, 2018.
Xiao, Y., Zhao, L., Dai, Y., Li, R., Pang, Q., and Yao, J.: Representing permafrost properties in CoLM for the Qinghai–Xizang (Tibetan) Plateau, Cold Regions Science and Technology, 87, 68-77, https://doi.org/10.1016/j.coldregions.2012.12.004, 2013.
Xie, W., Kimura, M., Takaki, K., Asada, Y., Iida, T., and Jia, X.: Interpretable Framework of Physics-Guided Neural Network With Attention Mechanism: Simulating Paddy Field Water Temperature Variations, Water Resources Research, 58, e2021WR030493, https://doi.org/10.1029/2021WR030493, 2022.
Zhang, G., Nan, Z., Hu, N., Yin, Z., Zhao, L., Cheng, G., and Mu, C.: Qinghai-Tibet Plateau Permafrost at Risk in the Late 21st Century, Earth's Future, 10, e2022EF002652, https://doi.org/10.1029/2022EF002652, 2022a.
Zhang, G., Nan, Z., Hu, N., Yin, Z., Zhao, L., Cheng, G., and Mu, C.: Qinghai-Tibet Plateau permafrost at risk in the late 21st century, OSF [code], https://doi.org/10.17605/osf.io/g7jqr, 2022b.
Zhang, X.: Vegetation map of the People's Republic of China (1:1 000 000), Geology Press, 2007.
Zhang, X., Sun, S. F., and Xue, Y.: Development and Testing of a Frozen Soil Parameterization for Cold Region Studies, Journal of Hydrometeorology, 8, 690-701, https://doi.org/10.1175/JHM605.1, 2007.
Zhao, L., Wu, Q., Marchenko, S. S., and Sharkhuu, N.: Thermal state of permafrost and active layer in Central Asia during the international polar year, Permafrost and Periglacial Processes, 21, 198-207, https://doi.org/10.1002/ppp.688, 2010.
Zhao, Y., Nan, Z., Cao, Z., Ji, H., and Hu, J.: Evaluation of Parameterization Schemes for Matric Potential in Frozen Soil in Land Surface Models: A Modeling Perspective, Water Resources Research, 59, e2023WR034644, https://doi.org/10.1029/2023WR034644, 2023.
Zheng, D., van der Velde, R., Su, Z., Wang, X., Wen, J., Booij, M. J., Hoekstra, A. Y., and Chen, Y.: Augmentations to the Noah Model Physics for Application to the Yellow River Source Area. Part I: Soil Water Flow, Journal of Hydrometeorology, 16, 2659-2676, https://doi.org/10.1175/JHM-D-14-0198.1, 2015.