JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH000682

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

A Hybrid Physics-Guided Deep Learning Modeling Framework for Predicting Surface Soil Moisture

Xuan Xi[1], Qianlai Zhuang[1,2], and Xinyu Liu[3]

[1]Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA, [2]Department of Agronomy, Purdue University, West Lafayette, IN, USA, [3]Department of Mathematics, Purdue University, West Lafayette, IN, USA

Key Points:

- A new deep learning modeling framework guided by physical processes was developed to predict surface soil moisture
- The physics-guided deep learning (PGDL) model outperforms the traditional process-based model and pure deep learning model in prediction accuracy
- Hydrological process consideration increases the interpretability and reliability of the PGDL model

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Q. Zhuang,
qzhuang@purdue.edu

Citation:

Xi, X., Zhuang, Q., & Liu, X. (2025). A hybrid physics-guided deep learning modeling framework for predicting surface soil moisture. *Journal of Geophysical Research: Machine Learning and Computation*, *2*, e2025JH000682. https://doi.org/10.1029/2025JH000682

Received 6 MAR 2025
Accepted 21 JUN 2025

Abstract Accurate prediction of surface soil moisture (SSM) is vital for understanding the complex interactions between terrestrial and atmospheric processes with significant implications for weather forecasting, agriculture, and water management. In this study, we introduce an innovative physics-guided deep learning (PGDL) model by integrating the process-based insights of the terrestrial ecosystem model (TEM) with the dynamic predictive capabilities of long short-term memory (LSTM) networks to improve SSM prediction. The PGDL model leverages the complementary strengths of the deterministic framework of TEM and the data-driven prowess of LSTM, providing predictions that are deeply rooted in physical processes while capturing complex patterns in data. We evaluated the PGDL model against traditional process-based (PB) models and pure deep learning (DL) approaches in both single-site and multisite simulations. For single-site simulations, we performed time-based partitioning on 7 sites with different vegetation types, whereas for multisite simulation, data from 13 sites within the same vegetation type were used for model training and evaluation. In single-site simulations, PGDL achieved significantly lower RMSE (0.04) and higher R^2 (0.66) compared to PB (RMSE: 0.13, R^2: −2.32) and DL (RMSE: 0.06, R^2: 0.36). In multisite simulations, PGDL (RMSE: 0.06, R^2: 0.63) also outperformed PB (RMSE: 0.09, R^2: 0.28) and DL (RMSE: 0.08, R^2: 0.38). Our results show that the PGDL modeling framework improves the predictive accuracy of DL models and the physical interpretability of PB models, which can serve as a robust tool to predict SSM dynamics.

Plain Language Summary In our study, we developed a new model to predict how wet the soil is at the surface, which is important for understanding how land interacts with the atmosphere. This information is crucial for better weather forecasting, farming, and managing water resources. Our new method combines traditional process-based models with advanced deep-learning techniques. By blending these approaches, our model not only relies on established physical principles but also learns from patterns in data that process-based methods might miss. We evaluated our approach through single-site and multisite simulations. For single-site simulations, we independently analyzed data from 7 sites each with a different vegetation type. For multisite simulation, we used data from 13 sites within the same vegetation type. Compared to traditional process-based and pure deep-learning models, our new model achieves higher accuracy and more physically consistent predictions. The success of our model highlights the advantage of combining the interpretability of traditional scientific methods with the flexibility of big data analytics. This could lead to better ways of predicting soil moisture, which is becoming increasingly important as the climate changes.

1. Introduction

Understanding and predicting surface soil moisture (SSM) is important in the field of Earth sciences due to its critical role in hydrological, ecological, and atmospheric dynamics. SSM acts as a pivotal link between the terrestrial water cycle and atmospheric conditions influencing weather patterns, climate regulation, and water availability for ecosystems and agriculture (Koster et al., 2004; Rodriguez-Iturbe, 2000; Seneviratne et al., 2010). Accurate measurements and predictions of SSM can significantly improve weather forecasting accuracy, agricultural productivity, and water resource management (Abowarda et al., 2021; Chatterjee et al., 2022). SSM affects evapotranspiration, soil respiration, and plant transpiration processes as well as the exchange of water and heat energy between the land surface and the atmosphere (Porporato et al., 2002; Zhou et al., 2021). SSM is also important to understanding the carbon cycle and its feedback on the Earth climate system (Green et al., 2019; Humphrey et al., 2021).

Despite its critical role, achieving accurate predictions of SSM is fraught with challenges stemming from the intricate interactions among soil characteristics, vegetation cover, and atmospheric conditions. The advancement in soil moisture prediction necessitates both empirical progress and theoretical refinement. Recent efforts have leveraged data-driven machine learning/deep learning (ML/DL) techniques (e.g., Adab et al., 2020; Cai et al., 2019; Celik et al., 2022) and traditional theory-driven process-based (PB) methods (e.g., Fan & Van Den Dool, 2004; Mo et al., 2012; Xia et al., 2014) to improve soil moisture predictions. Although both purely data-driven and theory-based approaches have their respective strengths, they each also possess limitations that may compromise their effectiveness in making SSM predictions.

The core strength of ML/DL lies in processing large amounts of data and deciphering the complex nonlinear patterns and interactions, thereby revealing insights into complex SSM dynamics influenced by multiple factors (Reichstein et al., 2019). However, the opacity of ML/DL models, which complicates the interpretability and understanding of their predictive causal mechanisms, significantly hinders their applications that require a deep understanding of process dynamics (Karpatne et al., 2017; Rasheed et al., 2022). ML/DL models often suffer from overfitting to training data due to their sensitivity to data quality and quantity, exacerbating poor generalization performance under unobserved conditions, and requiring large and carefully curated data sets to perform efficient training on data sets that are not always available (Hawkins, 2004). Furthermore, the success of ML/DL models in capturing complex nonlinear relationships does not necessarily imply the ability to predict future conditions under climate change where the underlying dynamics may be beyond the range of historical changes captured in the training data (Reichstein et al., 2019). The lack of transparency and difficulty in establishing causal relationships limits the applicability of ML/DL methods to fields where process understanding is crucial, a domain where PB models demonstrate superior performance (Clark et al., 2017).

In contrast, PB models elucidate the underlying mechanisms affecting SSM and its interaction with the climate system by explicitly representing physical processes (Wood et al., 2011). This inherent transparency and ability to incorporate theoretical understandings of the hydrological cycle allow the PB model to gain insight into the processes driving SSM dynamics. However, the reliability of PB models is often compromised by their parameter-related uncertainties, oversimplification of complex processes, and the need for extensive calibration against empirical data (Arhonditsis & Brett, 2004; Clark et al., 2016; Montanari, 2007). These issues may reduce their predictive accuracy and limit their transferability in different environments. In particular, the calibration process is often hampered by the scarcity of observational data and the subjectivity of adjusting model parameters to fit the observed data, causing performance to degrade under untested conditions (Gupta et al., 2008). Additionally, integrating new data and features into PB models encounters significant lag. This gap slows down the improvement of PB models reducing their usefulness and rapid response to complex environmental problems (Read et al., 2019). These key limitations of PB models are exactly the areas where ML/DL methods show advantages.

Therefore, the weaknesses of ML/DL models in terms of transparency and interpretability align with the strengths of PB models, which offer clear insights into hydrological processes and the theoretical basis behind predictions. Conversely, the challenges faced by PB models in parameter uncertainty and the need for extensive calibration contrast with the strengths of ML/DL approaches, which can efficiently process and learn from large data sets without explicit programming to capture the underlying process dynamics. This highlights a complementary relationship, suggesting that integrating ML/DL and PB approaches could harness the data processing and pattern recognition capabilities of the former with the process understanding and theoretical grounding of the latter, potentially overcoming the limitations of both to achieve more accurate, interpretable, and robust SSM predictions. Physics-guided machine learning (PGML) combines theory-driven PB models with data-driven ML models to leverage their complementary strengths. PGML is based on a new modeling paradigm—"Theory-Guided Data Science" (Karpatne et al., 2017)—to consider a hybrid modeling approach to remain true to accepted theory while also learning complex relationships from abundant data, thus integrating the strengths of both empirical and theoretical knowledge. PGML has been widely used and developed rapidly in many research fields (Willard et al., 2022). In terms of hydrology, the application of PGML has significantly enhanced the accuracy and interpretability of hydrological model predictions by integrating physical principles for deeper insights into water cycle dynamics, thus greatly improving water resources management under unpredictable climate change (Wang et al., 2020; Xie et al., 2021). Climate science has also made substantial progress through the application of PGDL, which combines atmospheric physics with deep learning to improve the prediction of extreme weather events and long-term climate change (Kashinath et al., 2021; Zanetta et al., 2023). In addition, PGML has

improved estimations of paddy field water temperature and agricultural nitrous oxide emissions providing interpretable forecasts for intricate agricultural dynamics (Liu et al., 2022; Xie et al., 2022).

There have also been many recent studies devoted to improving soil moisture prediction using PGDL methods. Breen et al. (2020) developed a hybrid artificial neural network that integrated dynamic weather and static spatial data, demonstrated comparable accuracy to satellite observations, and highlighted the value of meteorological time series in capturing the physical processes behind soil moisture changes. Geng et al. (2024) introduced a novel loss function based on physical mechanisms significantly improving soil moisture prediction performance through a physically guided long short-term memory (LSTM) network compared with traditional models. Both studies are mainly based on physics-guided loss function as the PGDL method. Although this approach shows promise in improving soil moisture prediction, its effectiveness is not universal. A major limitation is that this method oversimplifies complex hydrological interactions, resulting in an inability to account for nonlinear relationships between environmental variables and soil moisture, thus making generalization across different soil types and climate conditions difficult. Han et al. (2023) used physics-informed machine learning to generate a high-resolution SSM data set using remote sensing and meteorological data guided by knowledge of physical processes that affect soil moisture dynamics, but it was not based on complete physical processes and lacked sufficient physical interpretability. Li et al. (2024) developed a hybrid model based on an attention mechanism to effectively integrate the output of the PB model into the DL model and then improved soil moisture prediction using an ensemble model that combined the advantages of different hybrid models, but it failed to clearly combine external climate inputs and internal historical soil moisture data to build the model.

In this study, we developed a PGDL model for SSM predictions by combining a PB model, the terrestrial ecosystem model (TEM, Zhuang et al., 2004), and a DL model based on long short-term memory (LSTM, Hochreiter & Schmidhuber, 1997). By integrating these two models, this approach leverages the physical understanding of SSM processes from TEM and the ability of LSTM to learn complex patterns from data. Unlike existing studies, our PGDL model combines the explicit physical processes of the water balance equation from the PB model with the implicit relationships between soil moisture and its historical states. By constructing distinct neural network branches, our model directly incorporates both historical data and current relevant physical information into the SSM prediction. Compared with other existing methods, our PGDL model enhances the ability to handle complex hydrological interactions and nonlinear relationships, effectively integrates multiple data sources, and provides physical interpretability with better prediction performance.

In this study, we focus on site-level simulations and conduct two sets of experiments: single-site and multisite simulations. For the single-site simulations, we select seven sites that represent seven major vegetation types to ensure a broad representation of SSM dynamics. We partition the time series of each site to assess the ability of the PB, DL, and PGDL models to make predictions using data from the same site. For the multisite simulations, we select 13 sites with the same vegetation type to evaluate the model performance across different sites. We expect that the PGDL model will provide a more accurate, reliable, and physically consistent method for SSM prediction. By bridging the gap between DL models and PB understanding, our framework sets the stage for future explorations into the vast potential of hybrid modeling approaches in environmental science and beyond.

In this paper, Section 2 first describes the process of data collection and then presents the PB and DL models used for PGDL development and transitions to elaborating on the innovative creation of the PGDL model, concluding with the establishment of an experimental setup designed to evaluate this model. Section 3 presents our findings by contrasting model performances through time series analyses and scatter plot evaluations to discern the efficacy of each modeling approach. Finally, in Section 4, we provide a critical discussion comparing DL and PGDL models in the context of physical law compliance, the impact of limited training data on model performance, and strategies to enhance model interpretability and accuracy through feature importance analysis.

2. Methods

2.1. Data Collection

2.1.1. Observation of Surface Soil Moisture

In this study, we use in situ measured SSM from the International Soil Moisture Network (ISMN; Dorigo et al., 2011, 2013) as the label to train DL and PGDL models. The ISMN provides high-quality, in situ SSM measurements. Established as a centralized global database, ISMN integrates diverse soil moisture observations

Table 1
Information on the Selected Sites From ISMN in This Study

Simulation	Site	Vegetation type	Name	Network	Location
Single-site	1	Alpine tundra and polar desert	Fremontpass	SNOTEL	106.20°W, 39.38°N
	2	Boreal forest	MountainHome	SCAN	110.40°W, 40.37°N
	3	Temperate coniferous forests	Raggedmountain	SNOTEL	117.04°W, 47.86°N
	4	Temperate deciduous forests	McalisterFarm	SCAN	86.58°W, 35.07°N
	5	Grasslands	Phillipsburg	SCAN	99.33°W, 39.78°N
	6	Xeric shrublands	BlueCreek	SCAN	112.43°W, 41.93°N
	7	Xeric woodlands	Louismeadow	SNOTEL	111.76°W, 40.83°N
Multi-site	1	Grasslands	AdamsRanch#1	SCAN	105.42°W, 34.25°N
	2		ErosDataCenter	SCAN	96.62°W, 43.73°N
	3		FortReno#1	SCAN	98.02°W, 35.55°N
	4		Jordan	SCAN	107.13°W, 47.52°N
	5		LosLunasPmc	SCAN	106.77°W, 34.77°N
	6		Mandan#1	SCAN	100.92°W, 46.77°N
	7		Phillipsburg	SCAN	99.33°W, 39.78°N
	8		UwPlatteville	SCAN	90.39°W, 42.71°N
	9		FryCanyon	SNOTEL	115.93°W, 41.58°N
	10		LAPRELECREEK	SNOTEL	105.86°W, 42.44°N
	11		MedBow	SNOTEL	106.35°W, 41.38°N
	12		NAVAJOWHISKEYCK	SNOTEL	108.95°W, 36.18°N
	13		SOURDOUGHGULCH	SNOTEL	117.39°W, 46.24°N

from multiple networks and platforms ensuring a comprehensive and heterogeneous data set. This network encompasses data from various sensors and retrieval methods offering a robust representation of soil moisture dynamics across different geographical and climatic conditions.

All sites in this study were located in nonwetland soils within the continental United States (CONUS). For the single-site simulations, we selected seven sites, each representing a common vegetation type, to compare model performance across different vegetation conditions. For the multisite simulation, we selected one vegetation type, grassland, and selected 13 sites within this category. Grassland was chosen for two main reasons: first, it had enough available ISMN sites ensuring an adequate amount of training data; second, grassland exhibits a more typical SSM variation pattern. Soil moisture in this ecosystem is primarily regulated by precipitation and evapotranspiration resulting in relatively intuitive moisture fluctuations that mainly depend on shallow soil layers. The information on all selected sites from ISMN in this study is shown in Table 1.

The observed SSM from ISMN for each site is collected as a daily time series spanning from 1 January 2016 to 31 December 2021 yielding a total of 2,192 data points. The original ISMN SSM data were recorded in an hourly time step. We first averaged these hourly data to obtain daily time series for each site. The selection of these sites ensured that the resulting daily time series was high-quality, complete, and free of missing values.

2.1.2. Input Data

We prepared input data for each site, including spatially explicit information on soils, climate, and leaf area index (LAI). For the single-site experiment, only dynamic inputs were used for the DL and PGDL models, whereas for the multisite experiment, both static and dynamic inputs were incorporated to account for spatial variability. To ensure consistency and comparability across model simulations, we used the same inputs for all models.

For the climate data, we used the 20CRv3-ERA5 data set (ERA5 is the fifth-generation reanalysis of ECMWF (European Centre for Medium-Range Weather Forecasts)), a global climate data set with a resolution of 0.5° × 0.5° at daily time step from the Inter-Sectoral Impact Model Intercomparison Project (Lange et al., 2022). The climate data include precipitation (Pr, units: mm), surface downwelling shortwave radiation

(SWD, units: W/m^2), near-surface air temperature (T_{air}, units: °C), and near-surface relative humidity (units: %) where air temperature and relative humidity were used to calculate the vapor pressure (V_P, units: hPa) as another input. LAI data are obtained from ERA5 (specifically, ERA5 hourly data on single levels from 1940 to present, Hersbach et al., 2023) at a resolution of 0.25° × 0.25° at the hourly time step. We then converted the same hourly LAI into monthly time step as input for the PB model and into daily time step as input for the DL and PGDL models. For static input, we used elevation and soil texture, including soil sand, silt, and clay content (Zhuang et al., 2011). These variables vary spatially but remain fixed for each site, providing additional context information to support multisite simulations.

2.2. Model Description

2.2.1. Process-Based (PB) Model

TEM is a land ecosystem modeling framework that quantifies water, carbon, and nitrogen exchanges between the atmosphere and terrestrial ecosystems (Zhuang et al., 2004). In this study, we use its coupled water balance model (WBM; Zhuang et al., 2002), which can provide soil moisture simulation globally on a long-time scale. The WBM is divided into two parts for simulating soil moisture: one for general non-wetland soils and one for wetland soils. The focus of this study is on predicting soil moisture for general non-wetland soils. Accordingly, all references to soil moisture in this study pertain to non-wetland soils. In the non-wetland soils part of WBM, the soil profile is represented by six layers with different hydrologic characteristics. In this study, we predict the soil moisture time series in the first soil layer, which represents SSM in the model.

As the main water reservoir in the ecosystem, soils can be regarded as a "bucket." Precipitation fills the bucket, and this water is lost through evapotranspiration and runoff (Waring & Running, 2010). Therefore, changes to the water content of the soil profile (W_S) are determined using a water balance approach in TEM as (Zhuang et al., 2004)

$$\frac{\partial W_S}{\partial t} = I_F - E_V - E_S - D_R \quad (1)$$

where t means time, I_F is infiltration, E_V is evapotranspiration from the vegetation canopy, E_S is evaporation from the soil surface, and D_R is drainage from the first soil layer. SSM dynamics are influenced by the factors in Equation 1 and are assumed to vary according to the Richards equation with the SSM obtained simultaneously by numerically solving both equations using a tridiagonal system of equations (Zhuang et al., 2004).

2.2.2. Deep Learning (DL) Model

In this study, we use long short-term memory (LSTM; Hochreiter & Schmidhuber, 1997) as the DL model. LSTM models are a specific type of recurrent neural network (RNN) that can keep track of long-term information. Compared to standard RNNs, LSTM embeds "memory" into its architecture, which allows it to handle the types of long-term dependencies that are usually difficult for standard RNNs. The cell states of LSTM have gates that determine whether information can be passed from one of its multiple layers into the next cell state. In the PGDL model, we directly adopted the original LSTM network structure without making any specific modifications to the gate structure. Therefore, the function of the LSTM gate structure in the PGDL model is consistent with the original LSTM gate structure. Specifically, the LSTM unit has three typical gates: the input gate, the forget gate, and the output gate. The input gate controls how much of the new input information is added to the cell state. The forget gate decides how much of the previous cell state information is retained. The output gate determines how much of the cell state influences the output. These gates work together to allow the LSTM unit to effectively manage information and process time series data to make predictions.

This research fully benefits from the advantages of LSTM networks. First, they are architected essentially to recognize, remember, and predict based on long-term patterns in data series, which is critical for processing the large amounts of historical soil moisture data on which our study relies. Second, LSTM can distinguish between relevant and irrelevant past information and can predict the changing trend of soil moisture over time in a more detailed and accurate manner. By integrating LSTM into our hybrid model framework, we exploit its superior temporal data processing capabilities to capture the essence of complex temporal dynamics in soil moisture.

2.2.3. PGDL Model Development

We already know that SSM is calculated from I_F, E_V, E_S, and D_R according to the water balance equation (refer to Equation 1). In addition, previous research has shown that soil moisture exhibits "memory," indicating that the current SSM is also highly related to the previous SSM states (Koster & Suarez, 2001). Therefore, we assume that the target prediction SSM (W_S) is not only decided by an explicit physical model including I_F, E_V, E_S, and D_R but also has an implicit relationship with the history data $W_S(t-k)$. Based on this assumption, we propose a dynamic process formulation that combines the water balance equation with the historical SSM data:

$$\begin{cases} \dfrac{d}{dt}W_S(t) = I_F - E_V - E_S - D_R \\ W_S(t) = f(W_S(t-1), W_S(t-1), \ldots, W_S(t-k)) \end{cases} \tag{2}$$

where f is an unknown function and k, serving as a hyperparameter called sequence length in the LSTM structure, is the number of historical data of W_S used for the current prediction. The second line of Equation 2 describes the implicit relationship between current SSM and previous k SSM states. Inspired by the Adams-Moulton method (Hairer et al., 1993), which is a family of implicit multistep methods for solving ordinary differential equations, we combine the two formulations into

$$W_S(t) = \mathcal{N}\left(W_S{}^k, I_F{}^k, E_V{}^k, E_S{}^k, D_R{}^k\right) \tag{3}$$

where $W_S{}^k, I_F{}^k, E_V{}^k, E_S{}^k, D_R{}^k$ represent the historical data of the corresponding variables that

$$W_S{}^k = [W_S(t-1), W_S(t-2), \ldots, W_S(t-k)] \tag{4}$$

$$I_F{}^k = [I_F(t), I_F(t-1), \ldots, I_F(t-k+1)] \tag{5}$$

$$E_V{}^k = [E_V(t), E_V(t-1), \ldots, E_V(t-k+1)] \tag{6}$$

$$E_S{}^k = [E_S(t), E_S(t-1), \ldots, E_S(t-k+1)] \tag{7}$$

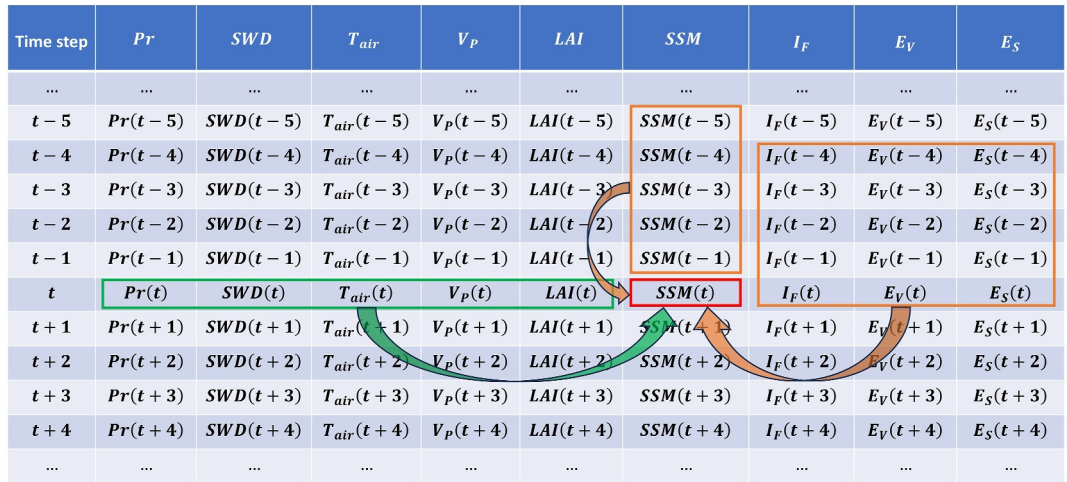
$$D_R{}^k = [D_R(t), D_R(t-1), \ldots, D_R(t-k+1)] \tag{8}$$

The Adams-Moulton method uses historical data to approximate the solution of the differential equation at the current time step. Similarly, our proposed model uses the historical and current data of physical variables and historical SSM data to predict the current SSM value.

The values of I_F, E_V, E_S, and D_R within the water balance equation (Equation 1) are calculated from the PB model using the dynamic physical inputs $P(t)$, including Pr, SWD, T_{air}, V_P, and LAI. Although Equation 1 formally satisfies the water balance condition and is widely used in hydrological modeling, it is often regarded as an empirical expression as it is difficult to fully characterize complex real-world processes. In particular, the drainage term D_R involves more complex subsurface hydrological processes, which are simplified and subject to structural uncertainty in the current PB model, making it difficult to accurately and comprehensively represent the relevant mechanisms. As a result, the estimation of D_R is more prone to instability and systematic bias than the other variables, thus limiting its effectiveness as an input feature for neural network training.

To address this issue and avoid relying on potentially inaccurate model outputs, we introduce an uncertainty term U^k in place of $D_R{}^k$. This term represents all unmodeled or difficult to estimate physical variables related to SSM (e.g., drainage) and is inferred from the same dynamic inputs $P(t)$, like I_F, E_V, and E_S. U^k provides a flexible, data-driven approach that enables the model to incorporate relevant physical drivers while compensating for the structural limitations of the PB model and maintaining physical consistency within the modeling framework. Therefore, the final model can be expressed as

$$W_S(t) = \mathcal{N}\left(W_S{}^k, I_F{}^k, E_V{}^k, E_S{}^k, P(t)\right) \tag{9}$$



Time step	Pr	SWD	T_{air}	V_P	LAI	SSM	I_F	E_V	E_S
...	...	...	...	...	...	...	...	...	...
$t-5$	$Pr(t-5)$	$SWD(t-5)$	$T_{air}(t-5)$	$V_P(t-5)$	$LAI(t-5)$	$SSM(t-5)$	$I_F(t-5)$	$E_V(t-5)$	$E_S(t-5)$
$t-4$	$Pr(t-4)$	$SWD(t-4)$	$T_{air}(t-4)$	$V_P(t-4)$	$LAI(t-4)$	$SSM(t-4)$	$I_F(t-4)$	$E_V(t-4)$	$E_S(t-4)$
$t-3$	$Pr(t-3)$	$SWD(t-3)$	$T_{air}(t-3)$	$V_P(t-3)$	$LAI(t-3)$	$SSM(t-3)$	$I_F(t-3)$	$E_V(t-3)$	$E_S(t-3)$
$t-2$	$Pr(t-2)$	$SWD(t-2)$	$T_{air}(t-2)$	$V_P(t-2)$	$LAI(t-2)$	$SSM(t-2)$	$I_F(t-2)$	$E_V(t-2)$	$E_S(t-2)$
$t-1$	$Pr(t-1)$	$SWD(t-1)$	$T_{air}(t-1)$	$V_P(t-1)$	$LAI(t-1)$	$SSM(t-1)$	$I_F(t-1)$	$E_V(t-1)$	$E_S(t-1)$
t	$Pr(t)$	$SWD(t)$	$T_{air}(t)$	$V_P(t)$	$LAI(t)$	$SSM(t)$	$I_F(t)$	$E_V(t)$	$E_S(t)$
$t+1$	$Pr(t+1)$	$SWD(t+1)$	$T_{air}(t+1)$	$V_P(t+1)$	$LAI(t+1)$	$SSM(t+1)$	$I_F(t+1)$	$E_V(t+1)$	$E_S(t+1)$
$t+2$	$Pr(t+2)$	$SWD(t+2)$	$T_{air}(t+2)$	$V_P(t+2)$	$LAI(t+2)$	$SSM(t+2)$	$I_F(t+2)$	$E_V(t+2)$	$E_S(t+2)$
$t+3$	$Pr(t+3)$	$SWD(t+3)$	$T_{air}(t+3)$	$V_P(t+3)$	$LAI(t+3)$	$SSM(t+3)$	$I_F(t+3)$	$E_V(t+3)$	$E_S(t+3)$
$t+4$	$Pr(t+4)$	$SWD(t+4)$	$T_{air}(t+4)$	$V_P(t+4)$	$LAI(t+4)$	$SSM(t+4)$	$I_F(t+4)$	$E_V(t+4)$	$E_S(t+4)$
...	...	...	...	...	...	...	...	...	...

Figure 1. An illustration for the dynamic inputs of W_S predicted at time step t (SSM(t)). In this example, we assume the sequence length (k) is 5. This example shows the data used to predict SSM(t) (encircled by red frames). The orange frames and arrows indicate the inputs for the long short-term memory branch, and the green frame and arrows indicate the dynamic inputs for the fully connected neural network branch within the physics-guided deep learning modeling framework.

Based on Equation 9, we develop a PGDL model as a hybrid modeling framework to predict SSM. First, we use the intermediate variables that are directly used to calculate SSM from the physical equation (i.e., I_F, E_V, and E_S) as the inputs to the LSTM network. In addition, we use the previous SSM value as another input to the LSTM network to ensure that the necessary historical information is delivered to the network in the same sequential format as the three intermediate variables. Therefore, LSTM performs feature extraction on four inputs in the modeling framework. The length of time steps that the LSTM processes these inputs is determined by k. Specifically, to predict $W_S(t)$ at time step t, the inputs to the LSTM will contain I_F, E_V, and E_S from time step $t-k+1$ to time step t, denoted as ${I_F}^k$, ${E_V}^k$, and ${E_S}^k$, and W_S from time step $t-k$ to time step $t-1$, denoted as ${W_S}^k$. In this way, the LSTM utilizes the past k time steps of data, including I_F, E_V, E_S, and W_S to make accurate predictions. Figure 1 shows the dynamic inputs of W_S predicted at time step t taking $k=5$ as an example. The detailed computational process of the LSTM is provided in Text S1 in Supporting Information S1.

Since the three intermediate variables cannot fully capture the physical processes to get SSM, we develop an additional fully connected neural network (FCNN) with the inputs as $P(t)$. For multisite simulations, we include static data as additional inputs to the FCNN allowing the model to account for spatial variability and site-specific characteristics, thus enhancing its ability to generalize across different sites. In this way, we take advantage of the ability of neural networks to learn the physical processes from $P(t)$ that are not involved in the current PB model. The FCNN branch contains four fully connected layers and adds nonlinearity through the activation function ReLU to capture complex data relationships and patterns.

The outputs of the LSTM and FCNN branches are then concatenated to form a comprehensive feature representation, which is mapped to the final output through an additional fully connected layer with the output size being 1, representing the SSM prediction value of the model. By concatenating the two outputs, the PGDL model can take full advantage of the LSTM's strengths in capturing temporal dependencies and the FCNN's ability to learn complex patterns of physical variables. A conceptual diagram of the PGDL modeling framework and its comparison with the PB model is shown in Figure 2.

We perform the preprocessing for the input features and labels before training. First, we perform normalization for the input features to ensure that all features are on the same scale. This could reduce the sensitivity of the model to the scale of the input features, thereby helping to accelerate model convergence, avoid vanishing or exploding gradients, and improve the generalization ability of the model. Specifically, we use Z-Score standardization: the mean value of the feature is returned to zero, and the standard deviation is normalized so that the feature conforms to the standard normal distribution. In addition, we applied a one-dimensional Kalman filter to

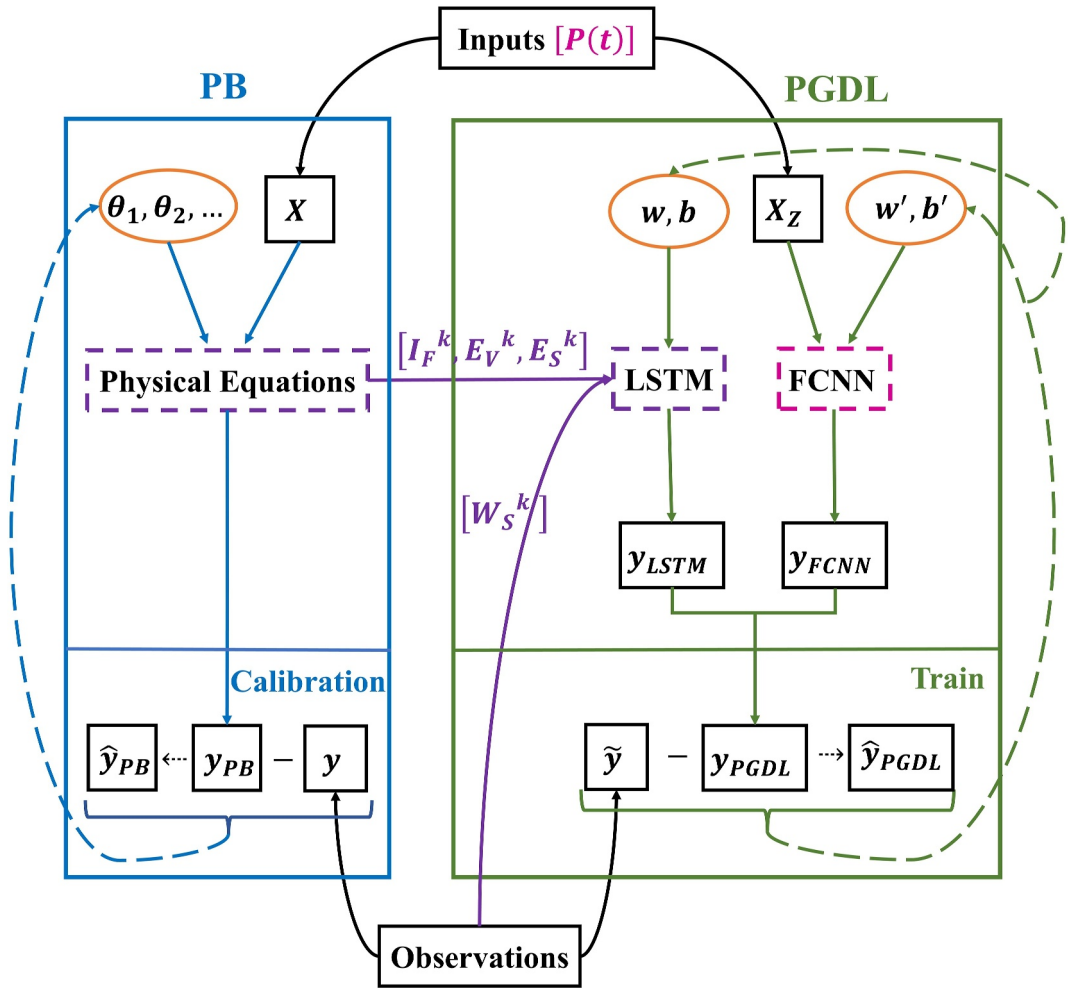


Figure 2. Conceptual diagram of the physics-guided deep learning (PGDL) model framework and comparison with the PB model. The blue part represents the workflow of the PB model, and the green part represents the workflow of the PGDL model. For the arrow lines, the solid lines represent the transfer of data, and the dotted lines represent the update of parameters. For the shapes, the black boxes represent data, the orange ellipses represent parameters, and the dashed boxes represent functional modules. The purple dashed box denotes the physical equation module in the PB model and its coupling with the PGDL model via the purple solid lines, which represent the flow of I_F^k, E_V^k, and E_S^k derived from the PB model and W_S^k from observations as inputs to the long short-term memory (LSTM) branch. The pink dashed box represents the fully connected neural network (FCNN) module in the PGDL model, which processes the inputs $P(t)$. This structure highlights the input components in Equation 9 and their corresponding pathways in the model. Both PB and PGDL models receive input (X), consisting of both static and dynamic features, and then obtain the predictions (y_{PB} and y_{PGDL}) through calculation with corresponding parameters. Finally, the parameters are updated to obtain the optimized values ($\hat{y}_{PB}$ and $\hat{y}_{PGDL}$) by comparing with observations (y). $\theta_{\sim}$ represents the parameters for the PB model. w and b denote the weight and bias parameters for the LSTM branch, and w' and b' denote the weight and bias parameters for the FCNN branch within the PGDL modeling framework. The differences between the two models are: (1) the PB model directly receives X, whereas the FCNN branch of the PGDL model receives normalized X (X_Z); (2) the PB model directly uses y, whereas the PGDL model uses Kalman filtered y ($\tilde{y}$).

the SSM observations used as training labels to reduce noise while preserving the overall trend. The filter was implemented under an identity state transition model with a process variance of 0.1 and a measurement variance of 1.0. These parameters were selected based on experimental testing to balance smoothing with signal fidelity. The filtered labels improve training stability by reducing extreme noise in the observations.

Table 2
The Calibrated Parameters Related to SSM Calculations in the PB Model

Simulation	Site (unit)	θ_{sat} (mm^3/mm^3)	k_{sat} (mm/s)	ψ_{sat} (mm)	b (unitless)	I_{RMAX} (unitless)	g_{max} (mm/s)	E_R (unitless)
Single-site	1	0.40	0.37	−120.00	5.10	0.20	13.50	0.28
	2	0.40	0.60	−147.22	3.18	0.22	15.20	0.50
	3	0.40	0.08	−120.79	3.89	0.18	16.70	0.25
	4	0.60	0.01	−112.84	3.77	0.27	5.25	0.35
	5	0.40	0.23	−116.84	5.49	0.30	4.22	0.38
	6	0.65	0.18	−120.07	3.72	0.29	4.13	0.40
	7	0.56	0.26	−117.89	3.13	0.26	3.90	0.30
Multisite	1	0.40	0.60	−116.53	6.41	0.21	4.20	0.40
	2	0.40	0.01	−119.90	5.87	0.31	4.20	0.40
	3	0.40	0.16	−119.56	9.08	0.26	3.50	0.50
	4	0.43	0.08	−175.87	4.70	0.31	4.20	0.40
	5	0.40	0.60	−120.26	5.17	0.31	4.20	0.40
	6	0.40	0.01	−120.02	11.12	0.31	4.20	0.40
	7	0.40	0.24	−116.83	5.48	0.31	4.20	0.40
	8	0.41	0.04	−119.79	7.40	0.30	4.20	0.40
	9	0.40	0.60	−132.69	4.42	0.31	4.20	0.40
	10	0.40	0.60	−120.08	3.27	0.26	3.50	0.50
	11	0.44	0.01	−119.58	11.40	0.21	4.20	0.40
	12	0.40	0.02	−120.00	11.40	0.21	4.20	0.40
	13	0.41	0.01	−114.57	7.90	0.24	4.20	0.40

2.3. Experiments

2.3.1. For PB Model

To ensure a fair comparison between the PB model and the DL/PGDL models, we first calibrate the PB model. This calibration is necessary because the accuracy of the SSM prediction of the PB model depends heavily on several key parameters. Specifically, we calibrate seven critical parameters: θ_{sat} (saturated volumetric water content), k_{sat} (saturated hydraulic conductivity), ψ_{sat} (saturated matric potential), b (Clapp-Hornberger parameter), I_{RMAX} (canopy rainfall interception parameter), g_{max} (maximum canopy conductance), and E_R (extinction coefficient of radiation). These key parameters serve as components of Equation 1 that simulate the change of SSM and jointly affect the SSM calculation in the PB model. θ_{sat} defines the maximum water-holding capacity of the surface soil directly affecting the soil's ability to retain moisture, which is critical for calculating relative saturation and influencing water movement and surface runoff. k_{sat} and ψ_{sat} work together to control water conductivity and retention. k_{sat} governs the efficiency of I_F and vertical water movement, whereas ψ_{sat} determines the matric potential gradient influencing soil water retention, fluidity, and surface runoff generation. The Clapp-Hornberger parameter b further refines these processes by describing how soil water retention characteristics change with varying moisture levels, thus affecting the discharge capacity and SSM dynamics. Additionally, I_{RMAX} controls the interception of rainfall by the canopy affecting rain throughfall and surface runoff, which in turn influences I_F and the daily SSM values. g_{max} and E_R play critical roles in regulating E_V and E_S by controlling canopy water conductance and radiation penetration through the canopy.

The calibration process was conducted using PEST (Model-Independent Parameter Estimation and Uncertainty Analysis, v17.2 for Linux) software, which is widely recognized in hydrological modeling for its ability to optimize model parameters by minimizing the difference between observed and simulated values (Doherty, 2004). We ensure consistency across models using the same SSM observational data that are employed for DL/PGDL training. The calibrated parameter values are presented in Table 2. The simulation from the calibrated PB model is then used for PGDL model training.

2.3.2. For DL and PGDL Models

The training of DL and PGDL models is implemented in Python with the PyTorch package. For both models, we perform cross-validation to fine-tune the hyperparameters by Bayesian optimization. There are seven hyperparameters: learning rate, batch size, sequence length, number of hidden sizes of LSTM, number of layers of LSTM, number of hidden sizes of FCNN, and dropout rate. We use the standard Adam optimizer (Kingma & Ba, 2014) for gradient descent and mean squared error as the loss function. During cross-validation, we adopted an early stopping strategy for each fold to prevent model overfitting and reduce unnecessary training time.

We applied different data partitioning strategies for training and validation processes for single-site simulation and multisite simulation. For single-site simulation, we divided the 6-year data of each site into training, validation, and test sets according to the year. Specifically, the first 5 years are used as training and validation set for cross-validation, and the sixth-year data is reserved as the test set. We adopt a time-series cross-validation technique called time series splitting to perform a four-fold cross-validation where each fold progressively expands the training set by 1 year while using the subsequent year as the validation set. In this way, each model training is performed on continuous historical data while predictions are made on subsequent, unseen data, maintaining the temporal order of the time series. To obtain more stable and reliable evaluations, we conducted five independent runs to evaluate the performance of the DL and PGDL models on the prediction task for each vegetation type. Each run starts with model initialization, and the training and test processes are carried out independently ensuring the independence of each experiment and the comparability of results. After completing all runs, the predictions from each run are averaged to obtain the final results. The subsequent analysis is based on the averaged results of the five runs, which can more accurately reflect the model performance on the test set.

For the multi-site simulation, we partitioned the data into training, validation, and test sets by site. First, we selected the entire data set from one of the 13 sites as the test set. The selection of the test site followed a two-step screening strategy to ensure both representativeness and generalization capability. In the first step, we used mean and standard deviation as metrics for representativeness. Specifically, we calculated the differences between the mean and standard deviation of SSM for each site and the overall SSM across all sites and then selected the top three sites with the smallest differences as candidates. In the second step, we evaluated the generalization capability of the three candidate sites from the perspectives of temporal correlation, extreme value error, and variance to select the most challenging site as the final test set. As a result, the FryCanyon (SNOTEL) site was chosen as the test set. We then performed hyperparameter optimization using leave-one-out cross-validation on the remaining 12 sites. Specifically, for each set of candidate hyperparameters, we left one site out as the validation set and used the remaining 11 sites as the training set, calculating the average validation error across all iterations as the performance metric. Finally, we trained the model on the 12 sites excluding the FryCanyon (SNOTEL) site and tested it on the test set.

For both single-site and multisite simulations, we apply the updated input method of single-step forecasting for iterative prediction of SSM time series on the test set. The core idea of this method is to use the output of the model as part of the input sequence for the next prediction after each prediction and gradually update the input data in this way. Specifically, every time a prediction is made, the SSM value of the next time step predicted by the model will be added to the input sequence, and the earliest data point in the input sequence will be removed to form a new input sequence for the next step. The main benefit of this method is that it can make full use of the model's ability to predict continuously. By dynamically updating the input data, the model can make predictions based on the latest information at every step, thereby improving the accuracy and real-time performance of the prediction. In addition, the single-step prediction method that updates the input also helps the model capture the dynamic changing characteristics of time series data, which is particularly important for understanding and predicting time-dependent data sequences. With this approach, we can perform effective SSM time series forecasting in changing data streams and avoid the observation data being involved in this iterative prediction process.

3. Results

3.1. Comparison of Model Performance in Capturing Soil Moisture Dynamics

We compared the predictions of PB, DL, and PGDL models for the SSM time series on the test set in both single-site and multisite simulations aiming to reveal which model more accurately captures the characteristics of SSM

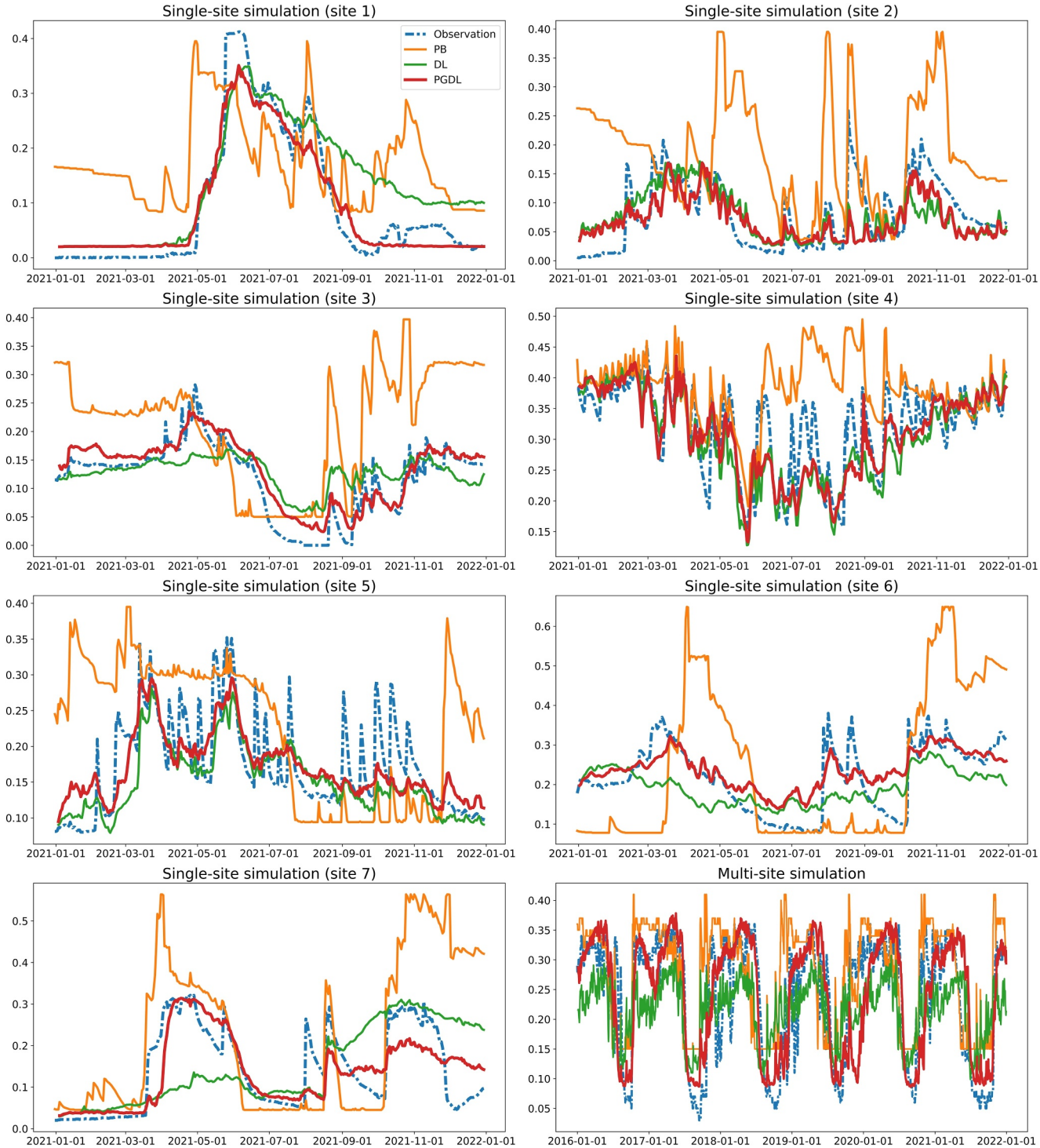


Figure 3. Comparison between surface soil moisture (SSM) observations and model predictions (units: m^3/m^3) on the test set. For each subplot, the *y*-axis is the SSM value, and the *x*-axis is the time. The first seven subplots show the results for the seven sites in the single-site simulation, and the last subplot shows the result for the test site in the multisite simulation.

dynamics. Figure 3 shows the comparison of the observed SSM time series and predictions by the three models on the test set. In each subplot, observations show the actual SSM dynamic changes over the test period, and the three models provide corresponding predictions. By comparing predictions with observations, we can evaluate the

Table 3
Statistical Values for the Model Predictions Compared to Observations

		PB		DL		PGDL	
Simulation	Site	RMSE	R^2	RMSE	R^2	RMSE	R^2
Single-site	1	0.13	−0.15	0.07	0.70	0.03	0.92
	2	0.14	−5.23	0.05	0.35	0.04	0.46
	3	0.13	−3.08	0.05	0.45	0.03	0.82
	4	0.09	−0.79	0.05	0.51	0.04	0.61
	5	0.12	−2.25	0.05	0.43	0.04	0.56
	6	0.17	−3.08	0.07	0.24	0.05	0.58
	7	0.16	−1.63	0.11	−0.14	0.06	0.64
	Mean	0.13	−2.32	0.06	0.36	**0.04**	**0.66**
Multisite		0.09	0.28	0.08	0.38	**0.06**	**0.63**

ability of each model to capture changes in the SSM time series. In Table 3, we use two key statistical indicators: root mean square error (RMSE) and coefficient of determination (R^2) to measure the performance of the models.

Figure 3 shows that for single-site simulation, the observed and simulated SSM values across different vegetation types exhibit dynamic differences. Among them, PGDL model predictions have the highest similarity with the observed data, especially in capturing peaks and fluctuations, showing a more refined response ability than the PB and DL models. In contrast, PB model predictions on most vegetation types deviate greatly from the observed values. During certain time periods at some sites, the PB model prediction is even a straight line, that is, it cannot simulate any dynamic sequence. The predictions of the DL model show improvement compared to the PB model in most vegetation types. However, either from an amplitude or variability perspective, the DL model cannot capture the dynamic changes of the observed SSM values as precisely as the PGDL model. For multisite simulation, all three models capture the interannual variation trend of SSM time series. However, the PGDL model outperforms the PB and DL models in capturing the amplitude of simulated SSM values particularly the annual maximum and minimum values demonstrating higher accuracy and stability.

Table 3 further shows that for single-site simulation, the PGDL model shows superior performance across all vegetation types. Its average RMSE (0.04, unit: m^3/m^3) is significantly lower than that of the PB and DL models, and its R^2 (0.66) is considerably higher. This demonstrates the superiority of the PGDL model in terms of simulation accuracy and reliability, indicating that the PGDL model consistently outperforms the PB and DL models across different vegetation types. For multisite simulation, the PGDL model also achieves the best performance with an RMSE of 0.06 and an R^2 of 0.63 outperforming the corresponding values of the PB and DL models. However, compared to single-site simulation, the prediction accuracy of both the DL and PGDL models decreases in the multisite simulation. This suggests that spatial heterogeneity among multiple sites affects the generalization ability of the models.

3.2. Evaluation of Model Performance Through Scatter Plot Analysis

We used the scatter fit evaluation method to compare and analyze the prediction bias of the PB, DL, and PGDL models in both single-site and multisite simulations. Each set of prediction results corresponds to three scatter plots, showing the comparison between the simulated values and observed values for the three models (Figure 4). To assess the long-term trend and prediction bias of the models, we applied a linear fit to the predicted and observed values of each model using the least squares method with the fitted line represented by a red dashed line. The accuracy and systematic bias of the models were evaluated by comparing the slope and intercept of the fitted line with the ideal values (slope = 1, intercept = 0). A slope close to 1 indicates that the model can accurately capture the trend of the data, whereas an intercept close to 0 indicates that there is no systematic bias within the model. Table 4 presents the average slope and intercept values of the fitted lines for the seven sites in the single-site simulation as well as the slope and intercept of the fitted line for the multisite simulation.

For both single-site and multisite simulations, the PGDL model's average slope is closest to 1, indicating that there is a strong positive correlation between the PGDL predictions and the observed values. At most sites of the single-site simulation, the fitting slope of the PGDL model is higher than that of the corresponding PB and DL models, suggesting that the PGDL model can make the most accurate dynamic predictions for different vegetation types. In addition, the average fitting intercept of the PGDL model is closest to 0, indicating that it has the smallest systematic bias in SSM predictions. Lower fitting intercept values also mean that the model is more accurate in its predictions for simulating extreme low humidity conditions.

For the seven sites of single-site simulation, the PB model's average fitting slope is lower than that of the PGDL model, but it shows better trend matching than the DL model, suggesting that the PB model can better simulate the changing trend of soil moisture in some cases, although its accuracy is not as accurate as the PGDL model. In addition, the average fitting intercept of the PB model is the highest among the three models, indicating larger deviations in predicting low humidity conditions. The higher fitting intercept value indicates that the PB model does not perform as well as the other two models in extreme climate conditions. For the DL model, its average

10.1029/2025JH000682

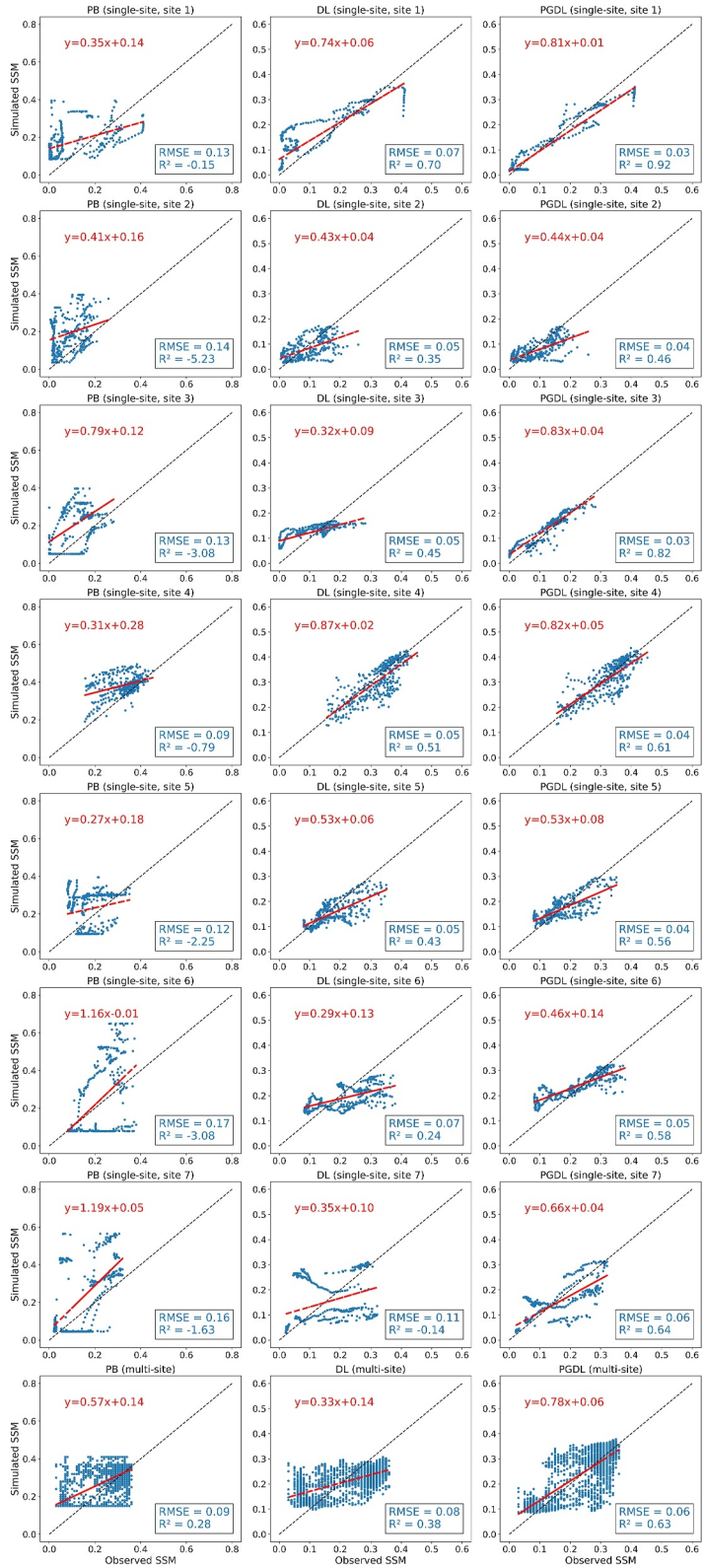


Figure 4.

fitting slope is the lowest among the three models, indicating its lower accuracy in capturing soil moisture change trends. This is because DL models have limitations in handling some specific types of data or complexity. Although the fitting slope accuracy of the DL model is not as good as the other two models, its average fitting intercept is only lower than the PGDL model, suggesting that the DL model performs better than the PB model especially when simulating soil moisture in dry conditions. The performance of these models in multisite simulation is basically consistent with that of single-site simulation.

Overall, the PGDL model is significantly better than the PB and DL models. Our results not only prove the overall superiority of the PGDL model but also reveal the respective advantages and limitations of the PB and DL models. The PB model's high fitting slope indicates that it is relatively robust in simulating changing trends. PB models are able to provide relatively accurate predictions under certain conditions especially simple or stable environmental conditions, because these situations may be more consistent with the physical process assumptions in the PB model. However, the high fitting intercept of the PB model indicates that it reflects a certain systematic bias in its predictions especially when predicting close to zero or low soil moisture values. Because PB models rely on simplified physical processes and parameterizations, it is difficult to accurately capture the dynamical soil moisture in extremely low humidity conditions or in the absence of strong environmental drivers. For the DL model, its low fitting intercept shows the advantage of the DL model in reducing prediction biases when there is no obvious input signal or under extremely low humidity conditions. However, the low fitting slope of the DL model indicates that its accuracy in capturing SSM trends is lower than that of the PB model. This is because when dealing with highly nonlinear and complex time series data especially under extreme climate changes or irregular environmental conditions, the DL models have difficulties capturing all dynamic changes due to limitations in the quality and quantity of training data. Although we adopted the LSTM network, which is better at processing time series and preprocessed labels for noise reduction, the performance of the DL model relies heavily on a large amount of high-quality training data. Therefore, when data are scarce or of low quality, the DL model may not be able to learn sufficiently accurate soil moisture change patterns.

By embedding physical mechanisms into DL models, our PGDL model not only captures complex data relationships but also ensures the physical feasibility and accuracy of model predictions. This combination of physics knowledge and data-driven methods enables PGDL to overcome the lack of adaptability of the PB model in complex environments and the deficiencies of the DL model in system deviation control in simulating SSM.

4. Discussions

4.1. Comparison of DL and PGDL Models in Representing Physical Processes

Our results indicate that the PGDL model provides more accurate SSM time series predictions at the site level than the DL model. However, our purpose in developing such a PGDL model is not only to provide more accurate predictions but, more importantly, we expect that after embedding the guidance of physical mechanisms, the PGDL will be more consistent with physical laws than the DL model. The water balance equation (Equation 1) used for SSM calculation in the PB model represent the basic physical processes that dominate SSM dynamics. In Equation 1, only I_F will increase SSM, whereas the other three terms (E_V, E_S, and D_R) will decrease SSM. Therefore, within a period of time, if I_F is continuously zero and E_V and E_S are both not continuously zero, SSM will not increase. Otherwise, it will violate the physical laws of water balance. Based on this, we conducted a model experiment. We first found all time periods that satisfy I_F to be continuously zero and E_V and E_S not to be continuously zero in each test set. For single-site simulation, we set the threshold to 7 days, while for multi-site simulation, it was set to 15 days to better capture long-term moisture fluctuation patterns. For all time periods that satisfy this condition, we further checked whether the SSM observations increase during this period. If not, we

Figure 4. Comparative analysis of surface soil moisture (SSM) time series observations and model predictions on the test set. Each row represents a site with the first seven rows corresponding to the seven sites from the single-site simulation and the last row corresponding to the test site from the multisite simulation. The three columns from left to right are the comparisons between PB, DL, and physics-guided deep learning model predictions and observed values. For each subplot, the *y*-axis is the simulated SSM values (units: m^3/m^3), and the *x*-axis is the observed SSM values (units: m^3/m^3). The red dotted line is a straight line fitted based on the least squares method describing the linear relationship between the observed values and the corresponding predicted values. The equation in red is the functional expression corresponding to the fitted straight line. The evaluation metrics, including root mean square error and R^2, are shown in the black box in the lower right corner quantifying the accuracy of the corresponding model predictions.

Table 4
Slope and Intercept Values of the Fitted Lines Comparing Observed and Predicted Values for PB, DL, and PGDL Models on the Test Set

Simulation	Metrics	PB	DL	PGDL
Single-site	Slope (mean)	0.64	0.50	0.65
	Intercept (mean)	0.13	0.07	0.06
Multisite	Slope	0.57	0.33	0.78
	Intercept	0.14	0.14	0.06

Note. For single-site simulations, the values represent the mean across the seven sites.

counted the number of times the SSM value predicted by the DL and PGDL models increases, respectively. This count is the number of times that the physical law of water balance is violated in the model predictions (Table 5).

We found that for both single-site and multisite simulations, the PGDL model violated the physical laws of water balance fewer times than the DL model (except for Site 1 in the single-site simulation, where neither model violated the physical laws). Such results indicate that the PGDL model is more consistent with the physical laws of water balance than the DL model when simulating soil moisture dynamics. This consistency is achieved through the feature extraction and representation process of LSTM and FCNN in the PGDL model. The LSTM branch captures temporal dependencies and change patterns in time series data, whereas the FCNN branch processes current physical information to capture complex nonlinear relationships. Despite not explicitly enforcing constraints such as the water balance equation, the PGDL model can effectively learn the physical laws expressed in the observations during the training process, thereby achieving better results than the traditional DL model.

The consistency of PGDL models with physical laws represents potential benefits. First, it increases confidence in SSM predictions because the behavior of the PGDL model is consistent with the underlying physical processes. By combining the strengths of LSTM and FCNN, the PGDL model can internalize physical processes during training leading to more accurate and reliable predictions. This consistency enhances the interpretability of the model making it easier for researchers to understand and verify the model outputs. Second, the PGDL model's compliance with physical laws improves its ability to extrapolate new scenarios with greater confidence. By learning the fundamental relationships that govern SSM dynamics, PGDL models can better handle unseen conditions, thereby reducing the risk of producing SSM predictions that do not comply with physical laws.

4.2. Effects of Training Data Sizes on DL and PGDL Model Performance in Single-Site Simulation

When exploring the performance of time series forecasting models, it is critical to understand the impact of training data volume on model forecast accuracy. The time span and quality of training data directly affect the reliability of forecast results. Due to the limited availability of data in practical applications, researchers are faced with the challenge of using limited data sets to train and optimize models. Therefore, evaluating the impact of training data of different time spans on model performance is crucial to building more robust and adaptable predictive models.

To explore the sensitivity of DL and PGDL models to changes in training data volume when processing time series data, this study designed a series of comparative experiments for single-site simulation aiming to compare the impact of different training data amounts on model prediction performance at the same site. Specifically, the data from year 2021 were used as the test set to show the comparison results. The difference is that in addition to using data from 2016 to 2020 (5 years in total) as the training set (as shown in Figure 3 and Table 3), we also used 2017 to 2020 (4 years), 2018 to 2020 (3 years), 2019 to 2020 (2 years), and only 2020 (1 year) as training sets. We then compared the prediction performance of the model under different training sets. This experiment was conducted at the seven sites representing distinct vegetation types ensuring the broad applicability and reliability of the research results (Table 6).

Table 5
The Number of Days That Do Not Satisfy the Physical Mechanism of Water Balance by DL and PGDL Models

Simulation	Site	DL model (#)	PGDL model (#)
Single-site	1	0	0
	2	4	2
	3	13	7
	4	6	4
	5	11	7
	6	6	3
	7	13	4
	Total	53	27
Multisite		70	42

Under the same number of training sets, the performance of the PGDL model is always better than that of the DL model (Table 6), indicating that the PGDL model has stronger data fitting and generalization capabilities. By integrating physical processes within the neural network architecture, PGDL models are more effective and robust in capturing SSM dynamics. In addition, both the DL and PGDL models exhibit the best average performance when using the most training data (i.e., 5 years of training data). As the amount of training data increases, the models learn more features and trends, thereby improving the accuracy and interpretability of predictions. PGDL can be used to extract information from large amounts of data more effectively and achieve higher predictive performance. In addition, the DL model shows better performance

Table 6
Comparisons of DL and PGDL Models Using Different Numbers of Years of the Training Set Across the Seven Sites in Single-Site Simulations

Training set (#)		5 years		4 years		3 years		2 years		1 year	
Model	Site	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
DL	1	0.07	0.70	0.12	0.10	0.12	0.05	0.06	0.77	0.03	0.94
	2	0.05	0.35	0.05	0.26	0.05	0.19	0.06	0.00	0.06	0.07
	3	0.05	0.45	0.08	−0.67	0.07	−0.40	0.05	0.42	0.03	0.70
	4	0.05	0.51	0.05	0.50	0.05	0.40	0.05	0.39	0.04	0.58
	5	0.05	0.43	0.05	0.51	0.04	0.54	0.04	0.53	0.05	0.32
	6	0.07	0.24	0.07	0.35	0.07	0.34	0.10	−0.49	0.09	−0.04
	7	0.11	−0.14	0.13	−0.64	0.10	−0.03	0.10	0.00	0.13	−0.59
	Mean	0.06	0.36	0.08	0.06	0.07	0.16	0.07	0.23	0.06	0.28
PGDL	1	0.03	0.92	0.07	0.67	0.05	0.83	0.04	0.87	0.03	0.92
	2	0.04	0.46	0.05	0.35	0.05	0.30	0.05	0.09	0.06	0.05
	3	0.03	0.82	0.04	0.59	0.03	0.81	0.03	0.73	0.04	0.68
	4	0.04	0.61	0.05	0.54	0.05	0.46	0.05	0.47	0.05	0.52
	5	0.04	0.56	0.05	0.51	0.05	0.51	0.05	0.47	0.06	0.28
	6	0.05	0.58	0.06	0.48	0.07	0.36	0.07	0.24	0.09	−0.12
	7	0.06	0.64	0.08	0.31	0.09	0.24	0.08	0.31	0.11	−0.12
	Mean	0.04	0.66	0.06	0.49	0.06	0.50	0.05	0.45	0.06	0.32

when using 2 years of training data compared to using 3 and 4 years. This suggests that the quality and relevance of the training data may be more important than the quantity in certain situations. Using smaller but more representative data sets can help the models avoid learning unnecessary noise or irrelevant patterns leading to better predictions. This is particularly relevant when the most recent data are more like the patterns observed in the test set, as they may provide more informative cues for predicting future trends.

Based on these findings, we recommend that when the data availability is scarce, prioritizing the quality and relevance of the training data over the quantity may yield better results. Strategies such as careful data selection, preprocessing, and augmentation can help maximize the information content of the available data, enabling the models to learn more effectively. Additionally, the PGDL model's consistent superiority over the DL model suggests that incorporating physical knowledge can provide a robust framework under limited data conditions making it a preferred choice for SSM prediction.

In summary, in single-site simulations, the PGDL model demonstrates a strong predictive ability and higher stability in using complex time series data even with reduced training data sets. Its superior performance over the DL model highlights the benefits of integrating physical principles into data-driven approaches, improving the accuracy of soil moisture prediction with limited data availability.

4.3. Enhancing Interpretability and Performance of PGDL

DL models often appear as a “black box” in their quantification. By analyzing the feature importance of the trained PGDL model, we can identify which input features have the greatest impact on the model predictions to reveal the driving factors behind the model predictions, thereby improving the transparency and credibility of the model.

There are eight dynamic features used for PGDL predictions: Pr, SWD, T_{air}, V_P, LAI, I_F, E_V, and E_S. We rank the importance of features by evaluating the impact of each feature on the model predictions. Based on the trained PGDL model, we add different multiples of its standard deviation to each feature to simulate different degrees of feature perturbation. Specifically, we perturb each feature with its corresponding standard deviation of 0.1 times, 0.5 times, 1 time, and 2 times as the perturbation amount and then use these perturbed features to re-perform PGDL prediction to observe the impact of feature perturbation. By comparing the difference between the

Table 7
The Importance Ranking of Features for the PGDL Model Under Four Situations (Adding 0.1 SD, 0.5 SD, 1 SD, and 2 SD Perturbation) in Single-Site Simulations

Site	0.1 SD perturbation	0.5 SD perturbation	1 SD perturbation	2 SD perturbation
1	[8, 6, 4, 3, 5, 1, 2, 7]	[8, 7, 6, 4, 2, 3, 5, 1]	[7, 2, 8, 4, 6, 3, 5, 1]	[7, 2, 1, 3, 8, 4, 6, 5]
2	[7, 8, 2, 4, 5, 1, 6, 3]	[7, 8, 2, 4, 5, 1, 6, 3]	[7, 8, 2, 4, 5, 1, 6, 3]	[7, 8, 4, 5, 2, 3, 1, 6]
3	[8, 7, 5, 3, 4, 1, 6, 2]	[8, 7, 5, 6, 3, 4, 1, 2]	[8, 7, 6, 5, 3, 4, 2, 1]	[8, 7, 6, 3, 5, 4, 2, 1]
4	[8, 7, 6, 5, 3, 1, 4, 2]	[8, 7, 6, 5, 3, 2, 4, 1]	[8, 7, 5, 6, 3, 2, 4, 1]	[8, 7, 5, 3, 2, 6, 4, 1]
5	[6, 8, 5, 4, 1, 2, 3, 7]	[8, 4, 6, 5, 7, 3, 1, 2]	[8, 6, 4, 5, 3, 7, 1, 2]	[8, 5, 6, 3, 4, 7, 1, 2]
6	[7, 4, 8, 6, 5, 2, 1, 3]	[8, 7, 4, 5, 6, 2, 3, 1]	[8, 7, 5, 4, 3, 2, 6, 1]	[8, 7, 5, 3, 2, 6, 4, 1]
7	[6, 4, 7, 5, 2, 1, 3, 8]	[6, 8, 4, 3, 7, 2, 1, 5]	[6, 4, 8, 7, 3, 1, 5, 2]	[6, 4, 8, 7, 5, 1, 2, 3]
Mean	[8, 6, 7, 4, 5, 1, 3, 2]	[8, 7, 6, 4, 5, 3, 2, 1]	[8, 7, 6, 4, 5, 3, 2, 1]	[8, 7, 5, 3, 6, 4, 2, 1]

Note. Numbers 1–8 in square brackets represent Pr, SWD, T_{air}, V_P, LAI, I_F, E_V, and E_S, respectively.

original output of the model and the output after perturbing the feature (measured as mean absolute error, MAE), we calculate an importance score for each feature under different standard deviation perturbations. This score reflects the degree to which the perturbation of this feature affects the predictions. The higher the score, the greater the impact of feature changes on the predictions. To determine which features are most critical to the model predictions, we rank the features based on their importance scores.

We evaluated the seven sites in the single-site simulation. By weighing the feature importance scores of all models, we obtained the comprehensive importance ranking of the features under each standard deviation multiple condition (Table 7). This feature importance analysis was only conducted for single-site simulations because the multisite simulation involved multiple sites with spatial heterogeneity leading to greater variation in learning patterns. This spatial variability makes it challenging to isolate the impact of each input feature on test set predictions, hindering the ability to draw meaningful conclusions. By focusing on the single-site simulation, we maintained spatial consistency between the training and test data ensuring more reliable and interpretable feature importance analysis.

Three features used in the LSTM part of the PGDL model (I_F, E_V, and E_S, corresponding to 6 to 8 in Table 7) basically rank higher in importance than five features used in the FCNN part (Pr, SWD, T_{air}, V_P, and LAI, corresponding to 1 to 5 in Table 7). This implies that the PGDL model relies more on the three input features of the LSTM when making predictions. This finding is consistent with the physical process of calculating SSM in the PB model where I_F, E_V, and E_S directly determines SSM from the water balance equation (Equation 1). The high importance of these features in the PGDL model indicates that the model has successfully learned the underlying physical processes that control SSM dynamics.

In addition, among the three features of the LSTM part, the PGDL model gives higher importance to E_V and E_S compared with I_F. This can be attributed to a faster response of the evaporation process to changes in environmental conditions since evaporation occurs almost every day and has a decisive influence on the daily changes in SSM. In contrast, water infiltration into surface soil is more intermittent occurring primarily after rainfall events. For the five features of the FCNN part, the PGDL model considers T_{air}, V_P, and LAI to be more important than Pr and SWD. This is consistent with the physical process of calculating SSM in the PB model (Bonan, 1996; Zhuang et al., 2002, 2004) where T_{air}, V_P, and LAI participate in the calculation of at least two of the three features in the LSTM part, whereas Pr and SWD only contribute to the calculation of I_F and E_S, respectively.

The feature importance analysis of the trained PGDL model not only verifies that the PGDL model learned data features are consistent with the physical process but also provides valuable insights into the relative impact of different input features on model predictions. By ranking features based on important scores, we can identify key drivers of soil moisture dynamics learned by the model, improve its interpretability, and guide future model development and data collection efforts.

4.4. Limitations and Future Directions of Multisite Simulation

In this study, we selected 13 sites under the same vegetation type (grassland) for multisite simulation and identified FryCanyon (SNOTEL) as the test set using a two-step strategy. The experimental results show that the PGDL model outperformed the PB and DL models on this test set confirming its advantage in capturing SSM dynamic changes. However, during cross-validation, we found that the performance of the DL and PGDL models was not stable, and in some cases (especially for the DL model), the prediction accuracy did not meet expectations. This indicates that the current multisite simulation has certain limitations.

One possible reason for this is that we screened the sites solely based on vegetation type. Vegetation type is generally considered an important factor influencing SSM dynamics, as it reflects the moisture response patterns and seasonal variations among sites. Therefore, we collected data by vegetation type expecting to better train and validate the model using data from multiple sites within the same vegetation type. However, the results suggest that vegetation type may not be sufficient to effectively represent the spatial variability of SSM across different sites. Even within the same vegetation type, significant differences in SSM dynamics can occur due to variations in climate, soil type, and topography. This spatial heterogeneity may limit the model's ability to generalize across different sites.

Additionally, the current validation strategy of multisite simulation may not be sufficient to fully evaluate the generalization ability of the model when there are large spatial differences in SSM. Spatial heterogeneity among sites can result in large differences between the validation and training sets leading to fluctuations in model performance during validation and testing. Moreover, DL models heavily rely on temporal continuity when learning time series patterns. When the time series patterns differ greatly among sites, the model may struggle to generalize effectively. Although the PGDL model improves generalization by incorporating physical constraints, its performance may still be limited when spatial differences are large and the static features are not sufficient to fully characterize the dominant drivers of SSM dynamics.

To address these limitations, more detailed spatial classification strategies can be explored in future studies. Although this study collected data based on the same and typical vegetation type, the results indicate that relying on a single classification feature may not be sufficient to fully capture the dynamic differences in SSM across different sites. Therefore, future research could consider including additional spatial features, such as soil type, climate region, and topographic characteristics, to perform more refined spatial clustering. This approach could improve the model's generalization ability and prediction accuracy leading to a more robust definition of training and test sets in multisite simulations and enhancing the applicability and stability of DL and PGDL models in spatially heterogeneous environments.

5. Conclusion

This study introduces a PGDL model that combines the deterministic physical processes of TEM with the predictive efficiency of LSTM networks. This innovative approach improves the accuracy and reliability of SSM predictions at seven sites representing different vegetation types in single-site simulations and 13 sites within the same vegetation type in multisite simulation in comparison with traditional PB model and independent DL model.

The leap in prediction accuracy is attributed to the fact that the PGDL model combines the ability of LSTM to capture long-term data dependencies with the physical process-based insights of TEM, thereby facilitating a more nuanced understanding and modeling of SSM dynamics. Compared with the DL model, the PGDL model is more consistent with the basic physical laws of water balance indicating its ability to learn knowledge about physical processes. Additionally, in single-site simulations, PGDL models exhibit enhanced robustness and prediction accuracy across varying amounts of training data resulting in better generalization and performance even with limited or varying data quality. Feature importance analysis further reveals that the PGDL model predominantly relies on input features directly related to the water balance equation emphasizing the model's ability to prioritize and learn from the most influential variables. This enhances the interpretability of the PGDL model making its prediction process more transparent and credible. Finally, we identified limitations within the multisite simulation in this study and provided insights and directions for future research to address these challenges.

By integrating physical processes into the neural network architecture, our PGDL model effectively bridges the gap between complex deep learning algorithms and traditional process-based methods, thereby providing a promising way to improve the accuracy, reliability, and interpretability of SSM predictions.

Data Availability Statement

The codes and data for this study are available at the Purdue University Research Repository (Xi et al., 2025).

Acknowledgments
Zhuang was funded by NASA through a subcontract from JPL 1609311.

References

Abowarda, A. S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., & Sun, Z. (2021). Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. *Remote Sensing of Environment*, *255*, 112301. https://doi.org/10.1016/j.rse.2021.112301

Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M., & Ghalhari, G. A. F. (2020). Machine learning to estimate surface soil moisture from remote sensing data. *Water*, *12*(11), 3223. https://doi.org/10.3390/w12113223

Arhonditsis, G. B., & Brett, M. T. (2004). Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Marine Ecology Progress Series*, *271*, 13–26. https://doi.org/10.3354/meps271013

Bonan, G. B. (1996). Land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and users guide. Technical note (No. PB-97-131494/XAB; NCAR/TN-417-STR). National Center for Atmospheric Research, Boulder, CO (United States). Climate and Global Dynamics Div.

Breen, K. H., James, S. C., White, J. D., Allen, P. M., & Arnold, J. G. (2020). A hybrid artificial neural network to estimate soil moisture using swat+ and SMAP data. *Machine Learning and Knowledge Extraction*, *2*(3), 16–306. https://doi.org/10.3390/make2030016

Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L., & Xue, X. (2019). Research on soil moisture prediction model based on deep learning. *PLoS One*, *14*(4), e0214508. https://doi.org/10.1371/journal.pone.0214508

Celik, M. F., Isik, M. S., Yuzugullu, O., Fajraoui, N., & Erten, E. (2022). Soil moisture prediction from remote sensing images coupled with climate, soil texture and topography via deep learning. *Remote Sensing*, *14*(21), 5584. https://doi.org/10.3390/rs14215584

Chatterjee, S., Desai, A. R., Zhu, J., Townsend, P. A., & Huang, J. (2022). Soil moisture as an essential component for delineating and forecasting agricultural rather than meteorological drought. *Remote Sensing of Environment*, *269*, 112833. https://doi.org/10.1016/j.rse.2021.112833

Clark, M. P., Bierkens, M. F., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., et al. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, *21*(7), 3427–3440. https://doi.org/10.5194/hess-21-3427-2017

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., et al. (2016). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research*, *52*(3), 2350–2365. https://doi.org/10.1002/2015wr017910

Doherty, J. (2004). *PEST model-independent parameter estimation user manual* (Vol. 3338, p. 3349). Watermark Numerical Computing.

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., et al. (2011). The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, *15*(5), 1675–1698. https://doi.org/10.5194/hess-15-1675-2011

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiova, A., Sanchis-Dufau, A. D., et al. (2013). Global automated quality control of in situ soil moisture data from the International Soil Moisture Network. *Vadose Zone Journal*, *12*(3), 1–21. https://doi.org/10.2136/vzj2012.0097

Fan, Y., & Van Den Dool, H. (2004). Climate Prediction Center global monthly soil moisture data set at 0.5 resolution for 1948 to present. *Journal of Geophysical Research: Atmospheres*, *109*(D10). https://doi.org/10.1029/2003jd004345

Geng, Q., Yan, S., Li, Q., & Zhang, C. (2024). Enhancing data-driven soil moisture modeling with physically-guided LSTM networks. *Frontiers in Forests and Global Change*, *7*, 1353011. https://doi.org/10.3389/ffgc.2024.1353011

Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., & Gentine, P. (2019). Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature*, *565*(7740), 476–479.

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes: International Journal*, *22*(18), 3802–3813. https://doi.org/10.1002/hyp.6989

Hairer, E., Nørsett, S. P., & Wanner, G. (1993). *Solving ordinary differential equations I: Nonstiff problems* (2nd ed.). Springer.

Han, Q., Zeng, Y., Zhang, L., Wang, C., Prikaziuk, E., Niu, Z., & Su, B. (2023). Global long term daily 1 km surface soil moisture dataset with physics informed machine learning. *Scientific Data*, *10*(1), 101. https://doi.org/10.1038/s41597-023-02011-7

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12. https://doi.org/10.1021/ci0342472

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., & Rozum, I. (2023). ERA5 hourly data on single levels from 1959 to present—Copernicus Climate Change Service (C3S) Climate Data Store (CDS).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Humphrey, V., Berg, A., Ciais, P., Gentine, P., Jung, M., Reichstein, M., et al. (2021). Soil moisture–atmosphere feedback dominates land carbon uptake variability. *Nature*, *592*(7852), 65–69. https://doi.org/10.1038/s41586-021-03325-5

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., et al. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2331. https://doi.org/10.1109/tkde.2017.2720168

Kashinath, K., Mustafa, M., Albert, A., Wu, J. L., Jiang, C., Esmaeilzadeh, S., et al. (2021). Physics-informed machine learning: Case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200093. https://doi.org/10.1098/rsta.2020.0093

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Koster, R. D., & Suarez, M. J. (2001). Soil moisture memory in climate models. *Journal of Hydrometeorology*, *2*(6), 558–570. https://doi.org/10.1175/1525-7541(2001)002<0558:smmicm>2.0.co;2

Lange, S., Mengel, M., Treu, S., & Büchner, M. (2022). ISIMIP3a atmospheric climate input data (v1.0). *ISIMIP Repository*. https://doi.org/10.48364/ISIMIP.982724

Li, L., Dai, Y., Wei, Z., Shangguan, W., Wei, N., Zhang, Y., et al. (2024). Enhancing deep learning soil moisture forecasting models by integrating physics-based models. *Advances in Atmospheric Sciences*, *41*(7), 1–16. https://doi.org/10.1007/s00376-023-3181-8

Liu, L., Xu, S., Tang, J., Guan, K., Griffis, T. J., Erickson, M. D., et al. (2022). KGML-Ag: A modeling framework of knowledge-guided machine learning to simulate agroecosystems: A case study of estimating N_2O emission using data from mesocosm experiments. *Geoscientific Model Development*, *15*(7), 2839–2858. https://doi.org/10.5194/gmd-15-2839-2022

Mo, K. C., Shukla, S., Lettenmaier, D. P., & Chen, L. C. (2012). Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction? *Geophysical Research Letters*, *39*(23). https://doi.org/10.1029/2012gl053598

Montanari, A. (2007). What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrological Processes: International Journal*, *21*(6), 841–845. https://doi.org/10.1002/hyp.6623

Porporato, A., D'odorico, P., Laio, F., Ridolfi, L., & Rodriguez-Iturbe, I. (2002). Ecohydrology of water-controlled ecosystems. *Advances in Water Resources*, *25*(8–12), 1335–1348. https://doi.org/10.1016/s0309-1708(02)00058-1

Rasheed, K., Qayyum, A., Ghaly, M., Al-Fuqaha, A., Razi, A., & Qadir, J. (2022). Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Computers in Biology and Medicine*, *149*, 106043. https://doi.org/10.1016/j.compbiomed.2022.106043

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., et al. (2019). Process-guided deep learning predictions of lake water temperature. *Water Resources Research*, *55*(11), 9173–9190. https://doi.org/10.1029/2019wr024922

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Rodriguez-Iturbe, I. (2000). Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamies. *Water Resources Research*, *36*(1), 3–9. https://doi.org/10.1029/1999wr900210

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Wang, N., Zhang, D., Chang, H., & Li, H. (2020). Deep learning of subsurface flow via theory-guided neural network. *Journal of Hydrology*, *584*, 124700. https://doi.org/10.1016/j.jhydrol.2020.124700

Waring, R. H., & Running, S. W. (2010). *Forest ecosystems: analysis at multiple scales*. Elsevier.

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, *55*(4), 1–37. https://doi.org/10.1145/3514228

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L. P. H., Bierkens, M. F., Blyth, E., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, *47*(5). https://doi.org/10.1029/2010wr010090

Xia, Y., Sheffield, J., Ek, M. B., Dong, J., Chaney, N., Wei, H., & Wood, E. F. (2014). Evaluation of multi-model simulated soil moisture in NLDAS-2. *Journal of Hydrology*, *512*, 107–125. https://doi.org/10.1016/j.jhydrol.2014.02.027

Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., & Shen, C. (2021). Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *Journal of Hydrology*, *603*, 127043. https://doi.org/10.1016/j.jhydrol.2021.127043

Xie, W., Kimura, M., Takaki, K., Asada, Y., Iida, T., & Jia, X. (2022). Interpretable framework of physics-guided neural network with attention mechanism: Simulating paddy field water temperature variations. *Water Resources Research*, *58*(5), e2021WR030493. https://doi.org/10.1029/2021wr030493

Xi, X., Zhuang, Q., & Liu, X. (2025). *A hybrid physics-guided deep learning modeling framework for predicting surface soil moisture. (Version 3.0)*. Purdue University Research Repository. https://doi.org/10.4231/XQ9H-1S38

Zanetta, F., Nerini, D., Beucler, T., & Liniger, M. A. (2023). Physics-constrained deep learning postprocessing of temperature and humidity. *Artificial Intelligence for the Earth Systems*, *2*(4), e220089. https://doi.org/10.1175/aies-d-22-0089.1

Zhou, S., Williams, A. P., Lintner, B. R., Berg, A. M., Zhang, Y., Keenan, T. F., et al. (2021). Soil moisture–atmosphere feedbacks mitigate declining water availability in drylands. *Nature Climate Change*, *11*(1), 38–44. https://doi.org/10.1038/s41558-020-00945-z

Zhuang, Q., McGuire, A. D., Melillo, J. M., Clein, J. S., Dargaville, R. J., Kicklighter, D. W., et al. (2011). Carbon cycling in extratropical terrestrial ecosystems of the northern hemisphere during the 20th century: A modeling analysis of the influences of soil thermal dynamics. *Tellus B: Chemical and Physical Meteorology*, *55*(3), 751–776. https://doi.org/10.3402/tellusb.v55i3.16368

Zhuang, Q., McGuire, A. D., O'neill, K. P., Harden, J. W., Romanovsky, V. E., & Yarie, J. (2002). Modeling soil thermal and carbon dynamics of a fire chronosequence in interior Alaska. *Journal of Geophysical Research*, *107*(D1), FFR-3. https://doi.org/10.1029/2001jd001244

Zhuang, Q., Melillo, J. M., Kicklighter, D. W., Prinn, R. G., McGuire, A. D., Steudler, P. A., & Hu, S. (2004). Methane fluxes between terrestrial ecosystems and the atmosphere at northern high latitudes during the past century: A retrospective analysis with a process-based biogeochemistry model. *Global Biogeochemical Cycles*, *18*(3). https://doi.org/10.1029/2004gb002239