Interpretable Soil Moisture Prediction with a Knowledge-guided Deep Learning Approach

Yanling Wang[1], Xiaolong Hu[1*], Yaan Hu[2], Leilei He[1], Lijun Wang[1], Wenxiang Song[1], Liangsheng Shi[1*]

[1]State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, China

[2]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, China

Corresponding author: Xiaolong. Hu (xlhu@whu.edu.cn), Liangsheng Shi (liangshs@whu.edu.cn)

Abstract. Soil moisture (SM) is a critical component of the hydrological cycle, but accurately predicting it remains challenging due to the nonlinearity of soil water transport, variability in boundary conditions, and the intricate nature of soil properties. Recently, deep learning has shown promise in this domain, typically by modeling temporal dependencies for soil moisture predictions. In this study, we propose non-local neural networks (NLNNs) to convert this problem into a single-time-step, simultaneous multi-depth soil moisture forecasting. The non-local operation design includes embedded Gaussian operations and disentangled knowledge-guided operations, resulting in two variants: the self-attention non-local neural network (SA-NLNN) and the knowledge-guided non-local neural network (KG-NLNN). The knowledge-guided non-local operation is designed to capture vertical soil moisture relationships by decomposing the influences on soil moisture at a given depth into four components, each governed by distinct physical processes. The models offer visual interpretability through learned non-local weights, which reveal interactions among soil moisture across different depths, thereby enabling a qualitative representation of inter-layer connectivity. Notably, the model guided by soil moisture transport knowledge yields more stable and reasonable interpretations. With in-situ observations, we demonstrate that our proposed models perform satisfactorily. The knowledge-guided non-local operations significantly enhance accuracy and reliability. Additionally, our models adapt to diverse time-scale situations while maintaining high computational efficiency. Both models exhibit robust noise resistance, with knowledge guidance enhancing KG-NLNN's noise resistance. In summary, our work addresses the

soil moisture prediction challenge in a novel way, highlighting the potential of NLNN and the importance of incorporating physic guidance in data-driven models.

Keywords: soil moisture; deep learning; non-local neural networks; knowledge-guided; visual interpretability

1.Introduction

Soil moisture plays an important role in hydrological processes, governing the exchange of water and energy fluxes between the atmosphere and the land (Vereecken et al., 2008). Accurate simulations of soil moisture dynamics hold great significance in various domains, including effective water resources planning and management, agricultural production, and flood disaster monitoring (Entekhabi et al., 1996; Koster et al., 2004; Zhang et al., 2018). However, precisely forecasting soil moisture dynamics poses challenges due to the nonlinearity of soil water transport (Richards, 1931), randomness in boundary conditions (Guswa et al., 2002), and the intricate nature of soil properties, including soil structure and hydraulic parameters (Vereecken et al., 2022). These factors contribute to strong spatio-temporal variabilities in soil moisture dynamics (Heathman et al., 2012). Traditionally, the simulation of soil moisture dynamics has primarily relied on physically based models, such as the soil-plant-atmosphere-water model (Saxton et al., 1974) and HYDRUS (Simunek et al., 2005). However, their implementation faces challenges in accurately estimating the required parameters (Bandai & Ghezzehei, 2021; Gill et al., 2006). What’s more, the current methodology struggles to accurately characterize soil structure at spatially relevant scales (Romero-Ruiz et al., 2018). This limitation complicates handling scenarios involving cracks, root water absorption, and other complexities, as illustrated in Figure 1. With advancements in technology and big data analysis capabilities, data-driven models have aroused increasing focus and appear to be more practical in soil moisture dynamics forecasting. For instance, researchers have discovered that both support vector regression and random forest show satisfactory results in soil moisture prediction while maintaining low computing costs (Gill et al., 2006; Prasad et al., 2019). Furthermore, the extreme learning machine (Huang et al., 2006) has demonstrated its capability to precisely predict soil moisture trends (Y. Liu et al., 2014).

In recent years, deep learning (Lecun et al., 2015) has gained considerable attention for its remarkable capabilities in fitting to complex data patterns. When predicting soil moisture, deep learning primarily relies on modeling temporal dependencies. The fundamental models handling sequential data fall into three categories: Recurrent Neural Networks (RNNs) (Elman, 1990), Convolution Neural Networks (CNNs) (LeCun, 1989), and Transformers (Vaswani et al., 2017). RNNs exploit temporal dependencies through recurrent operations, with Long Short-Term Memory (LSTM) networks demonstrating accurate soil moisture predictions (Fang et al., 2019). CNNs capture dependencies with repetitive convolutional operations and also yield satisfactory results in soil moisture dynamics modeling (Severyn & Moschitti, 2015; Shi et al., 2015). Both recurrent and convolutional operations process local neighborhoods in input data. Consequently, long-range dependencies are captured through repeated local operations, which is inefficient (L. Zhu et al., 2021). In contrast, Transformers process data in a more efficient way, owing to its core component – self-attention mechanisms. These mechanisms extract crucial long-range non-local information directly. For instance, Temporal Fusion Transformers with interpretable self-attention layers have shown significant improvements over existing benchmarks in multi-horizon time series forecasting (Lim et al., 2021). Furthermore, Transformers exhibit potential for effective soil moisture dynamics prediction with straightforward model structures (Y. Wang et al., 2023). Researchers are increasingly recognizing the potential of Transformers.

However, it is worth noting that current deep learning models often lack physical laws and interpretability. To bridge the gap between data-driven approaches and physics, physical principles can be embedded into loss functions or model architectures. Some researchers have added the residuals of governing physical equations to the loss function, giving rise to Physics-informed Neural Networks (PINN) (M. Raissi et al., 2019; Maziar Raissi et al., 2017). In terms of model architectures, Jiang et al. (2020) integrated the physical processes from a conceptual hydrological model into an RNN for runoff modeling. De Bézenac et al. (2018) incorporated advection-diffusion principles into the kernel design of a CNN to predict sea surface temperature. To date, most previous works have relied on traditional model structures, leaving a critical gap in reliable data-driven methods for soil moisture prediction. This underscores the necessity of transitioning toward soil science-informed machine learning models that use the power of data-driven techniques while integrating soil science knowledge during the training process to enhance reliability and generalizability (Minasny et al., 2024).

Considering that physical models calculate soil moisture content by iteratively using current soil profile states for stepwise predictions, we incorporate the spatial interactions of soil moisture within the profile into our machine learning model. We intend to update soil moisture at each depth based on the states of all depths, with predictions computed as a weighted aggregation of the previous states. When dealing with relationships between multiple variables, geometric deep learning (Bronstein et al., 2017) defines model invariances to enhance robustness and generalization. As an example, graph neural networks (GNNs) (Scarselli et al., 2008) utilize the adjacency matrix to aggregate node features and achieve local invariance. Wang et al. proposes a spatiotemporal graph convolutional network that models inter-station relationships to effectively predict soil moisture (W. Wang et al., 2025). While GNNs aggregate information through graph-structured neighborhood relationships, Non-local Neural Networks (NLNNs) directly model pairwise dependencies among all positions (X. Wang et al., 2018). This fully connected interaction pattern allows each position to directly interact with all other positions, thereby enabling the model to capture long-range global dependencies. The interaction weights are adaptively determined by the real-time soil moisture state in a fully data-driven manner. This fundamental difference reflects distinct inductive biases: GNNs rely on graph-structured message passing, whereas NLNNs explicitly model global interactions without neighborhood restrictions. For soil moisture dynamics, where relevant dependencies may exist between distant soil layers and vary over time, such global modeling capability is particularly beneficial. The non-local operation in NLNNs calculates responses at specific locations by aggregating features from all positions in the input feature map (X. Wang et al., 2018). This design allows NLNNs to flexibly model global relationships in a data-driven manner, making them suitable as a general modeling module for various tasks. Considering the complexity of interactions between multi-depth soil moisture, we introduce the NLNNs to capture spatially invariant soil moisture relationships across soil layers. Our objective is to model vertical heterogeneity and inter-layer connectivity without physical assumptions. Moreover, the weights computed through non-local operations provide qualitative interpretation for model learning mechanisms. NLNNs find wide application in image segmentation tasks and time series forecasting (P. Liu et al., 2019; Z. Zhu et al., 2019). As a representative of NLNNs, the Transformer is adept at processing various types of data, including images and video-related challenges (Guo et al., 2022; Khan et al., 2022; Lim et al., 2021; Z. Liu et al., 2021; Xie et al., 2021). Furthermore, NLNNs can serve as auxiliary blocks to enhance context modeling abilities (X. Wang et al., 2018; Yin et al., 2020). With the flexibility of non-local operation modifications, we can envision using NLNNs to simulate the characteristics of soil water dynamics in spatial distribution while ensuring interpretability.

In this study, we have integrated NLNNs to simulate in-profile soil moisture interactions and predict multi-depth soil moisture content without physical assumptions. Our aim is to achieve accurate and effective forecasts under diverse real-world scenarios, as depicted in Figure 1, while also providing qualitative description of intricate soil moisture dynamics, such as vertical heterogeneity and inter-layer connectivity. Specifically, we discard all assumptions on soil, root, or boundary conditions and instead attempt to learn the soil water dynamics directly from the data. Unlike traditional one-dimensional soil water flow models that often focus on adjacent-layer fluxes, our model captures complex vertical dependencies and non-uniform moisture redistribution across various depths, enhancing predictions in complex scenarios. We introduce the Self-Attention mechanism Non-local Neural Networks (SA-NLNN) to explore the potential of NLNN structures in soil moisture forecasting. Moreover, the Knowledge-Guided Non-local Neural Network (KG-NLNN) that incorporates soil water transport guidance into the non-local operation is proposed. We examine the models' interpretability using the synthetic data, while in-situ data is applied to assess the practicality and accuracy of the models. The key innovations of our study are as follows: First, unlike previous machine learning models that rely on time-series processing to capture temporal patterns, our study is designed based on a physically motivated assumption: the soil moisture profile at the current day, together with meteorological forcing, contains sufficient information to predict the soil moisture state of the following day. Therefore, the prediction task is formulated as a single-time-step problem involving multi-depth variables. This allows mutual compensation within the soil profile, enabling effective and precise soil moisture forecasts. The adaptability of NLNNs across various temporal and spatial scales is also demonstrated. Second, the learned non-local weights of the NLNN model can be visualized to provide qualitative information on soil properties inferred from soil moisture data. Each weight represents the relative influence of soil moisture at one depth on the moisture state at another depth in the subsequent time step, thereby reflecting vertical soil water interactions. The model interpretability is investigated using synthetic soil moisture data, including virtual examples of homogeneous soil, heterogeneous soil, two-layered soil, and soil with root water uptake. Third, incorporating knowledge-inspired concepts enhances model accuracy and reliability. When evaluating practical performance, we utilize in-situ soil moisture data sourced from the International Soil Moisture Network (ISMN) and compare our models with the benchmark LSTM model (Datta & Faroughi, 2023; Semwal et al., 2021; Y. Wang et al., 2023). To the best of our knowledge, this marks the first instance of employing NLNNs for interpretable soil moisture dynamics forecasting.

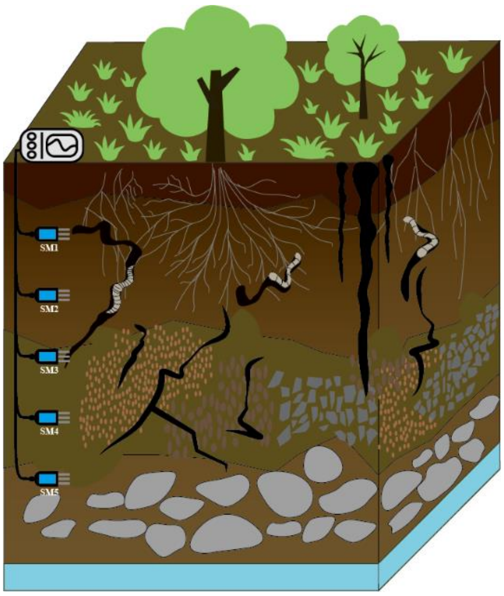

Figure 1. Examples of complex soil conditions related to soil texture and soil structure at the soil profile scale. SM3 is more related to SM1 other than SM2 or SM4, due to the existence of wormholes. The proposed non-local neural network is designed to understand that SM3 is highly correlated with SM1 (caused by fast water migration in wormholes) and less correlated with SM2 (caused by slow seepage under gravity).

The remainder of this study is organized as follows: Section 2 presents the NLNNs for soil moisture forecasting, including the SA-NLNN and KG-NLNN; Section 3 describes the synthetically generated soil moisture data and the in-situ data; Section 4 provides the model results and the interpretability analysis. Finally, the conclusion is drawn in Section 5.

2. Methodologies

2.1 Physical Background

The dynamics of soil moisture transport are fundamentally described by the Richards equation, a governing relation derived from the mass conservation law and the Buckingham-Darcy law (Buckingham, 1907). For one-dimensional uniform flow in homogeneous soil, and assuming the absence of preferential flow, this equation takes the following form:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial \mathrm{z}}\left[K\left(\frac{\partial \psi}{\partial z} + 1\right)\right] \tag{1}$$

where $\theta\ [cm^3 cm^{-3}]$ is the volumetric moisture content, $t\ [day]$ denotes the time, $z\ [cm]$ is the vertical coordinate (positive upward), $K\ [cm/day]$ is the unsaturated hydraulic conductivity, $\psi\ [cm]$ is the soil matric potential of water.

Based on this equation, the soil moisture profile at a subsequent time step evolves from the preceding profile. Infiltration and evaporation, driven by meteorological factors, directly influence surface soil moisture, which triggers a redistribution of moisture through the soil profile. Therefore, the multi-depth soil moisture at the next time step can be determined by both the current meteorological conditions and the soil moisture profile from the previous time step.

2.2 Model structures

According to section 2.1, we assume that the soil moisture within the profile at the next time step depends on both the current meteorological conditions and the soil moisture from the previous time step in our soil moisture forecasts at multiple depths. The NLNN models are designed to capture the potential interactions of soil moisture at different depths within the vertical profile (Figure 1), thereby making predictions that are closer to reality. Figure 2 illustrates the NLNN structure proposed for soil moisture dynamics prediction. The input data for the NLNN model, denoted as $X^t = [x_0^t, x_1^t, x_2^t, \dots, x_{n-1}^t, x_n^t]$, comprises a concatenation of soil moisture truth at n depths from the previous time step $\boldsymbol{sm}^t = [sm_1^t, sm_2^t, \dots, sm_n^t]^T$ and the upper boundary factor x_0^t obtained from meteorological conditions processing through an LSTM. Here, sm_n^t denotes the soil moisture at depth n and time t. The initial soil moisture content for the prediction is set to the truth from the preceding day. Specifically, this value is obtained from the physical model's output for the virtual scenario and from field observations for the real-world scenario.

Within our framework, we employ two types of non-local operations. The first, SA-NLNN, utilizes embedded Gaussian functions; it represents a novel application of the self-attention mechanism to capture vertical dependencies in soil moisture. The second model, KG-NLNN, is a newly proposed architecture where the non-local operation is decoupled based on the soil water transport mechanisms. In the NLNN structure, following the non-local operation and a residual connection, a fully connected neural network is employed to generate predictions for the soil moisture at each corresponding depth. This yields prediction denoted as, $\boldsymbol{sm}^{t+1'} = \left[sm_1^{t+1'}, sm_2^{t+1'}, \ldots, sm_{n-1}^{t+1'}, sm_n^{t+1'}\right]^{\mathrm{T}}$. The ground truth is represented as $\boldsymbol{sm}^{t+1} = [sm_1^{t+1}, sm_2^{t+1}, \ldots, sm_{n-1}^{t+1}, sm_n^{t+1}]^{\mathrm{T}}$.The model is trained by minimizing the error between predictions and the ground truth,

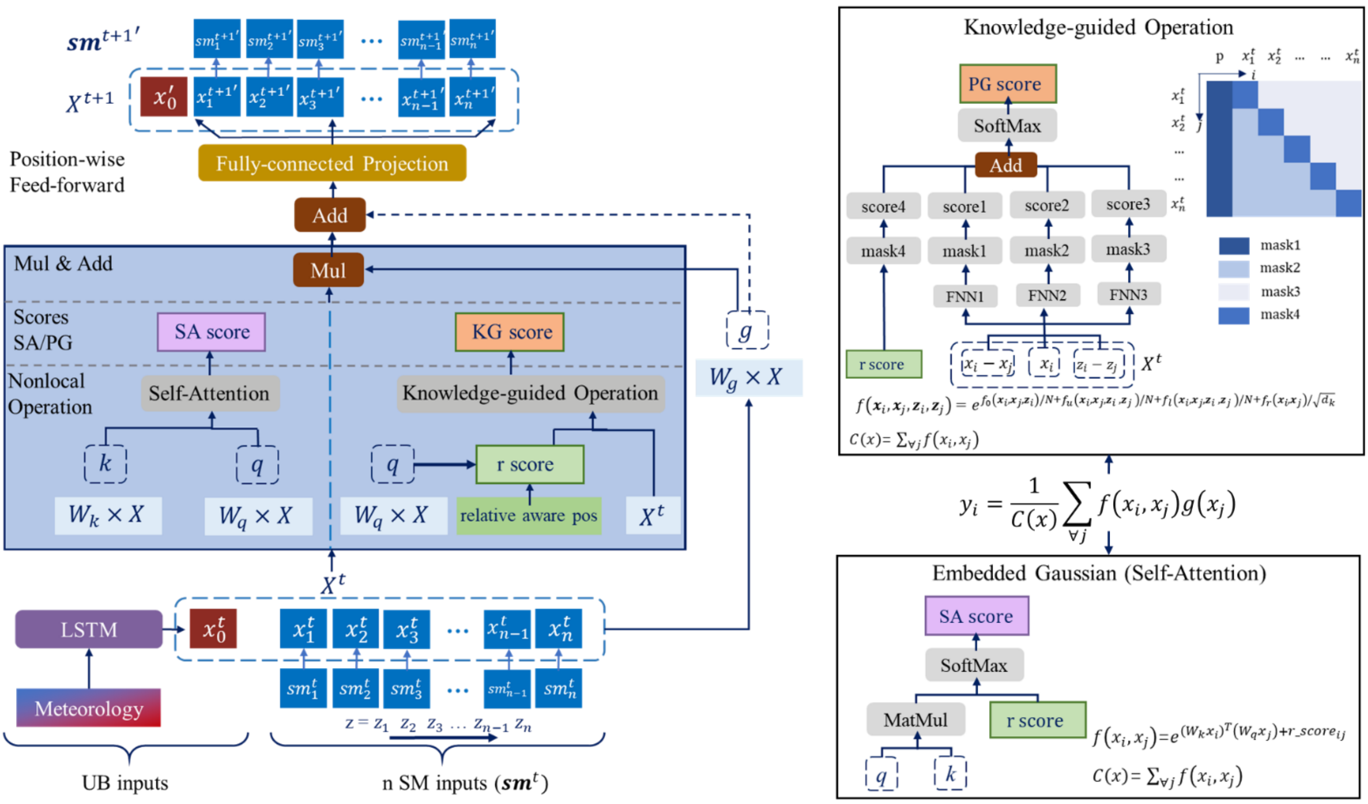

Figure 2. Left: non-local neural network structure for soil moisture forecasting. Right: embedded Gaussian operation and knowledge-guided non-local operation. RPE: relative position encoding. SA/KG score: non-local weights computed through embedded Gaussian operation and knowledge-guided operation. W_q, W_k and W_g are the weight matrixes to be learned for embeddings.

2.3 Non-local Operations

The general form of a non-local operation in NLNNs can be defined as follows (X. Wang et al., 2018):

$$\boldsymbol{y}_i = \frac{1}{\mathcal{C}(\boldsymbol{x})}\sum_{\forall j} f(\boldsymbol{x}_i, \boldsymbol{x}_j) g(\boldsymbol{x}_j) \tag{2}$$

Here i denotes the index of the output $\boldsymbol{y}$ for which the output value is being calculated, while j is the index that lists all conceivable positions in the input $\boldsymbol{x}$. The term $\boldsymbol{y}_i$ denotes the i_{th} component of the output $\boldsymbol{y}$. In this context, $\boldsymbol{x}$ represents the input data and $\boldsymbol{y}$ denotes the corresponding output, both sharing the same dimensionality. In this work, $\boldsymbol{x}$ represents the concatenation of input soil moisture data and upper boundary condition data, denoted as $\boldsymbol{sm}^t$. Accordingly, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ denote the soil moisture at the i_{th} and j_{th} depths at time step t, sm_i^t and sm_j^t. The output $\boldsymbol{y}$ corresponds to the predicted soil moisture at the next time step, denoted as $\boldsymbol{sm}^{t+1'}$, where $\boldsymbol{y}_i$ represents the predicted soil moisture content at the i_{th} depth, $sm_i^{t+1'}$. The computation of a generic non-local operation involves three components: the pairwise function f, the unary function g, and the normalization sum $\mathcal{C}(\boldsymbol{x})$. The function f calculates a scalar (representing relationship such as affinity) between i and all j, while the unary function g generates a representation of the input at position j. The aggregated response is then normalized by $\mathcal{C}(\boldsymbol{x})$. In this study, the form of g is restricted to a linear embedding: $g(\boldsymbol{x}_j) = W_g \boldsymbol{x}_j$, where W_g is a learnable weight matrix. The primary modification focuses on the pairwise function f. The $\mathcal{C}(\boldsymbol{x})$ is contingent on the design of f. Following the definition of attention heads from previous work on self-attention mechanisms (Vaswani et al., 2017), our NLNN models employ several operation heads to enhance the model's feature extraction and representation capabilities. The number of operation heads is denoted as n_{head}. Similar non-local operations are performed in each head, with some parameter matrices being unique. To form the output, results from each head are concatenated, and a parameterized linear transformation is applied.

The non-local operations offer flexibility by assuming various forms and can adapt to specific problem designs. This provides potential solutions for many complex situations. This flexibility stems from their ability to model global dependencies through data-dependent pairwise interactions. Among these formulations, the Transformer represents the most typical and widely used architectural instantiation, which models global dependencies through the query–key–value self-attention mechanism, multi-head attention, positional encoding, and feed-forward layers. From a more general perspective, the Non-local Neural Network can be viewed as a broader formulation of non-local dependency modeling, which computes interactions based on pairwise affinity functions without requiring the full Transformer architecture. In the following sections, we will introduce the classical embedded Gaussian operation, along with our knowledge-guided non-local operation designed for soil moisture dynamics.

2.3.1 Embedded Gaussian Operation:

Self-attention, a specific case of non-local operations within the embedded Gaussian version, is a key component of the Transformer architecture. It excels in processing data concisely and capturing intricate relationships, making it widely applied in various research areas (Devlin et al., 2019; Lim et al., 2021; Z. Liu et al., 2021). However, it overlooks the ordering of input, necessitating the incorporation of position information into the calculations to ensure accurate processing.

Common position encoding methods include absolute position encoding (Devlin et al., 2019; Gehring et al., 2017; Vaswani et al., 2017) and relative position encoding (Shaw et al., 2018). Absolute position encoding directly incorporates absolute position information pertaining to i or j and integrates it into the input. In contrast, relative position encoding focuses on the relative relationship between position i and j. Given the complexity of soil properties and the nature of soil moisture interactions, prioritizing the relative influence of soil moisture at each depth may prove more effective than relying on absolute position information in soil moisture analysis. In this approach, we utilize the relative position encoding similar to the method proposed by Shaw et al. (2018). The function f encompasses a Gaussian function of two embeddings along with the relative position representation associated with i and j. A self-attention mechanism with relative position encodings in each head can be defined as follows:

$$f(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{((W_k x_j)^{\mathrm{T}}(W_q x_i) + r\_score_{ij})/\sqrt{d_k}} \tag{3}$$

$$\mathcal{C}(\boldsymbol{x}) = \sum_{\forall j} f(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{4}$$

Here, W_q and W_k are the weight matrixes to be learned for embeddings. $\sqrt{d_k}$ denotes the scale factor, where d_k represents the dimension of the embeddings. r_score_{ij} is the relative position score computed using relative position encoding. Then the $\boldsymbol{y}_i$ can be calculated through Equation (1). The embedded Gaussian operation for soil moisture forecasts is illustrated in Figure 2.

In the relative position encoding, each relationship between two arbitrary positions i and j is represented by a learnable vector. Here, r_score_{ij} denotes an internal relative position score used in the non-local operation, rather than a model evaluation metric. Then, the r_score_{ij} is calculated as follows:

$$r\_score_{ij} = \left(\boldsymbol{a}_{i,j}\right)^T \left(W_q \boldsymbol{x}_i\right) \tag{5}$$

where $\boldsymbol{a}_{i,j}$ represents the relative position encoding utilized for r_score_{ij} computing. $\boldsymbol{a}_{i,j}$ is a parameter vector that needs to be trained. In the proposed SA-NLNN model, our trainable relative position encoding matrix A consists of $(n+1) \times (n+1)$ distinct elements. The matrix A needs to be learned through training:

$$A = \begin{pmatrix} \boldsymbol{a}_{0,0} & \cdots & \boldsymbol{a}_{0,n} \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_{n,0} & \cdots & \boldsymbol{a}_{n,n} \end{pmatrix} \tag{6}$$

In this model, all operation heads perform similar operations. W_q, W_k, and W_g are unique in each head. However, the relative position encoding can be shared across non-local operation heads.

2.3.2 Disentangled Knowledge-Guided operation:

In this work, we propose KG-NLNN, a model specifically designed for forecasting soil moisture at multiple depths in the soil profile, as depicted in Figure 2. The vertical movement of soil moisture exhibits a directional divergence: downward flow is driven primarily by gravity and constitutes a dissipation of potential energy, while upward movement is governed by capillary forces and other mechanisms acting against gravity. In this specific context, we employ a set of masks to decouple soil moisture interactions from different directions. The four masks in Figure 2 correspond to four key components: meteorological forcing, upper soil water influence, same-depth soil moisture effects, and lower soil water interactions, respectively. Meteorological forcing, upper soil water influence, and lower soil water interactions are modeled by fully connected networks with soil moisture content and depth differences as inputs, whereas same-depth soil moisture effects are represented via relative position encoding. This knowledge-guided architecture separates different moisture movement processes for independent learning, thereby enhancing the model's ability to capture complex relationships among soil moisture variables across the soil profile.

When analyzing the soil moisture at i_{th} depth, denoted as $\boldsymbol{y}_i$, its dynamics are influenced by several factors: upper boundary conditions represented by $\boldsymbol{x}_0$, upper soil moisture state at the previous time step,

$\boldsymbol{x}_u$, (where $u < i$, primarily donated by gravity), lower soil moisture $\boldsymbol{x}_l$, (where $l < i$, mainly affected by capillary), and the soil moisture at the same depth from the previous time step, $\boldsymbol{x}_i$. Since these four components are motivated by diverse physical mechanisms, they are defined in distinct forms within the non-local operation.

Before proceeding to the subsections, we provide a brief introduction to fully-connected neural networks (FNNs) that are utilized in the following sections. A two-layer fully-connected neural network can be defined as follows:

$$FNN(\boldsymbol{x}_{input}) = a_t(W_2(a_t(W_1\boldsymbol{x}_{input} + b_1) + b_2) \tag{7}$$

where a_t denotes the tanh activation function, and W_L and b_L represent the weight matrices and bias parameters to be learned in the L_{th} layer, respectively, where $L = 1,2$. $\boldsymbol{x}_{input}$ denotes the input vector of an FNN. According to the universal approximation theorem (Cybenko, 1989), a feedforward neural network with a single hidden layer is theoretically sufficient to approximate a wide range of nonlinear functions. In this study, a two-layer FNN is adopted to balance model expressiveness and computational efficiency. The hyperbolic tangent function is adopted as the activation function a.

The effect of upper boundary conditions on soil moisture at depth $\boldsymbol{z}_i$ is described by the function, $f_0(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{z}_i)$, which corresponds to three factors: $\boldsymbol{x}_0$, the meteorological factor; $\boldsymbol{x}_i$, the soil moisture at depth $\boldsymbol{z}_i$ from the previous time step; and $\boldsymbol{z}_i$, the depth of the concerned soil moisture. $\boldsymbol{z}_i$ denotes the i_{th} depth in the depth vector $\boldsymbol{z} = [z_0, z_1, \dots, z_n]^{\mathrm{T}}$, which corresponds to the input soil moisture data $\boldsymbol{sm}^t$. We utilize a two-layer FNN to describe this relationship:

$$f_0(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{z}_i) = FNN_0(\boldsymbol{x}_0, \boldsymbol{x}_i, \boldsymbol{z}_i), j = 0 \tag{8}$$

In considering the impacts of soil moisture in the upper layers and lower layers on soil moisture at depth $\boldsymbol{z}_i$, we propose $f_u(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{z}_i, \boldsymbol{z}_j)$ and $f_l(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{z}_i, \boldsymbol{z}_j)$ to calculate the effects. Both functions are determined by the disparity in soil moisture content $(\boldsymbol{x}_i - \boldsymbol{x}_j)$, the intrinsic soil moisture $\boldsymbol{x}_i$, and the distance between two positions $(\boldsymbol{z}_i - \boldsymbol{z}_j)$. As previously stated, two two-layer FNNs are employed in this section:

$$f_u(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{z}_i, \boldsymbol{z}_j) = FNN_u(\boldsymbol{x}_i - \boldsymbol{x}_j, \boldsymbol{x}_i, \boldsymbol{z}_i - \boldsymbol{z}_j), i > j \tag{9}$$

$$f_l(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{z}_i, \boldsymbol{z}_j) = FNN_l(\boldsymbol{x}_i - \boldsymbol{x}_j, \boldsymbol{x}_i, \boldsymbol{z}_i - \boldsymbol{z}_j), i < j \tag{10}$$

Additionally, we utilize relative position encodings to describe the soil water retention effect:

$$f_r(\boldsymbol{x}_i, \boldsymbol{x}_j) = r\_score_{ij}, i = j \tag{11}$$

where the relative position score r_score_{ij} is utilized for the water retention effect of soil moisture at a specific depth across two adjacent time steps. It can be calculated in Equation (4). Consequently, our position encoding matrix A_{PG}^{K} is a diagonal matrix comprising $(n + 1)$ distinct elements, which needs to be learned through training:

$$A_{PG} = \begin{pmatrix} \boldsymbol{a}_{0,0} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{a}_{n,n} \end{pmatrix} \tag{12}$$

According to the above, the impact on soil moisture at a fixed depth is harmoniously coordinated and integrated through the four components mentioned earlier, as illustrated in Figure 2. Therefore, the knowledge-guided non-local operation for soil moisture dynamics simulation can be defined as follows:

$$f(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{z}_i, \boldsymbol{z}_j) = e^{f_0(x_i, x_j, z_i)/N + f_u(x_i, x_j, z_i, z_j)/N + f_l(x_i, x_j, z_i, z_j)/N + f_r(x_i, x_j)/\sqrt{d_k}} \tag{13}$$

$$\mathcal{C}(\boldsymbol{x}) = \sum_{\forall j} f(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{z}_i, \boldsymbol{z}_j) \tag{14}$$

where N is the number of positions in $\boldsymbol{x}$, $\sqrt{d_k}$ denotes the scale factor. Then $\boldsymbol{y}_i$ can be calculated using Equation (1). All operation heads execute similar operations in this model. W_q utilized for r_score computing and W_g in $g(\boldsymbol{x}_j)$ are still unique in each head. The parameters of the FNNs are shared across non-local operation heads.

2.4 Boundary processing

In our soil moisture prediction task, the impact of the upper boundary conditions on soil moisture is partially simulated by an LSTM module (Hochreiter & Schmidhuber, 1997), as illustrated in Figure 2. We have selected six meteorological variables to characterize the influence of these upper boundary conditions: precipitation (P), air temperature (AT), long-wave radiation (LR), short-wave radiation (SR), relative humidity (RH), and wind speed (WS). These variables, denoted as $\boldsymbol{ub}^t = [P^t, AT^t, LR^t, SR^t, RH^t, WS^t]^{\mathrm{T}}$, are closely associated with the infiltration and evapotranspiration processes. Hydrologically, meteorological conditions from the previous time step (t–1) do not cease their influence immediately; rather, processes such as infiltration, lateral flow, and redistribution allow these conditions to continue affecting soil moisture at the subsequent time step t. Incorporating both time steps thus enables the model to capture cross-day causal relationships. A time step of 2 is used to keep the meteorological inputs concise while retaining adequate informational richness. Accordingly, the task of learning meteorological temporal dependencies is assigned to the LSTM network, which also justifies its use in processing boundary conditions. Following LSTM processing, the impact of the upper boundary conditions takes the form of sm_0^t, which is subsequently utilized in non-local operations in conjunction with the input soil moisture data $[sm_1^t, sm_2^t, \dots, sm_{n-1}^t, sm_n^t]^{\mathrm{T}}$ within the soil profile. The operation of an LSTM can be summarized as follows:

$$\boldsymbol{i}^t = a_s(W_i \cdot [\boldsymbol{h}^{t-1}, \boldsymbol{ub}^t] + b_i) \tag{15}$$

$$\boldsymbol{f}^t = a_s\left(W_f \cdot [\boldsymbol{h}^{t-1}, \boldsymbol{ub}^t] + b_f\right) \tag{16}$$

$$\boldsymbol{o}^t = a_s(W_o \cdot [\boldsymbol{h}^{t-1}, \boldsymbol{ub}^t] + b_o) \tag{17}$$

$$\widetilde{\boldsymbol{C}}^t = a_t(W_c \cdot [\boldsymbol{h}^{t-1}, \boldsymbol{ub}^t] + b_c) \tag{18}$$

$$\boldsymbol{c}^t = \boldsymbol{f}^t \cdot \boldsymbol{c}^{t-1} + \boldsymbol{i}^t \cdot \widetilde{\boldsymbol{C}}^t \tag{19}$$

$$\boldsymbol{h}^t = \boldsymbol{o}^t \cdot a_t(\boldsymbol{c}^t) \tag{20}$$

where W_i and b_i, W_f and b_f, W_o and b_o denote the deep learning parameters for the input gate, forget gate, and the output gate, respectively; W_c and b_c are the parameters for cell state updating; in addition, $\boldsymbol{i}^t$, $\boldsymbol{f}^t$ and $\boldsymbol{o}^t$ are the input gate, forget gate, and output gate at time t, respectively, and $\boldsymbol{c}^t$ is the memory cell state; $\boldsymbol{h}^t$ represents the hidden state; a_s is the sigmoid activation function, and a_t denotes the tanh activation function.

Through sequential processing, the last hidden state $\boldsymbol{h}^t$ in the output $[\boldsymbol{h}^{t-1}, \boldsymbol{h}^t]$ derived from input $[\boldsymbol{ub}^{t-1}, \boldsymbol{ub}^t]$, which encodes the upper boundary effect over two time steps, is adopted as the sm_0^t. In this study, the lower boundary conditions are disregarded due to the obstacles in observation.

2.5 Training Strategies

The objective of our model is to simultaneously predict soil moisture at multiple depths for the next time step. To achieve this, we define the loss function as the sum of squared errors between the model predictions and the corresponding ground truth of soil moisture content at different depths. The model is trained by minimizing this loss function:

$$\mathcal{L} = \sum_{t=0}^{B} \sum_{i=1}^{n} \left(sm_i^{t+1'} - sm_i^{t+1} \right)^2 \tag{21}$$

where n denotes the number of concerned soil moisture depths, and B is the training batch size, which is set to 100 in this study.

In this work, the collected data is divided into training, validation, and test sets in a time-ordered ratio of 6:2:2. For training, we employ the Adam optimizer (Kingma & Ba, 2015) with a learning rate of 0.001. The models are trained for a minimum of 2500 epochs, with 20 batches in each epoch. The validation set is utilized to select the best model and mitigate overfitting. Subsequently, the test set is then employed to evaluate the performance of the models. Each result is computed based on 10 replicates with different initializations. Regarding the model hyperparameter settings, in the non-local neural network, we set $d_k = d_q = 10, d_g = 16,$ and $n_{head} = 10$, where d_k, d_q and d_g represents the dimensions of the key, query and value (function g) components within the non-local block, respectively. n_{head} denotes the number of non-local heads. The LSTM consists of two stacked blocks, each configured with a hidden layer of 20 neurons. In the FNN adopted for KG-NLNN, we utilize 10 neurons in each hidden layer.

3. Data Descriptions

In our study, synthetic soil moisture data is generated to investigate the interpretability of these NLNN models. Additionally, we utilize the selected in-situ soil moisture data to assess the accuracy and practicability of our models.

3.1 Synthetic Data Description

The synthetic data are generated using the ROSS method (P J Ross, 2003; Peter J Ross, 2006). The Ross method is a rapid, non-iterative numerical scheme for soil moisture forward modeling. In our simulation, we create soil moisture content data for a 100 cm soil column with 1 cm intervals. For boundary conditions, the daily reference evapotranspiration (ET0) is calculated with the FAO Penman-Monteith method (Allen et al., 1998) in Wuhan coordinates to generate the synthetic data. As standardized in the FAO guidelines (Allen et al., 1998), actual evapotranspiration is the product of K_C and ET0, where K_C serves as a refined empirical parameter. When generating synthetic data, we applied this empirical coefficient method to derive a preliminary evapotranspiration estimate, adopting a coefficient value of 1.0 in this instance. The daily time series data of precipitation and calculated evapotranspiration are shown in Figure 3. The lower boundary condition is set as free drainage, and the initial moisture content of the soil column is set to a uniform value of 0.10. We generate three years of time series soil moisture data for this research.

In this section, we design four virtual cases of different configurations to investigate model interpretability, including homogeneous soil, heterogeneous soil, two-layered soil, and soil with root water uptake scenarios, as represented in Figure 4. When generating synthetic data in the case with root water uptake, the root depth is set to 50cm, and root density is vertically distributed evenly. Detailed soil property settings are given in Appendix A. Besides, we assess the adaptability across different time scales and observation locations using the available data.

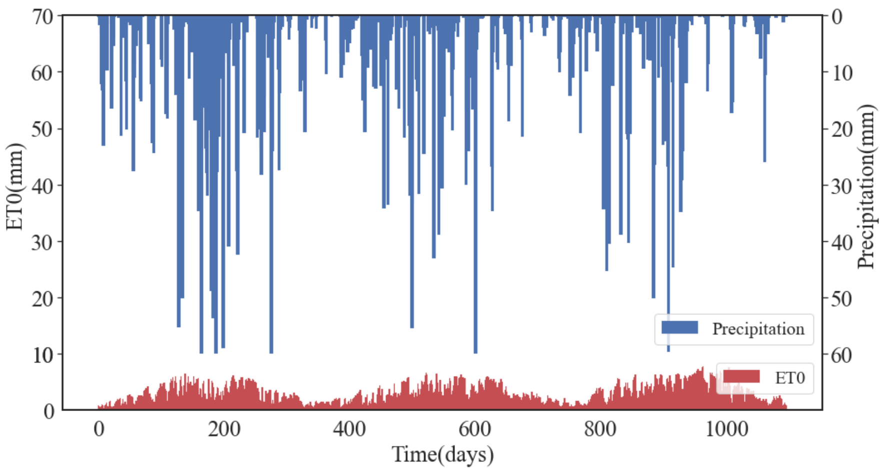

Figure 3. Daily time series precipitation and reference evapotranspiration data calculated at Wuhan coordinate for generating synthetic data.

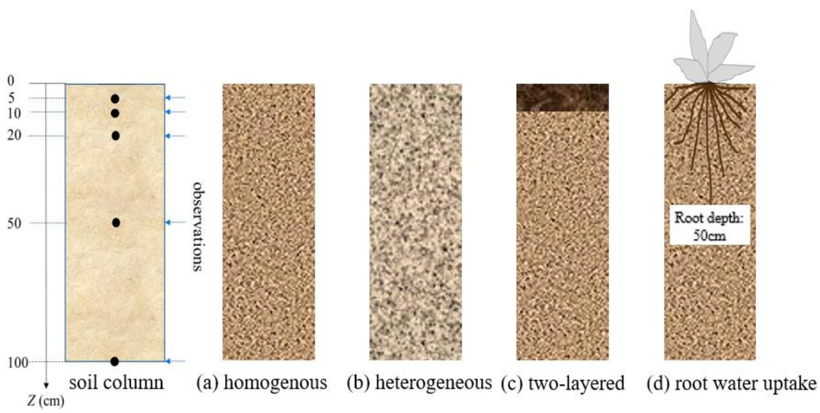

Figure 4. The virtual cases design, with homogeneous soil (a), heterogeneous soil (b), two-layered soil (c), and homogeneous soil with root water uptake (d).

3.2 In-situ Data Description

To comprehensively evaluate the proposed NLNN models, we carefully select soil moisture content observations from twenty sites within the International Soil Moisture Network (ISMN) (https://ismn.geo.tuwien.ac.at/en/). These sites are chosen based on geographical locations, soil textures, and land cover types. Detailed information for the selected sites is presented in Table 1, and their spatial locations are illustrated in Figure 5. These carefully selected sites encompass 16 soil types and 6 land cover species, providing a diverse range to assess the model's performance and its ability to adapt to complex soil situations. At each site, in-situ observations are required to include soil moisture observations at 5 standard depths (0.05m, 0.10m, 0.20m, 0.50m, 1.00m).

Table 1. Summary of main characteristics of twenty selected sites.

Number	Site	Sand	Silt	Clay	Land cover	Period	Lat.	Lon.
1	Kingston-1-W	85	10	5	Grassland	2012-2023	41.48	-71.54
2	Monahans-6-ENE	83	6	11	Shrub cover	2010-2022	31.62	102.81
3	Necedah-5-WNW	83	11	6	Grassland	2009-2022	44.06	-90.17
4	Shadow Mtns	79	10	11	Shrub cover	2013-2017	35.47	-115.72
5	Falkenberg	73	21	6	Cropland, rained	2003-2020	52.17	14.12
6	Kenai-29-ENE	54	38	8	Shrub cover	2012-2023	60.72	-150.45
7	AAMU-jtg	53	22	25	Grassland	2010-2022	34.78	-86.55
8	Darrington-21-NNE	53	22	25	Tree cover	2013-2019	48.54	-121.45
9	Palestine-6-WNW	49	27	24	Grassland	2009-2013	31.78	-95.72
10	Cullman	49	27	24	Mosaic Cropland	2006-2022	34.20	-86.80
11	Cape-Charles	49	27	24	Herbaceous cover	2011-2022	37.29	-75.93
12	LittleRiver	47	30	23	Grassland	2005-2020	31.50	-83.55
13	Montrose-11-ENE	43	35	22	Tree cover	2010-2023	38.54	-107.69
14	Coshocton-8-NNE	41	39	20	Grassland	2009-2016	40.37	-81.78
15	Bodega-6-WSW	39	38	23	Grassland	2011-2023	38.32	-123.08
16	Goodwell-2-SE	36	41	23	Grassland	2010-2022	36.57	-101.61
17	Riley-10-WSW	36	41	23	Shrub cover	2011-2021	43.47	-119.69
18	Joplin-24-N	35	41	24	Grassland	2010-2020	37.43	-94.58
19	Weslaco	34	45	21	Cropland, rained	2017-2021	26.16	-97.96
20	UpperBethlehem	32	38	30	Herbaceous cover	2008-2010	17.72	-64.80

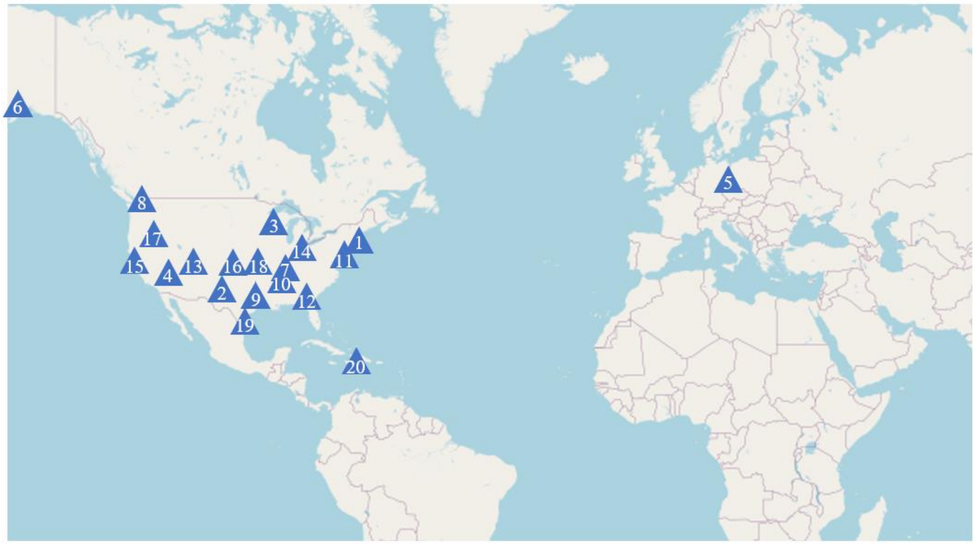

Figure 5. The spatial locations of twenty selected sites. The numbers on the sites correspond to the serial numbers in Table 1.

The meteorological inputs for our models include precipitation, atmospheric temperature, long-wave radiation, short-wave radiation, wind speed, and relative humidity, as mentioned above. These meteorological data are sourced from the NASA Prediction of Worldwide Energy Resources project (https://power.larc.nasa.gov/). Based on the latitude and longitude coordinates of each station, we downloaded the corresponding point-scale, daily-resolution meteorological datasets. Detailed information about this can be found at (https://power.larc.nasa.gov/docs/methodology/data/sources/). Unfortunately, due to challenges in obtaining groundwater level observations, changes in the lower boundary conditions are not considered in this study.

4. Results and discussions

In this study, we systematically examine and analyze our models from three perspectives. Initially, we assess the essential capabilities of models, including accuracy and uncertainty, using both synthetic data and in-situ observations. Subsequently, we apply simulated soil moisture data under diverse virtual scenarios to evaluate our model's interpretability and its ability to provide qualitative interpretations depicting soil moisture interaction mechanisms across diverse depths within the profile. Finally, we investigate the impacts of varying temporal scales, noise levels, and observation locations on our non-local neural networks.

To explore the forecasting ability of our models over time series, we examine predictions for 1, 3, and 7 days ahead at selected sites, as well as 1, 3, 7, and 15 days ahead for simulated data. We generate predictions iteratively. The evaluation standards in this work comprise the mean absolute error (MAE) and the root mean square error (RMSE). Both MAE and RMSE quantify the deviation between the predictions and the ground truth. However, RMSE exhibits greater sensitivity to outliers due to its squaring of deviations, which amplifies the impact of extreme values, while MAE offers a smoother average error value. These metrics are calculated as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N_s} \left| T_i - \hat{T}_i \right|}{N_s} \tag{22}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_s} \left(T_i - \hat{T}_i \right)^2}{N_s}} \tag{23}$$

where $\hat{T}_i$ and T_i represent the predictions and the ground truth, respectively; $\overline{T}_i$ is the average of the ground truth; N_s is the test sample size. Here, T denotes the soil moisture content [%] which needs to be calculated. All the compared models are trained and evaluated using the same datasets, input variables, and evaluation metrics to further ensure consistency and fairness in the comparison.

When conducting uncertainty analysis, evaluating confidence bounds becomes challenging because most deep learning neural networks are essentially deterministic models. To address this, many researchers utilize the bootstrap aggregating (bagging) method (Breiman, 1996) to analyze model predictive uncertainty (Kornelsen & Coulibaly, 2014). The bagging method involves training multiple neural network models using subsets of the training set, all with identical architecture. To create the training subset for each model, a statistical bootstrap approach is employed. For each subset, we randomly select individual input vectors from the entire training set with replacement, ensuring that each subset contains the same number of elements as the entire training set. After training, we obtain an ensemble of trained models, each trained with a unique training subset. The final output and uncertainty estimates are then derived from the mean and standard deviation of this ensemble.

To explore the impact of noise on our models using the synthetic data, we apply the zero-mean Gaussian noise with a variance of 1:

$$\dot{\theta} = \theta + \eta * \mathcal{N}(0,1), \tag{24}$$

where $\dot{\theta}$ is the volumetric soil moisture content with noise [%], and θ is the synthetic volumetric soil moisture content. Three noise levels are tested (η = 0.5, 1.0, 2.0) in this work.

In our investigation of model interpretability, the visualized non-local weight maps generated from the output play a crucial role as evaluation standards. According to Equation (2), the normalized weights $\frac{1}{\mathcal{C}(x)} f(\boldsymbol{x}_i, \boldsymbol{x}_j)$ quantify the relative influence of soil moisture at depth j on the prediction at depth i. These normalized interaction weights reflect how strongly soil moisture information from different depths on the previous day contributes to the predicted soil moisture at a given depth on the following day. These weight maps may provide qualitative interpretations depicting intricate mechanisms of soil water dynamics. The color brightness on the weight distribution map signifies the level of interaction strength among upper boundary conditions and soil moisture across different depths. Therefore, analyzing the weight matrix map is essential for gaining insights into the learning mechanisms of our NLNN models.

4.1 Interpretability analysis

Before the models can be applied to real-world scenarios, their stability and interpretability must first be analyzed. In this section, we explore the interpretability of the NLNN models by designing several scenarios that generate synthetic data. These simulated cases primarily involve variations in soil properties, including homogeneous soil, heterogeneous soil, two-layered soil, and soil with root water uptake scenarios. We benchmark the soil moisture prediction tasks against the LSTM model, widely used in time series forecasting (Datta & Faroughi, 2023; Ding et al., 2019; Siami-Namini et al., 2019). Specifically, the LSTM model takes two forms tailored for different data processing approaches: LSTM_T, which utilizes input data from the previous four time steps to predict soil moisture content at the next time step. It follows a configuration similar to that in previous work (Y. Wang et al., 2023). These predictions rely on modeling temporal dependencies. In contrast, LSTM_I replaces the non-local operations in the architecture shown in Figure 2 with LSTM modules, thereby modeling interactions among soil water layers. It represents the predictive capabilities achievable by a single-time-step LSTM.

With the synthetic data, we investigate the model performance and interpretability through the weight matrix maps and delve into their learning mechanisms across diverse scenarios.

Figure 6 displays the RMSE results for 1, 3, 7, and 15-day forecasts of four models, and the MAE values of four simulated scenarios are summarized in Appendix C. As shown in Figure 6, the LSTM_T model achieves very high accuracy in 1-day predictions, but its performance deteriorates rapidly over longer periods. As for the other models, NLNNs and LSTM_I exhibit comparable performance. The knowledge-guided model KG-NLNN exhibits lower variance and maintains greater stability in RMSE, especially in the 15-day prediction task. The integration of knowledge guidance proves crucial in ensuring model stability.

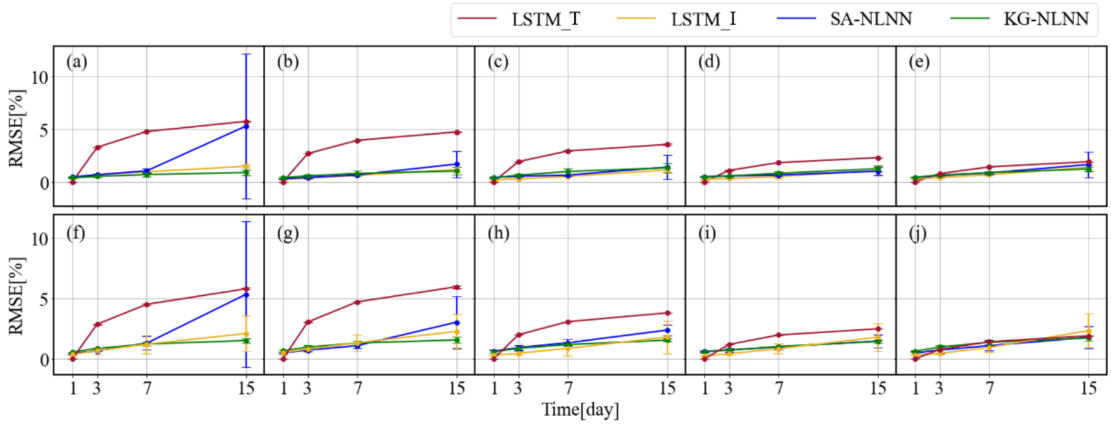

Figure 6. The RMSE results for 1, 3, 7, and 15-day for heterogeneous soil(a-e), and two-layered soil (f-j). The error bar indicates the standard deviations of the RMSE, which are computed via ten training replicates.

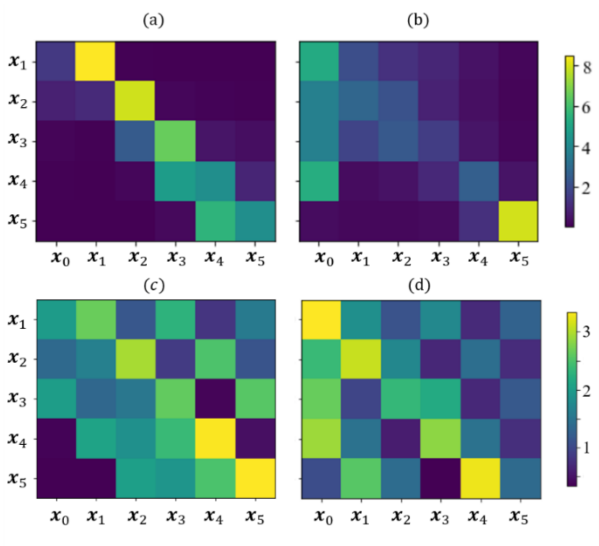

Figure 7. The non-local weight maps in homogeneous simulated soil scenarios through KG-NLNN (a) Ks = 0.25 (b) Ks = 10.49, and SA-NLNN (c) Ks = 0.25 (d) Ks = 10.49

Figure 7 depicts the weight matrix maps generated by KG-NLNN and SA-NLNN models for homogeneous soil scenarios varying saturated hydraulic conductivity (Ks) values. These maps represent the term $\frac{f(x_i,x_j)}{C(x)}$ calculated through non-local operations. Each element at position (i,j) represents the impact of soil moisture at depth $\boldsymbol{z}_j$ at the previous time on the soil moisture content at depth $\boldsymbol{z}_i$. Notably, when $j=0$, it signifies the influence of upper boundary conditions on soil moisture across various depths. The brightness level corresponds to the strength of this influence, with higher brightness indicating a stronger impact. Homogeneous soil scenarios with different Ks values are used to examine variations in the non-local weight matrices. The weight maps produced by the KG-NLNN model exhibit clear and stable spatial patterns across different Ks, whereas the SA-NLNN results appear relatively chaotic, indicating that a knowledge-guided structural design can serve as a valuable enhancement.

Differences in hydraulic conductivity govern soil water flow velocity, leading to variations in the time required for water to reach different depths. These differences shape the structure of the weight maps and give rise to the distinct patterns observed in Figures 7(a) and 7(b). For instance, loam (Ks = 0.25) exhibits slow infiltration, so its moisture content is easily influenced by adjacent layers in Figure 7(a). In contrast, sand (Ks=10.49) allows rapid infiltration, resulting in deeper soil moisture being affected directly by meteorological factors. Although the proposed model does not involve any parameterization nor perform a quantitative description of soil hydraulic parameters, it nevertheless provides insights into these hydraulic properties to some extent.

Additionally, two-layer soil scenarios are employed in which the soil properties of the upper and lower layers are exchanged to further investigate changes in the non-local weight matrices. Figure 8 depicts the weight matrix maps generated by KG-NLNN and SA-NLNN models for two-layered soil scenarios. The saturated hydraulic conductivity of the two soil types varies significantly, with distinct characteristics influencing water transport and drainage, as recorded in Appendix A. Figure 8 presents the weight matrix maps generated through KG-NLNN and SA-NLNN. Some soil structural information, such as stratification, can be reflected from the soil moisture interactions in Figure 8(a)(b). In the scenario where sand is beneath loam, water gradually released from the loam layer can quickly reach various depths of the sand below. Consequently, soil moisture in the lower layers is primarily influenced by the upper loam. As shown in Figure 8(a), the moisture in the lower layer (0.10m, 0.20m, 0.5m, 1.0m) is notably influenced by the moisture at 0.05m. Conversely, with sand above loam, the upper sand rapidly drains water, and the water from the upper sand is absorbed and held by the lower loam. Therefore, soil moisture in the lower layers is mainly affected by the adjacent upper layer, as shown in Figure 8(b). This layered pattern in the weight map serves as a qualitative indicator of soil texture. Although the weights do not have a direct quantitative relationship with the soil hydraulic parameters, they can reflect the difference in hydraulic conductivity between the layers and reveal which layer is more permeable.

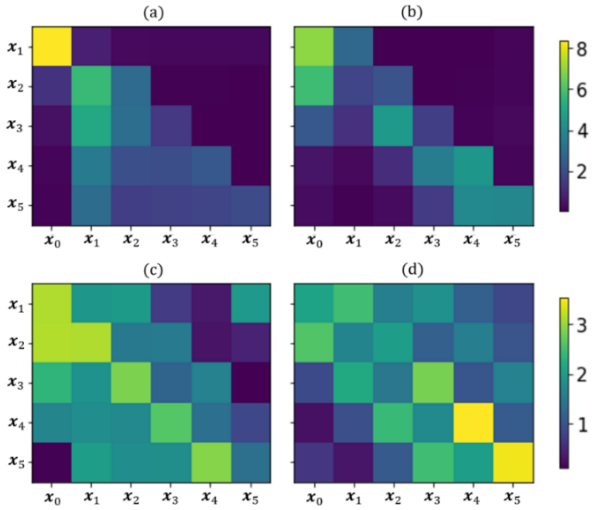

Figure 8. The non-local weight maps in two-layered simulated stratified soil scenarios through KG-NLNN (a) loam above sand (b) sand above loam, and SA-NLNN (c) loam above sand (d) sand above loam.

As a result, both NLNN models achieve satisfactory soil moisture forecasts in the simulated scenarios. Furthermore, the models have advanced the interpretability of machine learning through non-local weight matrix maps. Notably, KG-NLNN offers more reliable qualitative descriptions of soil properties via weights visualizations, highlighting the importance of knowledge guidance.

4.2 Performance evaluation

In this section, we evaluate the performance of the SA-NLNN and KG-NLNN models using in-situ observations from twenty ISMN sites. The performance of LSTM_T, LSTM_T, SA-NLNN, and KG-NLNN is evaluated at five different depths (0.05m, 0.1m, 0.2m, 0.5m, 1.0m). Notably, our NLNN models predict soil moisture for all five depths simultaneously, whereas LSTM_T models each depth separately. When comparing our models with physical models, the inherent methodological differences between machine learning and physical models make fair and direct comparisons with standard knowledge-based modeling particularly challenging. We therefore limit our comparison to a preliminary assessment in Appendix B.

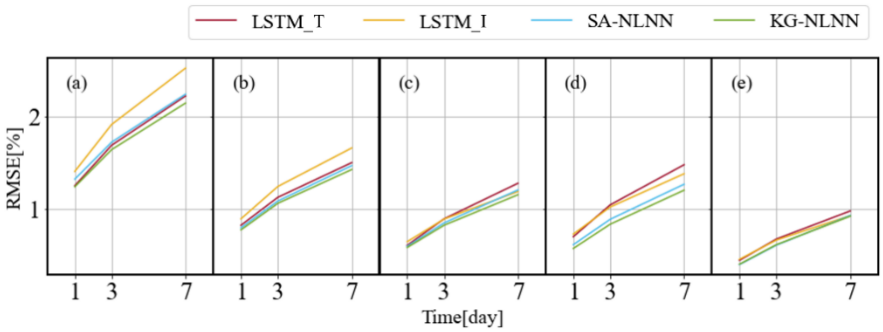

Figure 9. Comparison of mean RMSE for LSTM_T, LSTM_I, SA-NLNN, and KG-NLNN. The values are averaged across twenty research sites and presented separately for each of the five soil depths: 0.05m(a), 0.10m(b), 0.20m(c), 0.50m(d), 1.00m(e).

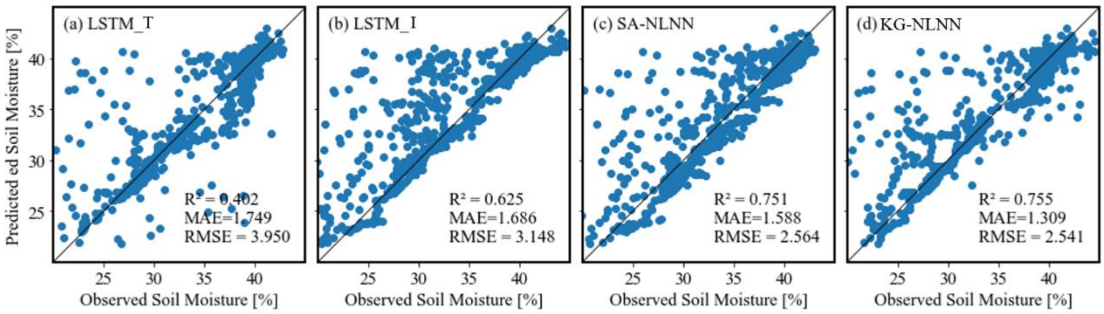

Figure 10. Scatter plots of the soil moisture observations and 7-day predictions generated from (a) LSTM_T, (b) LSTM_I, (c) SA-NLNN, and (d) KG-NLNN at UpperBethlehem.

Table 2. The MAE [%] values for 1, 3, and 7-day forecasts across the four models across twenty research sites at 5 distinct depths, based on ten repeated trainings. The bold values indicate the best performance for each metric across the models.

depth/m	MAE											
	KG-NLNN			SA-NLNN			LSTM_T			LSTM_I		
	1d	3d	7d	1d	3d	7d	1d	3d	7d	1d	3d	7d
0.05	**0.391**	**0.600**	**0.893**	0.440	0.666	0.979	0.737	1.074	1.515	0.808	1.203	1.713
0.10	**0.392**	**0.603**	**0.900**	0.431	0.659	0.972	0.498	0.726	1.027	0.506	0.771	1.113
0.20	0.397	0.607	0.900	0.431	0.648	0.947	**0.356**	0.558	0.844	0.357	**0.547**	**0.787**
0.50	**0.392**	**0.601**	**0.896**	0.432	0.648	0.962	0.405	0.632	0.955	0.403	0.620	0.909
1.00	0.394	0.602	0.885	0.422	0.641	0.943	0.245	0.386	0.597	**0.243**	**0.385**	**0.592**

Table 2 displays the MAE values across twenty selected sites, considering forecasts for 1, 3, and 7 days from the four models at five distinct depths. These results are derived from ten repeated trainings, and the corresponding RMSE results are presented in Figure 9. From MAE results, we observe that both LSTM_1 and LSTM_4 perform well in deep soil moisture predictions. Meanwhile, our proposed NLNN models consistently demonstrate superior accuracy at depths from 0.05m to 0.5m. Regarding RMSE, the KG-NLNN model stands out as the best model in most situations. Figure 10 depicts the correlation between the 7-day soil moisture predictions and observations of the test set for LSTM-4, LSTM-1, SA-NLNN, and KG-NLNN. The density of scatter plots serves as an indicator of model reliability (Datta & Faroughi, 2023). The KG-NLNN model exhibits superior performance in soil moisture prediction compared to the other models, suggesting the stability of our model over longer prediction periods. The comparison between KG-NLNN and SA-NLNN underscores the value of incorporating soil water transport mechanisms into of decoupled non-local operations. Nevertheless, a limitation of the proposed NLNN models lies in their forecasts for moisture content at 1.0m. This limitation could be attributed to the absence of consideration for lower boundary conditions in our study.

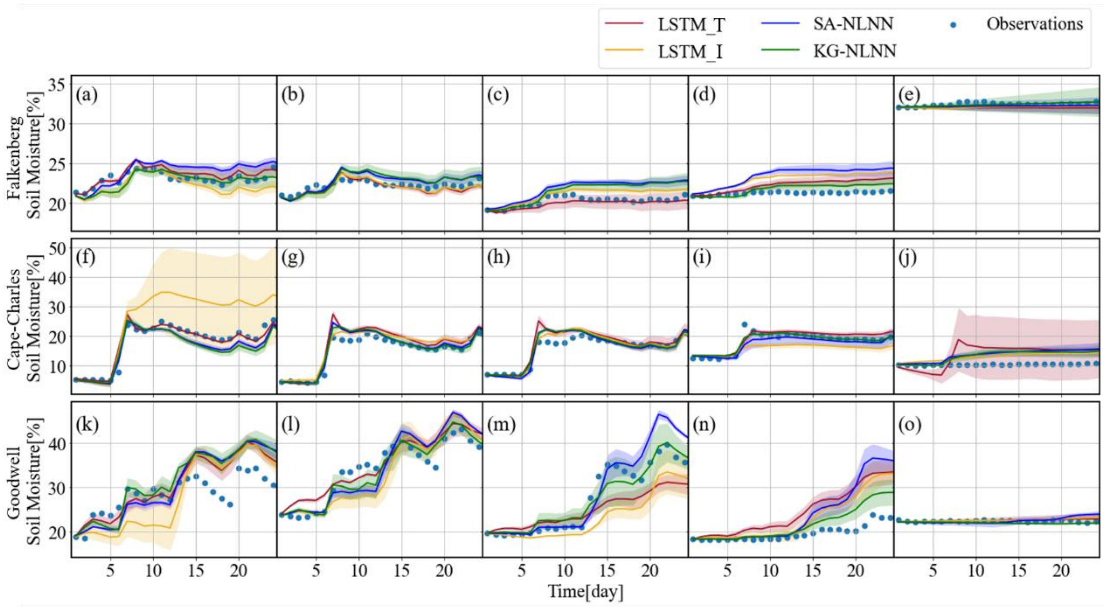

Figure 11. The autoregressive 24-day predicted soil moisture time series of 5 depths with LSTM_I, LSTM_T, KG-NLNN and SA-NLNN at Falkenberg (a-e), Cape-Charles (f-j), and Goodwell (k-o). The shaded region represents the confidence interval of the models, spanning 1 standard deviation.

Regarding how NLNN model predictions change over time, Figure 11 displays the autoregressive 24-day predicted time series soil moisture data for the NLNN models across three sites: Falkenberg, Cape-Charles, and Goodwell. The shaded region represents the confidence interval of the models, spanning 1 standard deviation. The LSTM-based models exhibit relatively greater uncertainty in predictions. However, it is evident that both models perform satisfactorily and stably, with the proposed KG-NLNN model being closer to the observations. Considering the temporal accumulation of autoregressive errors in extended soil moisture forecasting, we provide additional long-term prediction results in Appendix B for comprehensive evaluation.

According to section 4.1, the non-local weight maps can be qualitatively related to the soil properties, demonstrating the interpretability of the model. In real-world cases, even with limited soil information from the site in Table 1, we can combine the weight maps with the measured soil texture data for our analysis. Figure 12 illustrates the non-local weight matrix maps for the Falkenberg, Cape-Charles, and UpperBethlehem sites, generated by the KG-NLNN model. These maps remain stable during repeated training, with discernible variations among the three sites. They offer qualitative interpretations related to soil properties. In Figure 12(a), it is seen that at Falkenberg site, soil moisture at different depths is primarily influenced by upper boundary conditions and upper layer soil moisture. Figure 12(b) shows that at Cape-Charles site, soil moisture is mainly affected by upper boundary conditions and soil moisture at the same depth from the previous time step. Figure 12(c) depicts the strong soil water retention effect at UpperBethlehem site, soil moisture is mainly related to its own state at the previous time step. By combining Table 1, we can see that the non-local weight maps are consistent with the soil texture information. From Falkenberg to UpperBethlehem site, as the soil texture changes from sandy to clay, the learnt water retention capacity in Figure 12 increases from low to high. Consequently, the non-local weight maps are able to capture different physical mechanisms of different sites from the measurement data.

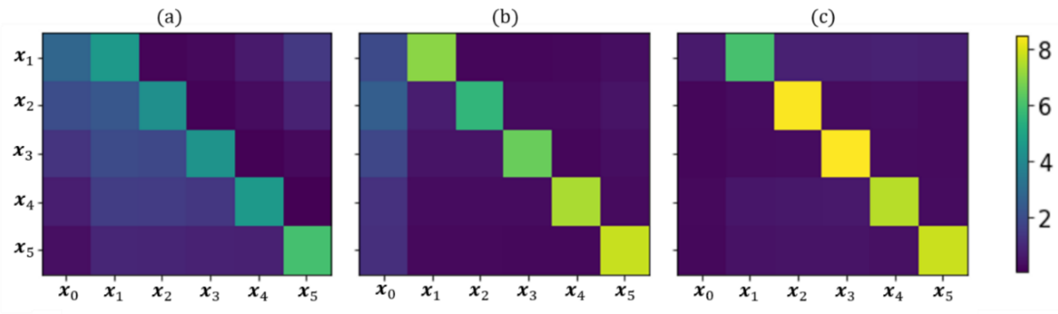

Figure 12. The non-local weight maps through the KG-NLNN at three typical sites, (a) Falkenberg, (b) Cape-Charles, and (c) UpperBethlehem.

In summary, our NLNN models achieve precise and efficient soil moisture predictions across diverse scenarios, as validated by comparisons with LSTMs using in-situ observations. Their multi-depth modeling strategy enhances overall accuracy through complementary interactions. The proposed KG-NLNN model delivers accurate predictions with low uncertainty, while also providing qualitative descriptions of the intricate soil properties. This performance underscores the necessity of incorporating soil water transport knowledge guidance in non-local operation design.

4.3 Effects of the noise levels, time scales, and observation positions

In addition to model accuracy and interpretability, our non-local neural network exhibits adaptability in prediction tasks across different time scales. In this section, we have conducted tests involving different noise levels, time intervals, and observation positions. To further investigate the impact of noise on our NLNN models, we have employed five different noise levels (0.5, 1.0, 2.0, 5.0, 10.0) and compared the NLNN model performance with LSTM models. The RMSE results for soil moisture prediction at 0.05m, 0.10m, 0.20m, 0.50m, and 1.00m are presented in Figure 13. The LSTM_T model demonstrates poor noise resistance and long-term forecasting capability. The other three models perform similarly under low-noise conditions, with LSTM_I even exhibiting some advantage. However, as the noise level increases, NLNN models demonstrate better robustness. Notably, the knowledge-guided NLNN is particularly stable, consistent with its performance on in-situ soil moisture data.

When investigating the KG-NLNN model's performance at the 0.2-day, 0.5-day, and 1-day time intervals within homogenous soil, a subtle difference emerges in the weight map generated by the KG-NLNN model, as illustrated in Figure 14. Despite a decrease in accuracy with longer time intervals, the model consistently achieves satisfactory results. The results reflect the adaptability of the model to diverse time scales.

When the number of observation locations increases to 10 (at depths of 0.05m, 0.1m, 0.2m, 0.3m, 0.4m, 0.5m, 0.6m, 0.7m, 0.8m, 0.9m), the MAE values for soil moisture 1, 3, 7, and 15-day forecasts of the NLNN models across five depths are summarized in Table 3. The uniform augmentation of measurements significantly enhances the prediction accuracy of SA-NLNN, while having minimal impact on the performance of KG-NLNN. This suggests that the knowledge guidance allows for lower requirements on soil moisture measurements. In scenarios with uniformly augmented observations, SA-NLNN may prove more efficient.

In conclusion, both the NLNN models achieve accurate and reliable soil moisture predictions under diverse scenarios. They can adapt to tasks across different time scales. The SA-NLNN performs better under uniformly distributed observations, while the KG-NLNN demonstrates stronger noise resistance.

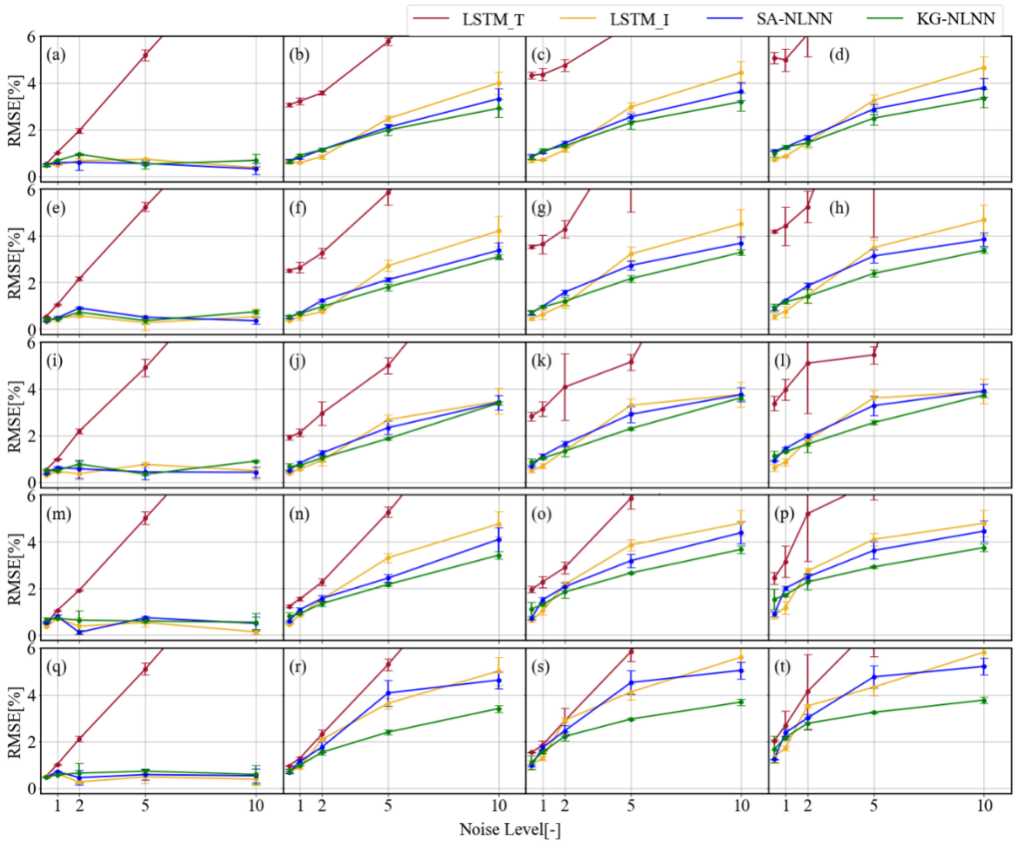

Figure 13. The RMSE results for 1, 3, 7, and 15-day at 0.05m(a-d), 0.10m(e-h), 0.20m(i-l), 0.50m(m-p) and 1.0m(q-t) in the homogenous soil under increasing noise levels. The error bar indicates the standard deviations of the RMSE,

which are computed via ten training replicates. Note: portions of the red curves are truncated where the error significantly exceeds this range, reflecting its relatively lower predictive accuracy.

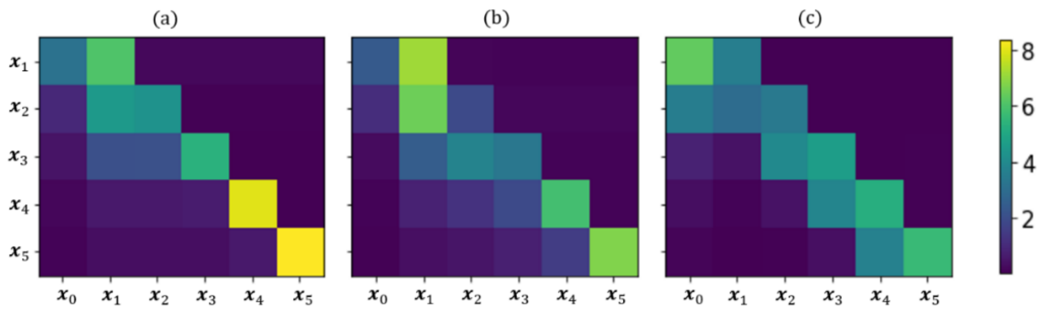

Figure 14. The non-local weight maps of the KG-NLNN model at different time scales at 0.2-day (a), 0.5-day (b), and 1.0-day (c) in the homogenous soil.

Table 3. The MAE [%] values for 1, 3, 7, and 15-day forecasts of the proposed KG-NLNN model and SA-NLNN model at 5 depths with 10 depth measurements under the homogenous soil scenario. The bold values indicate the best performance for each metric across the models.

Depth/m	Homogeneous soil							
	KG-NLNN				SA-NLNN			
	1d	3d	7d	15d	1d	3d	7d	15d
0.05	**0.327**	**0.470**	**0.645**	**0.817**	0.394	0.588	0.906	1.657
0.10	0.280	0.407	0.602	**0.825**	**0.250**	**0.350**	**0.535**	0.892
0.20	0.331	0.564	0.979	1.419	**0.221**	**0.300**	**0.418**	**0.604**
0.50	0.174	0.258	0.380	0.581	**0.148**	**0.204**	**0.302**	**0.502**
1.00	**0.108**	0.180	0.300	0.493	0.118	**0.174**	**0.259**	**0.460**

5. Conclusions

In this study, we employ the deep learning model NLNNs to achieve precise and efficient soil moisture predictions under diverse scenarios without relying physical assumptions., while providing qualitative interpretation for complex soil moisture dynamics, such as vertical heterogeneity and inter-layer connectivity. In light of the accuracy and parameter estimation challenges in physical models, and the credibility concerns in machine learning models, we have introduced a framework that integrates both accuracy and mechanistic insight. Our method leverages in-profile soil moisture interactions across various depths. Consequently, the soil moisture prediction task is reformulated as a single-time-step prediction task that involves multi-depth soil moisture variables. In this way, we apply the self-attention-based model SA-NLNN to explore the potential of the NLNN structure. Expanding on this framework, we disentangle the non-local operation into four components to create the KG-NLNN model according to the soil water transport knowledge. By comparing our NLNNs with the LSTM model using synthetic data and in-situ observations, we demonstrate that both our NLNN models achieve precise and effective forecasts, providing an alternative possibility for soil moisture simulations. The knowledge-guided model KG-NLNN exhibits the best performance and remains stable with low uncertainty. The physical knowledge guidance in non-local operations significantly enhances the model's accuracy and reliability.

Additionally, our proposed models offer qualitative interpretations related to the soil properties. Through the investigation of various virtual scenarios -- including homogeneous soil, heterogeneous soil, two-layered soil, and soil with root water uptake -- we observe that both the KG-NLNN and SA-NLNN models perform well in different soil conditions. The qualitative interpretations derived from soil moisture data generated by KG-NLNN facilitate descriptions of soil textures. When testing with in-situ data, we find that the KG-NLNN model also provides interpretations consistent with real soil vertical heterogeneity without physical assumptions. This highlights the importance of integrating knowledge-guided assistance into model design. Moreover, we have assessed the model's performance under different noise conditions, observation positions, and time scales. Both NLNN models exhibit robustness to noise, and the knowledge guidance enhances noise resistance. Besides, NLNN model demonstrates adaptability to diverse time scales. When observations are evenly distributed, the SA-NLNN shows significant improvements compared to KG-NLNN, while maintaining high computational efficiency.

Nevertheless, the model faces challenges that necessitate future improvements. Its training and application are site-specific, limiting its transferability. Further research is required to enhance its applicability across different sites. Specifically, difficulties arise in estimating soil moisture content at deep layers, possibly due to the lack of consideration for the groundwater boundary. Incorporating lower boundary conditions into the model could address this limitation. Additionally, multi-objective network training may benefit from more effective strategies and more precise loss function designs. Introducing constraints at multiple time steps holds promise for achieving more stable results. Finally, further refinement of the non-local operation may enhance the model's performance. What’s more, the proposed network framework is architecturally flexible and modular, making it customizable for diverse research requirements. Beyond soil moisture, the NLNN-based strategy could be readily extended to other systems, such as solute transport in groundwater. We encourage the exploration of such specialized structures to address various coupled physical or hydrological problems across different scales.

AUTHOR CONTRIBUTION

Y. Wang: Conceptualization, Methodology, Software, Writing–original draft. X. Hu: Writing – review & editing, Supervision. Y. Hu: Supervision. L. He: Writing – review & editing. L. Shi: Writing – review & editing, Supervision. L. Wang: Writing – review & editing. W. Song: Methodology, Writing – review & editing.

ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Program of China (2021YFC3201203) and the National Natural Science Foundation of China (Grant 52179038 and Grant U2243235).

CODE/DATA AVAILABILITY

The data and codes used in this paper are available on the website (https://doi.org/10.5281/zenodo.10408929).

Appendix A

The parameters used to generate the synthetic data are recorded in Table A1 and Table A2:

Table A1. The van Genuchten soil hydraulic parameters (van Genuchten, 1980) used for synthetic data generation.

Case Design	Homogenous soil	Heterogeneous soil	Two-layered soil	Soil with root water uptake
θ_r $[-]$	0.078	0.078	0.078	0.078
θ_s $[-]$	0.43	0.43	0.43	0.43
α $[cm^{-1}]$	3.6	3.6	3.6	3.6
n $[-]$	1.56	1.56	1.56	1.56
K_s $[cm\ day^{-1}]$ $(0-10cm)$	0.250	Table A2	0.250	0.250
K_s $[cm\ day^{-1}]$ $(10-100cm)$	0.250	Table A2	10.49	0.250
l $[-]$	0.5	0.5	0.5	0.5
Presence of plant	False	False	False	True

Table A2. The soil hydraulic conductivity of the heterogeneous scenario.

depth$[cm]$	K_s $[cm\ day^{-1}]$									
$0-10$cm	0.226	0.270	0.241	0.263	0.222	0.226	0.263	0.221	0.262	0.276
$10-20$cm	0.230	0.226	0.217	0.226	0.249	0.203	0.229	0.196	0.207	0.202
$20-30$cm	0.200	0.239	0.244	0.253	0.251	0.248	0.203	0.225	0.206	0.205
$30-40$cm	0.241	0.223	0.197	0.227	0.218	0.256	0.258	0.294	0.308	0.242
$40-50$cm	0.242	0.155	0.177	0.184	0.218	0.230	0.225	0.211	0.207	0.252
$50-60$cm	0.285	0.338	0.351	0.345	0.317	0.355	0.333	0.343	0.322	0.320
$60-70$cm	0.261	0.272	0.306	0.279	0.319	0.250	0.262	0.224	0.240	0.269
$70-80$cm	0.269	0.300	0.276	0.250	0.267	0.233	0.240	0.249	0.207	0.233
$80-90$cm	0.202	0.209	0.208	0.248	0.231	0.232	0.245	0.258	0.250	0.222
$90-100$cm	0.254	0.211	0.201	0.203	0.186	0.213	0.233	0.196	0.247	0.213

Appendix B

This section presents a preliminary comparison between the NLNN model and the physics-based soil moisture model derived from Richards' equation.

The Ross method (P J Ross, 2003; Peter J Ross, 2006) is a rapid, non-iterative numerical scheme for soil moisture forward modeling based on Richards' Equation. For boundary conditions, the daily reference evapotranspiration (ET0) is calculated with the FAO Penman-Monteith method (Allen et al., 1998). As standardized in the FAO guidelines (Allen et al., 1998), actual evapotranspiration is the product of K_C and ET0, where K_C serves as a refined empirical parameter. When generating synthetic data, we applied this empirical coefficient method to derive a preliminary evapotranspiration estimate, adopting a coefficient value of 1.0 in this instance. We first utilize 10 days of site historical data to invert the site-specific soil hydraulic parameters (α, n,K_s) through data assimilation with the ensemble Kalman filter (EnKF) method (Evensen, 2003) within the Ross framework. These parameters are then applied in the Ross method to obtain a fast solution of one-dimensional Richards' equation, enabling the forecasting of soil moisture dynamics.

In the real-world experiments, we selected three sites: Falkenberg, Cape-Charles, and Goodwell, with distinctly different soil textures and land covers, as recorded in Table 1 in the manuscript. Figure B1 illustrates the autoregressive 24-day predicted time series soil moisture data for the KG-NLNN model and Ross-EnKF across these three sites. The MAE results are recorded in Table B1. It is seen that soil moisture forecasts obtained by KG-NLNN are closer to real observations, compared to the traditional Ross-EnKF method.

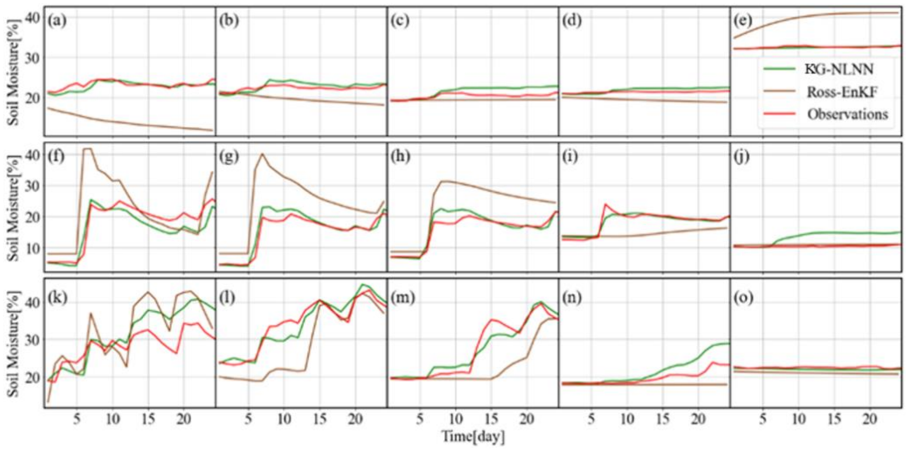

Figure B1. The 24-day predicted soil moisture time series of 5 depths with KG-NLNN and Ross-EnKF at Falkenberg (a-e), Cape-Charles (f-j), and Goodwell (k-o).

However, it should be noted that the data assimilation process in Ross-EnKF did not update soil infiltration parameters, potentially disadvantaging the physical model. What's more, the proposed approaches cannot predict soil moisture at arbitrary depths and times as the physical models. The fundamental differences between machine learning and physical modeling make fair, direct comparisons with standard methods both critical and difficult.

Table B1. The MAE [%] values for 24-day forecasts of the proposed KG-NLNN model and Ross-EnKF model

	Falkenberg	Cape-Charles	Goodwell
KG-NLNN	0.681	1.766	1.998
Ross-EnKF	4.395	5.484	3.840

Moreover, our machine learning approach exhibits autoregressive error accumulation in long-term soil moisture predictions—a limitation not observed in knowledge-based modeling. As demonstrated by the 120-day autoregressive forecasts (Figure B2), while model uncertainty gradually accumulates with prediction time, it remains within acceptable bounds. Importantly, the knowledge-guided KG-NLNN model maintains significantly greater stability across the entire prediction horizon.

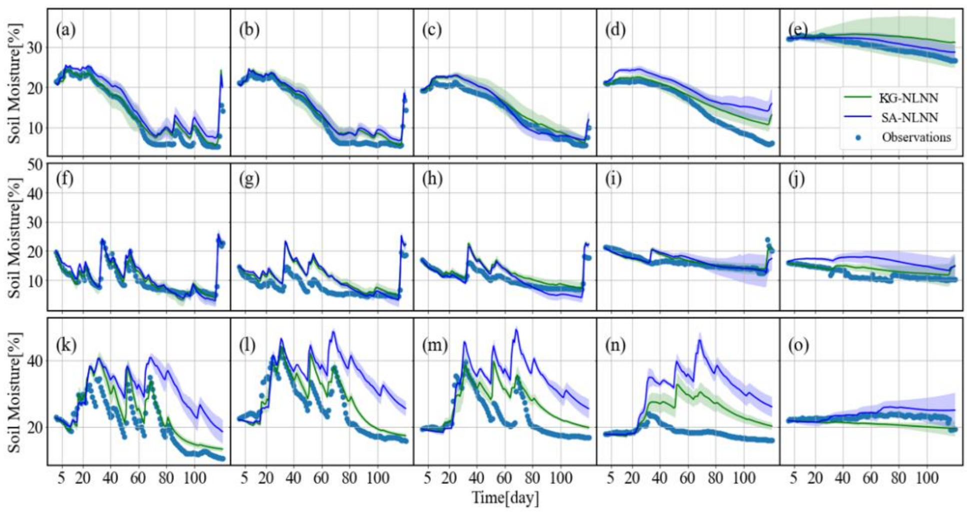

Figure B2. The 120-day predicted soil moisture time series of 5 depths with KG-NLNN and SA-NLNN at Falkenberg (a-e), Cape-Charles (f-j), and Goodwell (k-o).

Appendix C

Table C1. The MAE [%] values for 1, 3, 7, and 15-day forecasts of LSTM_T LSTM_I, the proposed KG-NLNN and SA-NLNN model at 5 depths under four designed scenarios.

Depth/m	KG-NLNN															
	homogeneous				heterogeneous				two-layered				root water uptake			
	1d	3d	7d	15d	1d	3d	7d	15d	1d	3d	7d	15d	1d	3d	7d	15d
0.05	0.235	**0.327**	**0.433**	**0.539**	**0.259**	**0.372**	**0.510**	**0.652**	0.449	0.680	0.945	**1.170**	0.528	0.747	0.996	1.224
0.10	0.313	0.451	0.627	0.788	0.306	0.431	0.593	0.749	0.521	0.745	0.995	**1.191**	0.409	0.544	0.685	0.825
0.20	0.342	0.533	0.776	1.016	0.305	0.488	0.736	0.971	0.433	0.649	0.901	**1.179**	0.623	0.852	1.056	1.212
0.50	0.235	0.357	0.545	0.782	0.253	0.399	0.630	0.952	0.334	0.518	0.774	1.098	0.375	0.598	0.870	1.182
1.00	0.203	**0.312**	**0.445**	**0.647**	0.244	0.397	0.618	0.934	0.368	0.625	0.969	1.329	0.278	0.455	0.749	**1.120**
Depth/m	SA-NLNN															
	homogeneous				heterogeneous				two-layered				root water uptake			
	1d	3d	7d	15d	1d	3d	7d	15d	1d	3d	7d	15d	1d	3d	7d	15d
0.05	0.328	0.470	0.686	1.039	0.363	0.524	0.750	1.840	0.327	0.505	0.836	2.210	0.536	0.918	2.150	6.702
0.10	0.249	0.375	0.580	0.957	0.220	0.314	0.477	0.851	0.390	0.569	**0.808**	1.465	0.322	0.447	0.675	1.480
0.20	0.262	0.366	0.519	0.820	0.292	0.389	0.482	**0.648**	0.487	0.696	0.945	1.350	0.379	0.546	0.775	1.861
0.50	0.209	0.291	0.414	0.566	0.265	0.337	0.431	**0.623**	0.327	0.483	0.708	**1.018**	0.344	0.485	0.687	1.502
1.00	0.245	0.376	0.575	0.807	0.282	0.430	0.640	0.941	0.336	0.530	0.810	**1.250**	0.297	0.482	0.820	1.748
Depth/m	LSTM_T															
	homogeneous				heterogeneous				two-layered				root water uptake			
	1d	3d	7d	15d	1d	3d	7d	15d	1d	3d	7d	15d	1d	3d	7d	15d
0.05	**0.009**	1.503	2.791	3.874	**0.010**	1.520	2.829	3.936	**0.015**	1.422	2.795	4.106	**0.020**	1.649	3.187	4.641
0.10	**0.007**	1.176	2.237	3.184	**0.007**	1.202	2.282	3.240	**0.015**	1.479	2.879	4.179	**0.014**	1.255	2.449	3.584
0.20	**0.008**	0.786	1.630	2.380	**0.010**	0.782	1.628	2.384	**0.012**	0.836	1.735	2.561	**0.013**	0.801	1.671	2.479
0.50	**0.006**	0.406	0.942	1.483	**0.008**	0.373	0.872	1.375	**0.008**	0.400	0.933	1.476	**0.009**	0.403	0.939	1.482
1.00	**0.008**	0.266	0.662	1.116	**0.007**	0.266	0.664	1.121	**0.006**	0.258	0.644	1.103	**0.006**	0.267	0.667	1.136
Depth/m	LSTM_I															
	homogeneous				heterogeneous				two-layered				root water uptake			
	1d	3d	7d	15d	1d	3d	7d	15d	1d	3d	7d	15d	1d	3d	7d	15d
0.05	0.318	0.440	0.590	0.845	0.346	0.484	0.656	0.948	0.264	**0.451**	**0.771**	1.313	0.343	**0.462**	**0.600**	**0.804**
0.10	0.135	**0.202**	**0.319**	**0.528**	0.149	**0.249**	**0.408**	**0.699**	0.359	**0.542**	0.863	1.436	0.274	**0.365**	**0.491**	**0.707**
0.20	0.120	**0.174**	**0.262**	**0.444**	0.138	**0.217**	**0.359**	0.669	0.218	**0.320**	**0.545**	1.072	0.159	**0.238**	**0.366**	**0.594**
0.50	0.128	**0.177**	**0.274**	**0.494**	0.142	**0.207**	**0.341**	0.640	0.179	**0.293**	**0.542**	1.090	0.173	**0.276**	**0.443**	**0.742**
1.00	0.214	0.350	0.594	1.075	0.188	**0.288**	**0.465**	**0.865**	0.180	**0.293**	**0.578**	1.343	0.242	**0.393**	**0.672**	1.203

References:

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, *300*(9), D05109.

Bandai, T., & Ghezzehei, T. A. (2021). Physics-Informed Neural Networks With Monotonicity Constraints for Richardson-Richards Equation: Estimation of Constitutive Relationships and Soil Water Flux Density From Volumetric Water Content Measurements. *Water Resources Research*, *57*(2). https://doi.org/10.1029/2020WR027642

De Bézenac, E., Pajot, A., & Gallinari, P. (2018). Deep learning for physical processes: Incorporating prior scientific knowledge. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123–140.

Buckingham, E. (1907). Studies on the movement of soil moisture.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314.

Datta, P., & Faroughi, S. A. (2023). A multihead LSTM technique for prognostic prediction of soil moisture. *Geoderma*, *433*(March), 116452. https://doi.org/10.1016/j.geoderma.2023.116452

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*(Mlm), 4171–4186.

Ding, Y., Zhu, Y., Wu, Y., Jun, F., & Cheng, Z. (2019). Spatio-Temporal attention lstm model for flood forecasting. *Proceedings - 2019 IEEE International Congress on Cybermatics: 12th IEEE International Conference on Internet of Things, 15th IEEE International Conference on Green Computing and Communications, 12th IEEE International Conference on Cyber, Physical and So*, 458–465. https://doi.org/10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00095

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Entekhabi, D., Rodriguez-Iturbe, I., & Castelli, F. (1996). Mutual interaction of soil moisture state and atmospheric processes. *Journal of Hydrology*, *184*(1–2), 3–17. https://doi.org/10.1016/0022-1694(95)02965-6

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*, 343–367.

Fang, K., Pan, M., & Shen, C. (2019). The Value of SMAP for Long-Term Soil Moisture Estimation with the Help of Deep Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(4), 2221–2233. https://doi.org/10.1109/TGRS.2018.2872131

Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. In *International conference on machine learning* (pp. 1243–1252). PMLR.

van Genuchten, M. T. (1980). A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Science Society of America Journal*, *44*(5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Gill, M. K., Asefa, T., Kemblowski, M. W., & McKee, M. (2006). Soil moisture prediction using support vector machines. *Journal of the American Water Resources Association*, *42*(4), 1033–1046. https://doi.org/10.1111/j.1752-1688.2006.tb04512.x

Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., et al. (2022). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, *8*(3), 331–368.

Guswa, A. J., Celia, M. A., & Rodriguez-Iturbe, I. (2002). Models of soil moisture dynamics in ecohydrology: A comparative study. *Water Resources Research*, *38*(9), 5-1-5–15. https://doi.org/10.1029/2001wr000826

Heathman, G. C., Cosh, M. H., Merwade, V., & Han, E. (2012). Multi-scale temporal stability analysis of surface and subsurface soil moisture within the Upper Cedar Creek Watershed, Indiana. *Catena*, *95*, 91–103. https://doi.org/10.1016/j.catena.2012.03.008

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

Huang, G. Bin, Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(1–3), 489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Jiang, S., Zheng, Y., & Solomatine, D. (2020). Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning. *Geophysical Research Letters*, *47*(13). https://doi.org/10.1029/2020GL088229

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. *ACM Computing Surveys (CSUR)*, *54*(10s), 1–41.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Kornelsen, K. C., & Coulibaly, P. (2014). Root-zone soil moisture estimation using data-driven methods. *Water Resources Research*, *50*(4), 2946–2962.

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.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y. (1989). Generalization and network design strategies. *Connectionism in Perspective*, *19*(143–155), 18.

Lim, B., Arık, S., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Liu, P., Chang, S., Huang, X., Tang, J., & Cheung, J. C. K. (2019). Contextualized non-local neural networks for sequence learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 6762–6769).

Liu, Y., Mei, L., & Ki, S. O. (2014). Prediction of soil moisture based on Extreme Learning Machine for an apple orchard. *CCIS 2014 - Proceedings of 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems*, 400–404. https://doi.org/10.1109/CCIS.2014.7175768

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). Swin Transformer. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9992–10002. Retrieved from https://ieeexplore.ieee.org/document/9710580/

Minasny, B., Bandai, T., Ghezzehei, T. A., Huang, Y. C., Ma, Y., McBratney, A. B., et al. (2024). Soil Science-Informed Machine Learning. *Geoderma*, *452*(October), 117094. https://doi.org/10.1016/j.geoderma.2024.117094

Prasad, R., Deo, R. C., Li, Y., & Maraseni, T. (2019). Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *Catena*, *177*(February), 149–166. https://doi.org/10.1016/j.catena.2019.02.012

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Raissi, Maziar, Perdikaris, P., & Karniadakis, G. E. (2017). Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations, (Part I), 1–22. Retrieved from http://arxiv.org/abs/1711.10566

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Journal of Applied Physics*, *1*(5), 318–333. https://doi.org/10.1063/1.1745010

Romero-Ruiz, A., Linde, N., Keller, T., & Or, D. (2018). A review of geophysical methods for soil structure characterization. *Reviews of Geophysics*, *56*(4), 672–697.

Ross, P J. (2003). Modeling soil water and solute transport—Fast, simplified numerical solutions. *Agronomy Journal*, *95*(6), 1352–1361.

Ross, Peter J. (2006). Fast solution of Richards' equation for flexible soil hydraulic property descriptions. *Land and Water Technical Report, CSIRO*, *39*(06).

Saxton, K. E., Johnson, H. P., & Shaw, R. H. (1974). Modeling Evapotranspiration and Soil Moisture. *Transactions of the American Society of Agricultural Engineers*, *17*(4), 673–677. https://doi.org/10.13031/2013.36935

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE Transactions on Neural Networks*, *20*(1), 61–80.

Semwal, V. B., Gupta, A., & Lalwani, P. (2021). An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition. *Journal of Supercomputing*, *77*(11), 12256–12279. https://doi.org/10.1007/s11227-021-03768-7

Severyn, A., & Moschitti, A. (2015). UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification. *SemEval 2015 - 9th International Workshop on Semantic Evaluation, Co-Located with the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2015 - Proceedings*, (SemEval), 464–469. https://doi.org/10.18653/v1/s15-2079

Shaw, P., Uszkoreit, J., & Vaswani, A. (2018). Self-attention with relative position representations. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *2*, 464–468. https://doi.org/10.18653/v1/n18-2074

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, *2015-Janua*, 802–810.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The performance of LSTM and BiLSTM in forecasting time series. In *2019 IEEE International conference on big data (Big Data)* (pp. 3285–3292). IEEE.

Simunek, J., Van Genuchten, M. T., & Sejna, M. (2005). The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. *University of California-Riverside Research Reports*, *3*, 1–240.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., & Hopmans, J. W. (2008). On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resources Research*, *46*(4), 1–21. https://doi.org/10.1029/2008WR006829

Vereecken, Harry, Amelung, W., Bauke, S. L., Bogena, H., Brüggemann, N., Montzka, C., et al. (2022). Soil hydrology in the Earth system. *Nature Reviews Earth and Environment*, *3*(9), 573–587. https://doi.org/10.1038/s43017-022-00324-6

Wang, W., Wei, Y., Hao, L., Wei, Z., & Zhao, T. (2025). Soil moisture forecasting in wireless sensor networks via spatiotemporal graph convolutional networks, (May 2024), 1–17. https://doi.org/10.1002/vzj2.70000

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7794–7803).

Wang, Y., Shi, L., Hu, Y., Hu, X., Song, W., & Wang, L. (2023). A comprehensive study of deep learning for soil moisture prediction. *Hydrology and Earth System Sciences Discussions*, *2023*, 1–38.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Advances in Neural Information Processing Systems*, *15*(NeurIPS), 12077–12090.

Yin, M., Yao, Z., Cao, Y., Li, X., Zhang, Z., Lin, S., & Hu, H. (2020). Disentangled non-local neural networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16* (pp. 191–207). Springer.

Zhang, C., Liu, J., Shang, J., & Cai, H. (2018). Capability of crop water content for revealing variability of winter wheat grain yield and soil moisture under limited irrigation. *Science of the Total Environment*, *631*, 677–687.

Zhu, L., She, Q., Li, D., Lu, Y., Kang, X., Hu, J., & Wang, C. (2021). Unifying Nonlocal Blocks for Neural Networks. *Proceedings of the IEEE International Conference on Computer Vision*, 12272–12281. https://doi.org/10.1109/ICCV48922.2021.01207

Zhu, Z., Xu, M., Bai, S., Huang, T., & Bai, X. (2019). Asymmetric non-local neural networks for semantic segmentation. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 593–602. https://doi.org/10.1109/ICCV.2019.00068