

Dear Editor [Bo Guo](#):

We are grateful to receive comments and suggestions on our manuscript of '[EGUSPHERE-2025-4440](#)' from you and the reviewers. We have carefully considered these comments and made significant revisions. All the modifications in the revised manuscript are highlighted in blue color.

The responses to the reviewer's comments are also appended for your convenience.

We hope the revision will be satisfactory. If you have any further questions or concerns, please let us know. Once again, we appreciate your time on this manuscript.

Yours sincerely,

Liangsheng Shi

Wuhan University

Referee #1

The authors have addressed all my concerns; however, I recommend that they thoroughly review and correct the citation format of the entire manuscript.

Response:

We sincerely thank the reviewer for their careful reading and constructive feedback. We have thoroughly reviewed the citation format throughout the manuscript

Referee #2

This manuscript proposes a novel application of non-local neural networks for multi-depth soil moisture prediction and presents promising results on both synthetic and in-situ datasets. The idea of reformulating the problem as a single-step, profile-wise prediction and visualizing non-local weights to infer inter-layer interactions is valuable and interesting. The manuscript is well organized, but may need improvement in conceptual clarity, methodological transparency, and consistency of physical interpretation. Please see below for my comments.

Response:

We sincerely thank the reviewer for taking the time to carefully read our manuscript and for providing constructive comments. We have carefully considered all the suggestions provided and made revisions to improve conceptual clarity, methodological transparency, and consistency of physical interpretation throughout the manuscript.

Major comments:

1. Ambiguity of “Interpretability” and “Physics / Knowledge Guidance”: A central claim of the paper is that the proposed framework is interpretable and physics- or knowledge-guided, but these concepts are not clearly or rigorously defined.

Response1:

Thank you for your comment. To clarify, the interpretability of our model primarily refers to the ability to qualitatively analyze the learned non-local weight matrices, which reveal how soil moisture at different depths interacts within the profile. This interpretability is structural and qualitative, rather than quantitative or strictly linked to physical parameters.

Similarly, by “knowledge-guided” or “physics-guided,” we refer to the fact that the KG-NLNN architecture incorporates domain knowledge in its model design rather than directly enforcing physical equations or hydraulic parameters. Specifically, the non-local interaction mechanism is designed to reflect vertical soil moisture relationships by decomposing the influences on soil moisture at a given depth into four components, namely meteorological forcing, upper-layer soil moisture, same-layer soil moisture, and lower-layer soil moisture. These components are governed by different physical processes and thus contribute differently. The proposed model does not require any soil hydraulic parameterization or PDE constraints; instead. We have revised the manuscript to clarify these points in Lines 17-23:

“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 quantitative representation of inter-layer connectivity. Notably, the model guided by soil moisture transport knowledge yields more stable and reasonable interpretations.”

2. Methodological Clarity and Internal Consistency: Several aspects of the model formulation and workflow are insufficiently specified or internally inconsistent.

Response2:

Thank you for your comment. Following the reviewer's suggestions, we have carefully reviewed the manuscript and made corresponding revisions to improve methodological clarity and internal consistency. The model formulation, workflow description, and related explanations have been revised and clarified in the appropriate sections, as detailed in the subsequent responses. We believe these changes have improved the overall coherence of the manuscript.

3. Fairness and Transparency of Model Comparisons: The manuscript claims superior performance of KG-NLNN over LSTM and SA-NLNN, but it is unclear whether the comparisons are fair and well controlled.

Response3:

Thank you for this comment. We would like to clarify that we have made deliberate efforts to ensure fair and controlled comparisons among KG-NLNN, LSTM, and SA-NLNN. In Response 8, we provide detailed information on the number of trainable parameters and the training epochs for all compared models. These settings were carefully adjusted to keep model complexity and training budgets comparable across different architectures, thereby minimizing potential biases arising from overparameterization or unequal training effort.

All models were trained and evaluated using the same datasets, input variables, and evaluation metrics to further ensure consistency and fairness in the comparison. We have clarified these points in the revised manuscript in Lines 413-414 to improve transparency regarding the experimental design and model comparison procedure.

4. Interpretation of Weight Maps and Physical Meaning: The visual analysis of non-local weight matrices is central to the manuscript's interpretability claims, but the physical interpretation is currently speculative.

Response4:

Thank you for your comment. We acknowledge that the physical interpretation of the non-local weight matrices is primarily qualitative rather than strictly quantitative. The interpretability analysis is conducted using controlled virtual test cases, which are designed to examine whether the learned non-local interactions reflect meaningful soil moisture interactions. Because the proposed model does not explicitly impose soil hydraulic equations or parameterize soil hydraulic properties, a direct quantitative correspondence between the learned non-local weights and specific physical parameters is not intended. Instead, the non-local weights are interpreted as relative indicators of inter-layer influence and interaction strength.

Nevertheless, the results show that the learned weight patterns capture qualitatively consistent features of soil water movement, such as variations in vertical interaction strength under contrasting soil conditions. In this sense, the weight maps provide a qualitative description of soil properties rather than a quantitative physical characterization. We have revised the manuscript to clarify this distinction and to avoid overinterpreting the learned weights as direct representations of physical parameters. The corresponding descriptions are listed in Lines 18-23, 112-115, and section 4.1 in the manuscript.

5. Claims of Generalizability and Flexibility: The manuscript claims that the framework is flexible and easily customizable, yet also indicates that models are trained site-specifically.

Response5:

Thank you for your comment. We would like to clarify that the claimed flexibility and customizability refer to the NLNN framework itself, rather than to the direct transferability of a trained model across sites. The NLNN framework is designed to be flexible and task-adaptive, allowing users to incorporate different structural designs or domain knowledge according to specific applications. The corresponding descriptions are recorded in Lines 98-100. Within this general framework, we developed the KG-NLNN architecture specifically for soil moisture prediction.

In this study, the KG-NLNN model is trained and evaluated at a single site, with the primary objective of achieving accurate soil moisture predictions under realistic soil conditions by leveraging the interactions among soil moisture at different depths within the profile. The model is intended to capture site-specific soil moisture dynamics to improve predictive performance under local environmental conditions. Due to the strong spatial heterogeneity of soil properties and the complexity of real-world environments, the dynamics learned by a neural network at one site may not directly generalize to another site. While cross-site application may provide relatively stable predictions, high predictive accuracy achieved at one location cannot be guaranteed at another. Developing models with high cross-site generalizability is a focus of our future research.

Specific comments:

Title: Does the “interpretable” mean that the deep learning model is interpretable, or the modeling results are interpretable? Additional, what are the new findings, which the interpretable approach can obtain while traditional models cannot.

Response 6:

Thank you for your comment. Regarding the 'interpretable', we refer to the ability of our model to provide qualitatively interpretable non-local weights while performing high-accuracy predictions. These weights reflect the interactions among soil moisture at different depths, thereby offering a qualitative description of soil properties. In terms of machine learning interpretability, this belongs to intrinsic or model-based interpretability, specifically weight-based interpretability, because the model structure itself produces these non-local weights that can be directly analyzed to understand soil moisture dynamics.

Compared with traditional physical models, our approach achieves higher accuracy in short-term predictions without relying on physical assumptions or governing equations, thus avoiding explicit parameterization of soil hydraulic properties, while still capturing qualitative information about soil moisture dynamics. Compared with conventional machine learning models, our model not only delivers stable predictions but also provides interpretable non-local weight maps, making the results more reliable and insightful. This combination of predictive performance and intrinsic interpretability distinguishes our approach from existing models.

Line 20 and 25: As pointed in the prior comments, “guided by physics” and “physics guidance” can be misleading.

Response7:

Thank you for your comment. We have replaced these terms with more precise descriptors: 'guided by soil moisture transport knowledge' and 'knowledge guidance', and have checked throughout the manuscript.

Line 20: While the authors showed that KG-NLNN outperformed the others, they should justify the

fairness of the comparisons e.g., the number of input and tuning parameters are similar across the models.

Response8:

Thank you for your comments. We have included detailed information about the proposed NLNN models in comparison to the standard LSTM model, concerning their complexity and parameterization. Notably, we have revised the names of the LSTM models according to your request: LSTM_4 has been changed to LSTM_T, and LSTM_1 to LSTM_I. The details can be found in Response 41.

In terms of model complexity, Table R1 lists the training time required for training each model over 100 epochs and the total number of epochs needed. The SA-NLNN, with its efficient matrix operations, exhibits the lowest computational cost and highest efficiency. While the KG-NLNN model incurs relatively higher computational expenses, it remains within an acceptable range. Regarding model parameterization, we have provided the number of parameters requiring training in Table R1. It's worth noting that the LSTM model are constructed for each depth, necessitating a multiplication factor of 5 for model parameters at each site. This indicates that our proposed NLNN models maintain an overall small number of parameters. This indicates that our model has a comparable number of parameters to the LSTM model, making the comparison fair.

Table R1. Comparisons of required training time, epochs, and parameters number of four models for each site.

	KG-NLNN	SA-NLNN	LSTM_T	LSTM_I
Training Time (s/100epochs)	40	12	10	8
Total Epochs	2000	2000	1000	1000
Total time (1000s)	80	24	50	40
Parameters per Site	7120	7529	5781	11242

Line 50: What is an “extreme learning machine”?

Response9:

An Extreme Learning Machine (ELM) (Huang et al., 2006) is a type of single-hidden-layer feedforward neural network (SLFN) that stands out for its unique training mechanism. Unlike traditional deep learning models that rely on gradient-based backpropagation, an ELM randomly initializes the hidden layer weights and biases, which remain fixed during training. The output weights are then analytically determined through a simple matrix inversion (the Moore-Penrose generalized inverse). This method has been widely used in various machine learning applications (Liu et al., 2014).

Line 70: The introduction of physics-informed neural networks appears to be distractive

Response10:

Thank you for your suggestion. We have revised this section by removing the detailed introduction of PINNs and retaining only a brief definition in Lines74-80. The aim is to emphasize that most current models are based on traditional structures, and that integrating domain knowledge into model architectures remains critically important.

Line 75: The citation formats are not consistent.

Response11:

We have removed it.

Line 90: Are you referring to time stepping or iterative solver at each time step? The current statement does not provide a clear explanation of how the spatial interactions were done, nor the difference between the current method and physical methods. Please elaborate

Response12:

Thank you for your comment. We are referring to the spatial interactions between different soil depths within a single time step, rather than the temporal time-stepping or an iterative numerical solver.

Unlike physical models (e.g., those based on the Richards equation) that often require iterative solvers to handle non-linear spatial coupling at each time step, our NLNN approach captures these interactions through a non-local mechanism. Specifically, the spatial interaction is achieved by a weighted aggregation where the state of each layer is updated by considering the states of all other layers simultaneously. This allows the model to represent vertical heterogeneity and inter-layer connectivity in a single forward pass, while maintaining physical consistency through the knowledge-guided masks.

We have revised (now Lines84-87) to explicitly clarify this mechanism:

“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.”

Line 90: “Temporal variations ...” does not logically flow in the current context.

Response13:

Thank you for your suggestion. We have removed this part.

Line 95: What does it mean by “GNN rely on explicit, pre-defined graph structures”?

Response14:

Thank you for your comment. In the context of soil moisture modeling, Graph Neural Networks (GNNs) typically require an adjacency matrix to be defined a priori to represent the spatial relationships between different soil layers or locations. This structure is "explicit" because it is manually constructed based on physical proximity (e.g., connecting only adjacent layers) or predefined heuristics. Consequently, the model's ability to capture spatial interactions is constrained by this fixed topology.

In contrast, our proposed NLNN does not require a predefined graph. Instead, it computes dynamic, implicit dependencies between all soil depths simultaneously using a non-local attention mechanism. This allows the model to capture correlations that might be missed by a predefined, locally-connected graph structure.

Line 100: The NLNNs sound like a GNN that considers hopping (message aggregation) on the entire graph. Can you clarify if that is the case? If not, please elaborate the difference then.

Response15:

Thank you for your comment. Both NLNNs and GNNs involve information aggregation across different nodes (soil layers). However, there are differences in their mechanisms: GNNs typically operate on a predefined and explicit graph structure, the underlying topological relationships remain static. In contrast, NLNNs dynamically compute a weight for every input instance. This means the "edges" between soil layers are not fixed but are generated on-the-fly based on the current data state. We have added a clarifying statement in Lines 92-96:

“While GNNs rely on explicit, pre-defined graph structures where topological relationships remain fixed and neighbor nodes typically share uniform transformation rules, Non-local Neural Networks (NLNNs) dynamically compute global dependencies (X. Wang et al., 2018). In NLNNs, interaction weights are nonlinearly determined by the real-time soil moisture state rather than by a static topology.”

Line 100: How generalizable the model will be?

Response16:

Thank you for your comment. We clarify that Non-local Neural Networks (NLNNs) constitute a flexible modeling paradigm that has been widely used across different domains, such as image recognition, and can also be incorporated as a modular component (e.g., similar to a residual block) within larger network architectures (Wang et al., 2018).

In this study, the soil moisture prediction model is designed as a task-specific implementation under the NLNN framework. Therefore, while the underlying NLNN module is general and reusable, we do not claim that the proposed soil moisture prediction model is directly generalizable to other tasks without modification.

We have revised the sentences in Lines96-100:

“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”

Line 115: What “interpretability” are you referring to? What are the new findings that were obtained due to this “interpretability”?

Response17:

Thank you for your comment. We have revised this word into “qualitative description” for better expression.

"interpretability" refers to the ability of our model to provide qualitatively interpretable non-local weights while performing high-accuracy predictions. These weights reflect the interactions among soil moisture at different depths, thereby offering a qualitative description of soil properties.

By analyzing the learned non-local weight maps, we find that the model is able to provide a qualitative characterization of soil property, which differs from traditional machine learning models. For example, in the synthetic case of “sand over loam,” the learned weights correctly reflect the layered soil structure (Section 4.1). These findings demonstrate that our model is not a "black box" but a "grey box" that makes predictions based on learned patterns, enhancing the trustworthiness of deep learning in hydrological applications.

Line 120: “three-dimensional soil column” Did you model the three-dimensional flow problem?

Response18:

Thank you for your comment. We have not solved a full three-dimensional (3D) flow numerical problem. Instead, our model remains a one-dimensional (1D) vertical forecasting framework.

The phrase “three-dimensional soil column” was intended to emphasize that the model captures complex and non-uniform water interactions that occur in real soils, such as macropore flow or fracture-induced bypass flow. Our NLNN framework uses non-local operations to represent these 3D preferential pathways that influence vertical moisture redistribution.

We have revised these sentences to prevent misleading:

“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.”

Line 125: Why is the current model not time series data processing?

Response19:

Thank you for your comment. While many machine learning models aim to identify temporal patterns within long historical sequences, our framework is inspired by the physical modeling paradigm (e.g., the Richards equation). The soil moisture state of the next day is fundamentally determined by the current profile's moisture state and the immediate meteorological forcing. 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.

Therefore, our model treats the task as a state-transition problem rather than a pattern-matching time-series problem. By focusing on a single-time-step input of multi-depth variables, we emphasize the inter-layer interactions. We have provided more descriptions in Lines 124-127:

“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.”

Line 130: What are the physical meanings of the weights and how will they help interpret the results?

Response20:

Thank you for your comment. In this study, the non-local weights learned by the NLNN model are primarily used for qualitative interpretation. Each weight represents the relative influence of soil moisture at one depth on the moisture at another depth in the subsequent time step, thereby reflecting vertical soil water interactions. By visualizing these learned weights, the model results can be interpreted in terms of interactions within the soil profile and can provide qualitative insights into soil properties (such as layered structure). We have revised the sentences in Lines 130-134:

“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.”

Line 135: Again, the citation format should be corrected

Response21:

We have revised it.

Figure 1: I would suggest providing the full name of “SM” before using the acronym

Response22:

Thank you for your comment. We have added the full name “soil moisture (SM)” in Abstract before using the acronym.

Line 150: Should it be ”manuscript” or “paper”?

Response23:

We have revised it into “The remainder of this study is organized as follows:” in Line148.

Line 155: What does it mean by “under the assumptions that preferential flow”?

Response24:

Thank you for your comment. We have revised this sentence into “and assuming the absence of preferential flow”.

Line 170: The assumption appears the opposite of the statement in Line 90. Can you clarify which one should be the case?

Response25:

Thank you for your comment. This belongs to the same assumption, which is the soil moisture content status of the previous day's soil profile and the soil moisture content of the next day's weather forecast profile.

Line 180: Are both methods new? Please clarify.

Response26:

Thank you for your comment. SA-NLNN is not a fundamentally new architecture; rather, it adapts the established self-attention mechanism (widely used in Transformers) to a new application (soil moisture prediction). In contrast, KG-NLNN is a novel architecture proposed in this study. It modifies the standard non-local operation by decoupling it into four components, each designed to represent an aspect of soil water transport. We have clarified the novelty in the revised manuscript: “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.”

Line 185: The predictions did not include “ sm_{t+1} ” in Figure 2. Additionally, the vector size of the predictions is not consistent with that of the ground truth.

Response27:

Thank you for your comment. We have revised these descriptions here.

“This yields prediction denoted as, $\mathbf{sm}^{t+1'} = [sm_1^{t+1'}, sm_2^{t+1'}, \dots, sm_{n-1}^{t+1'}, sm_n^{t+1}']^T$.”

Figure 2: Was SA score, KG score done separately or together? The current plot appears to show they are two comparts within one framework.

Response28:

Thank you for your comment. The SA score and KG score were calculated separately as they belong to two different models: SA-NLNN and KG-NLNN, respectively. Specifically, they represent the outputs of two different non-local operations implemented under the same general NLNN framework. The current plot combines them to provide a visual demonstration that they belong to the same framework while employing different non-local operations. To make this clearer, we have updated Figure 2 to explicitly distinguish between the two models.

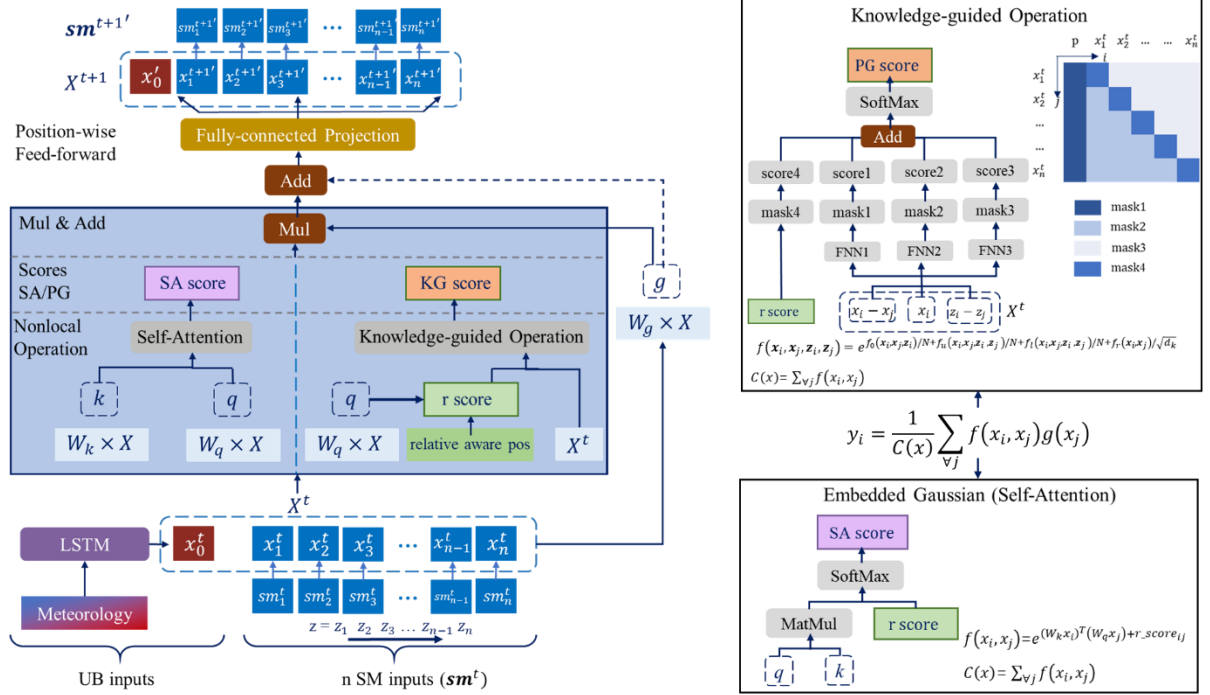


Figure R1(Figure 2 in the manuscript). Left: non-local neural network structure for soil moisture forecasting. Right: embedded Gaussian operation and knowledge-guided non-local operation. RPE: relative position encoding. SA/KG score: non-local weights computed through embedded Gaussian operation and knowledge-guided operation. W_q , W_k and W_g are the weight matrixes to be learned for embeddings.

Figure 2: For the fully connected NN, what are the dimensions of nodes and layers?

Response29:

Thank you for your comment. This module (Fully-connected NN) in Figure 2 is a single-layer linear transformation (depth = 1) without a non-linear activation function. Its primary role is to project the concatenated features from multiple heads back into the target feature space.

Node Dimensions: The output dimension is determined by the product of the embedding dimension and the number of heads. Specifically, the dimension is $d_g \times n_{head}$, where d_g represents the feature dimension of the function g (value space) and n_{head} is the number of operation heads (Line 340). This design project the concatenated multi-head outputs back to the original feature space.

To avoid potential confusion, we have renamed this component to "fully-connected projection" in the revised manuscript in the updated Figure 2.

Equation (2): Please define y_i .

Response30:

Thank you for your comment. We have clarified the definition of in the manuscript in Line198:
“The term y_i denotes the i_{th} component of the output y .”

Line 210: Does the weight matrix W_g depend on the number of locations being examined and their specific depths? In other words, will the W_g be directly applied to other problems?

Response31:

Thank you for your comment. To clarify, the weight matrix W_g is implemented as a point-wise linear transformation. Regarding dimensions: The size of W_g depends only on the channel dimensions ($C_{in} = 1, d_g = 16, n_{head}=10$) and is independent of the number of locations. Thus, the model's architecture can handle inputs of varying number of depths.

Regarding generalization: Although the architecture is flexible, the weights of W_g in this study are specifically optimized for a particular station to capture its unique temporal/spatial patterns. While the framework is directly applicable to other problems, the learned parameters are task-specific, reflecting the distinct data distributions of different stations.

Line 215: “It is evident that xxx”: NLNN reads like a transformer, please elaborate the difference.

Response32:

Thank you for your comment. The NLNN is a general framework, and the self-attention mechanism in Transformers can be regarded as a special case of a NLNN, as already clarified in the SA-NLNN section in Lines221-225. The proposed KG-NLNN further builds upon this framework through task-specific modifications tailored to soil water modeling. Therefore, KG-NLNN represents a customized, task-driven architectural extension. Overall, both KG-NLNN and the Transformer-based model (SA-NLNN) can be regarded as members of the NLNN family.

Line 240-250: Is the r_score new? If not, I would either clarify that in the main text or move them to the SI.

Response33:

Thank you for your comment. The r_score is an internal score used within the non-local operation to model relative positional effects, which is conceptually needed for the SA score and KG score computing in the model. It is an integral part of the model architecture and is required to compute non-local weights. Therefore, it is neither intended as a performance measure nor suitable to be moved to the Supporting Information.

To avoid misunderstanding, we have revised the manuscript to clarify the role of r_score in Line241:
“Line 240-250: Is the r_score new? If not, I would either clarify that in the main text or move them to the SI.”

Line 255: “The four masks in Figure 2 ...” Is this conceptually similar to PCA?

Response34:

We would like to clarify the role of the four masks in Figure 2. Each mask is used to partition the computed attention matrix into four parts, with each part processed using a distinct computation method before being combined to produce the final attention output. This mechanism is not

conceptually equivalent to PCA-based decomposition. Unlike PCA, which performs linear orthogonal projections for variance-based feature decomposition, our masks are structural partitions within the attention computation, designed to enable different calculation strategies for different subsets of the attention matrix. The goal is computational disentanglement, not statistical dimensionality reduction.

Line 255: “Each of these components ...” looks not consistent with the figure

Response35:

Thank you for your comments. We have revised the text to clarify that only meteorological forcing, upper soil water influence, and lower soil water interactions are modeled using FNNs, while the same-depth soil moisture effects are described by relative positional encoding in Lines255-260, making the description consistent with Figure 2.

“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.

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.”

Line 270: Can the two layers be defined as “deep” learning?

Response36:

Thank you for your comments. We agree that a standalone two-layer fully connected network does not constitute "deep" learning in the traditional sense. However, we would like to clarify that these two layers are not the entirety of our model; they represent only a specific functional module designed to qualitatively estimate interaction intensity (the "masks") as a component to the non-local operations.

The overall framework, which integrates these modules with non-local neural networks (NLNNs), operates as an end-to-end deep learning architecture. We intentionally kept this specific module shallow (two layers) to prevent an excessive increase in model complexity and to ensure trainability. Using a deeper sub-network at this stage would significantly complicate the optimization process and lead to diminishing returns in predictive accuracy.

Line 270: Commonly used in your research may not justify the decision. I would add references to support your choice.

Response37:

In the revised manuscript, we have updated the text to provide support for our choice of a two-layer feedforward neural network (FNN) and revised the sentences in Lines274-277:

“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.”

Specifically, we reference the universal approximation theorem (Cybenko, 1989), which provides a theoretical guarantee that that a feedforward neural network with a single hidden layer can approximate a wide range of nonlinear functions when non-polynomial activation functions such as the hyperbolic tangent are used. Building upon this theoretical foundation, a two-layer FNN is

adopted in this study as a practical design choice to achieve an appropriate balance between model expressiveness and computational efficiency.

Line 300: Can you justify the reason for such selection?

Response38:

Thank you for your comment. The six selected meteorological variables—precipitation (P), air temperature (AT), long-wave radiation (LR), short-wave radiation (SR), relative humidity (RH), and wind speed (WS)—were chosen because they are the primary drivers of soil water balance processes. Specifically, Precipitation (P) represents the direct water input to the soil system, while Air temperature (AT), long-wave radiation (LR), short-wave radiation (SR), relative humidity (RH), and wind speed (WS) jointly control potential evapotranspiration, which is the main pathway of soil water loss.

These variables are therefore widely adopted in hydrological and soil moisture modeling studies, as they effectively capture the dominant physical mechanisms governing infiltration and evapotranspiration.

Line 330: Can you show the training loss and validation loss over epochs?

Response39:

Thank you for your comment. Here we have provided the training loss and validation loss for the proposed NLNN models.

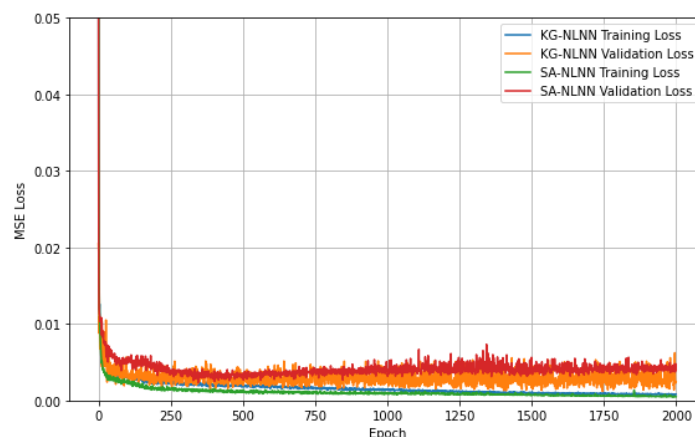


Figure R2. The training loss and validation loss for the KG-NLNN and SA-NLNN models.

We provide the training and validation loss curves for both KG-NLNN and SA-NLNN in the manuscript. As shown in the figure, the training losses of both models decrease rapidly at the beginning of training and then stabilize. Notably, the training loss of SA-NLNN decreases faster than that of KG-NLNN. The validation losses of both models also decrease initially, with a slight increase observed in the later stages of training. Overall, the validation loss of KG-NLNN remains lower than that of SA-NLNN throughout the training process.

Line 425: “The color brightness on the ...” Can you explain how this can be understood conceptually?

Response40:

Thank you for your comment. Conceptually, the color brightness represents the relative contribution of soil moisture at different depths from the previous day to the predicted soil moisture at a given

depth on the next day. According to Equation (2), the prediction y_i is computed as a weighted aggregation of information from all depths j , where the normalized weights $\frac{1}{c(x)}f(x_i, x_j)$ quantify the influence of the j -th depth on the i -th prediction. A higher color brightness therefore indicates a larger influence coefficient, meaning that soil moisture at that depth plays a more important role in determining the predicted value.

$$y_i = \frac{1}{c(x)} \sum_{x_j} f(x_i, x_j)g(x_j) \quad (2)$$

We have added the descriptions in Lines430-434:

“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 $\frac{1}{c(x)}f(x_i, x_j)$ 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.”

Line 435: Why LSTM_1 is introduced after LSTM_4? How about how about 2 and 3? I would suggest using a more descriptive naming instead of using numbers.

Response41:

Thank you for your comment. LSTM_4 is introduced first because it represents the standard usage of LSTM models for soil moisture prediction, where temporal dependencies are modeled using a fixed look-back window. This configuration has been widely adopted in previous studies and thus serves as a conventional baseline.

To avoid potential confusion regarding the numbering, we have revised the manuscript by adopting more descriptive model names and reorganizing the presentation accordingly. Specifically, LSTM_4 is now referred to as LSTM_T, which utilizes input data from the previous four time steps to predict soil moisture content, focusing 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.

“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.”

Figure 7: What do the weights represent?

Response42:

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. The elements at position (i,j) represent the impact of soil moisture at depth z_j at the previous time on soil moisture content at depth 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. We have added the descriptions in Lines 467-472.

Line 460: “Figure 7 depicts the weight matrix maps ...” How about the homogeneous and heterogeneous cases?

Response43:

Thank you for your comments. We have generated the weight matrix maps for both homogeneous and heterogeneous soil cases as requested in Figure R3. However, the discussion of these cases is less informative compared to the layer-based examples presented in the manuscript. Therefore, we do not include these cases in the main discussion of the manuscript.

For homogeneous soils, the weight maps indicate that soil moisture at a given depth primarily depends on the previous day's soil moisture at the same depth as well as contributions from the upper layers. For heterogeneous soils, the weight maps are less regular than those for homogeneous soils, which may be related to the randomness in the soil hydraulic conductivity (Ks) used in generating the examples.

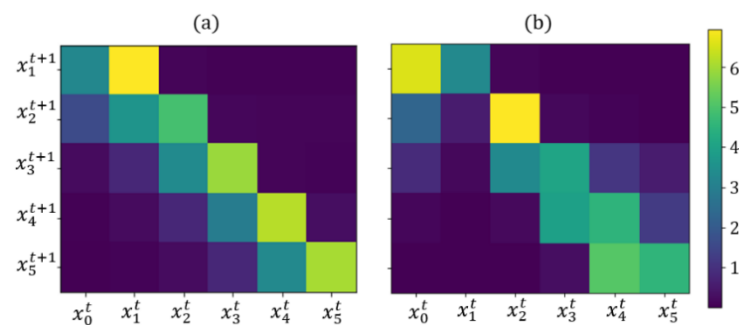


Figure R3. The non-local weight maps in two-layered simulated stratified soil scenarios through KG-NLNN (a) homogeneous soil (b) heterogeneous soil.

The two-layered scenarios, together with homogeneous soil cases under different Ks scenarios, already clearly demonstrate the model's interpretability and its ability to capture soil water interactions across depths. Therefore, we have kept the original discussion focused on these examples, which provide the most insightful illustration of model behavior.

Line 470: What case is being discussed? What is the difference between Figure 7 and Figure 8.

Response44:

Thank you for your comments. We have revised this section according to response 46.

Figure 7 focuses on homogeneous soil scenarios with different saturated hydraulic conductivity (Ks)

values and is intended to explore whether the KG-NLNN model can qualitatively characterize soil properties. It demonstrates that KG-NLNN can qualitatively represent variations in soil hydraulic conductivity. However, this representation provides qualitative insight only and does not imply a direct quantitative correspondence with actual soil hydraulic parameters.

Figure 8, on the other hand, now presents the weight visualizations of the KG-NLNN and SA-NLNN models for the layer-based example to further investigate their ability to describe soil structure. These results show that KG-NLNN produces stable weights that qualitatively reflect soil structure, whereas the weights generated by SA-NLNN are highly unstable and fail to provide meaningful physical interpretation.

Line 480: “the weight map of the SA-NLNN model appears slightly chaotic” Can you perform a test on more layers, with increasing or decreasing K_s to see if this trend holds?

Response45:

Thank you for your comment. We explicitly provide the non-local weight matrices generated by the SA-NLNN model, as shown in the corresponding figures. We examined multiple homogeneous soil scenarios with both increasing and decreasing saturated hydraulic conductivity (K_s) values. The results indicate that, despite changes in K_s , the non-local weight maps produced by SA-NLNN remain irregular and lack stable spatial patterns across different depths. This behavior suggests that the observed instability of the SA-NLNN weight maps is not case-specific but persists across different soil hydraulic conditions. This indicates that a knowledge-guided structural design can serve as a valuable enhancement.

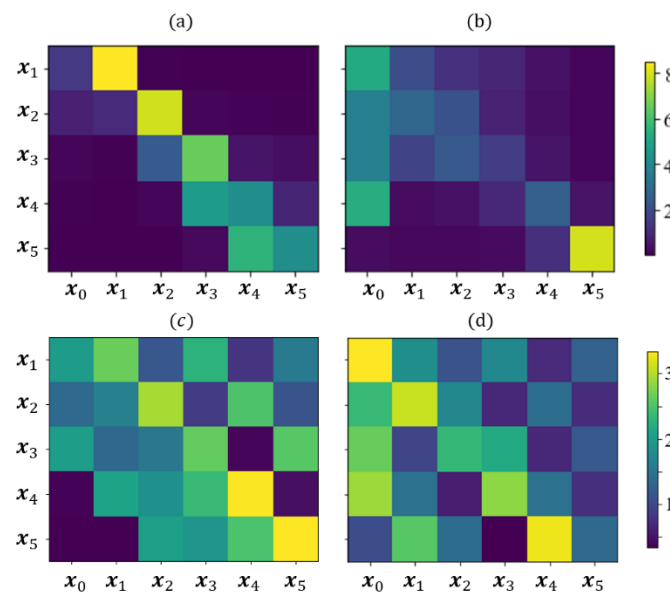


Figure R4(Figure 7 in the manuscript). The non-local weight maps in homogeneous simulated soil scenarios through KG-NLNN (a) $K_s = 0.25$ (b) $K_s = 10.49$, and SA-NLNN (c) $K_s = 0.25$ (d) $K_s = 10.49$.

Line 490: What is the reason that the homogeneous case was shown first in figure 4, but here was discussed after the two-layer case?

Response46:

Thank you for your valuable suggestion. We agree that the presentation order could be confusing,

and we will reorganize this section to improve the flow, ensuring that the discussion order of the homogeneous and two-layer cases is consistent and more intuitive in Lines 465-500:

“

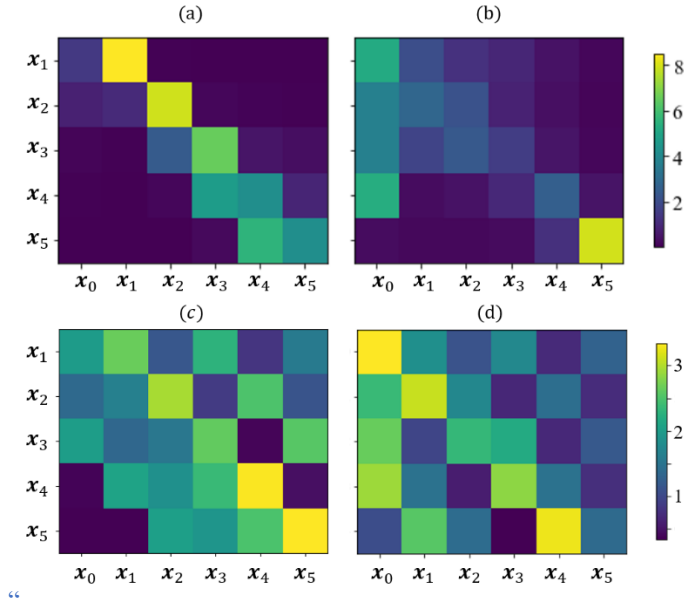


Figure 7. The non-local weight maps in homogeneous simulated soil scenarios through KG-NLNN (a) $K_s = 0.25$ (b) $K_s = 10.49$, and SA-NLNN (c) $K_s = 0.25$ (d) $K_s = 10.49$

Figure 7 depicts the weight matrix maps generated by KG-NLNN and SA-NLNN models for homogeneous soil scenarios varying saturated hydraulic conductivity (K_s) 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 z_j at the previous time on the soil moisture content at depth 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 K_s 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 K_s , 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 ($K_s = 0.25$) exhibits slow infiltration, so its moisture content is easily influenced by adjacent layers in Figure 7(a). In contrast, sand ($K_s=10.49$) 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.

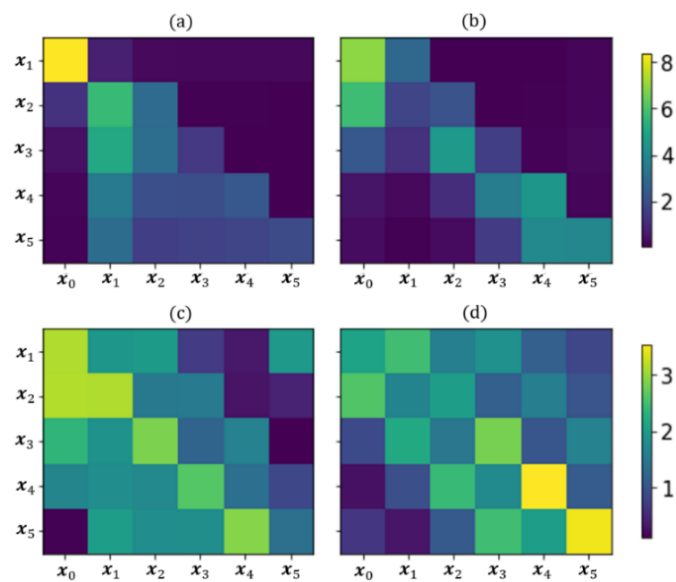


Figure 8. The non-local weight maps in two-layered simulated stratified soil scenarios through KG-NLNN (a) loam above sand (b) sand above loam, and SA-NLNN (c) loam above sand (d) sand above loam. ”

Line 510: Why is the comparison challenging? Can you elaborate the reasoning?

Response47:

The comparison between the proposed ML-based framework and conventional physical models (e.g., Ross model with EnKF) is inherently challenging due to differences in their modeling paradigms and data requirements:

Optimization Complexity: In physical models like Ross, typically only a subset of soil hydraulic parameters is optimized, while boundary conditions and complex hydraulic functions are difficult to calibrate simultaneously using traditional numerical methods. In contrast, neural networks inherently optimize a high-dimensional parameter space, including latent representations of boundary conditions, leading to higher precision in localized contexts.

Accuracy vs. Stability: Data-driven models excel at capturing fine-grained, short-term fluctuations, providing superior precision within the training distribution. However, they are prone to error

accumulation in long-term autoregressive forecasting. Conversely, physical models provide lower precision but maintain long-term stability and physical consistency.

Data Dependency: The ML framework requires more training data, whereas physical models rely on accurate soil property measurements and theoretical assumptions.

Overall, we contend that a perfectly comparison between machine learning frameworks and physics-based models remains inherently challenging due to their disparate data requirements and distinct operational strengths. To address this, Appendix B presents the most rigorous and equitable comparison possible under current conditions, ensuring that both paradigms are evaluated on as level a playing field as their fundamental differences allow.

Figure 9: Are the RMSE average across both depths and sites?

Response48:

We clarify that the RMSE values shown in Figure 9 are averaged across the twenty research sites for each specific depth. It is not a single average across all depths. To avoid confusion, we have revised the figure caption to explicitly state the averaging method.

“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).”

Table 3: Why were some numbers bold?

Response49:

The bold values in Table 3 indicate the best predictive performance (the lowest MAE) among all compared models for each specific metric and depth. We have added a clarifying sentence to the table caption to ensure this is clear to the reader.

“Table 3. 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.”

Figure 11: Can you label the sites on the figures?

Response50:

Thank you for this suggestion. To improve the readability of the time-series plots, we have now added site labels directly onto the figure panels. Specifically, we labeled each group of soil depths with their corresponding site names (Falkenberg, Cape-Charles, and Goodwell) at the top of each row

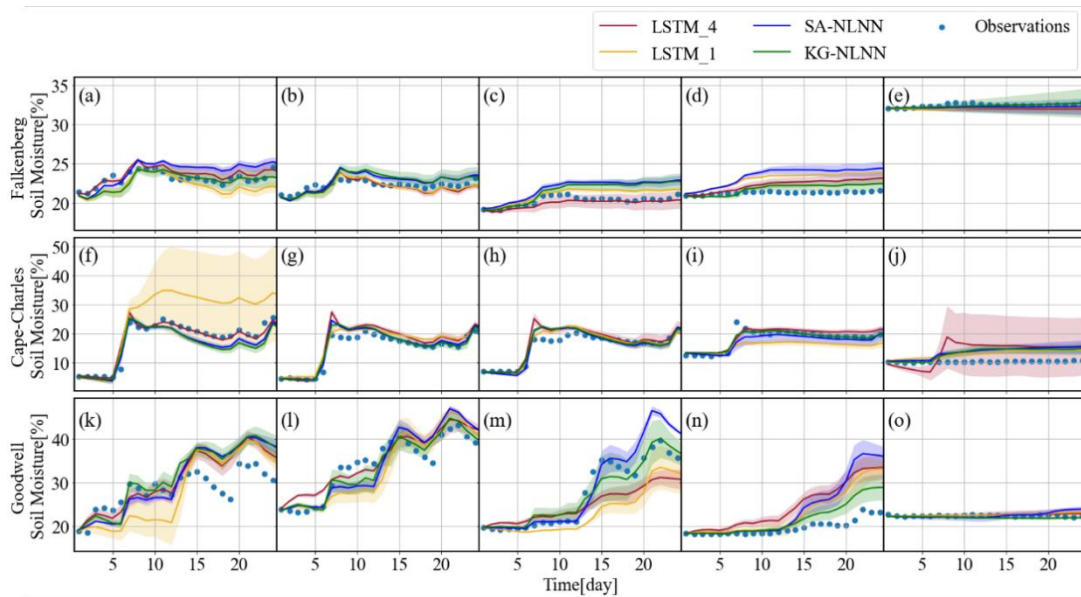


Figure R5 (Figure 11 in the manuscript). The autoregressive 24-day predicted soil moisture time series of 5 depths with LSTM_1, LSTM_4, KG-NLNN and SA-NLNN at Falkenberg (a-e), Cape-Charles (f-j), and Goodwell (k-o). The shaded region represents the confidence interval of the models, spanning 1 standard deviation.

Line 590: Can you quantify the accuracy?

Response51:

To quantify the impact of time intervals on model accuracy, we have added a new analysis (see Figure R6) showing the trend of RMSE as the time interval increases.

Our analysis reveals that model error grows with larger time intervals. This degradation in accuracy is primarily attributed to the data aggregation effect: at larger time scales, soil moisture data are typically processed as daily or weekly means. This averaging process smoothes out the temporal fluctuations (e.g., rapid responses to precipitation events) that are present in high-frequency data. Consequently, the model loses the high-resolution information necessary for precise prediction, leading to increased RMSE.

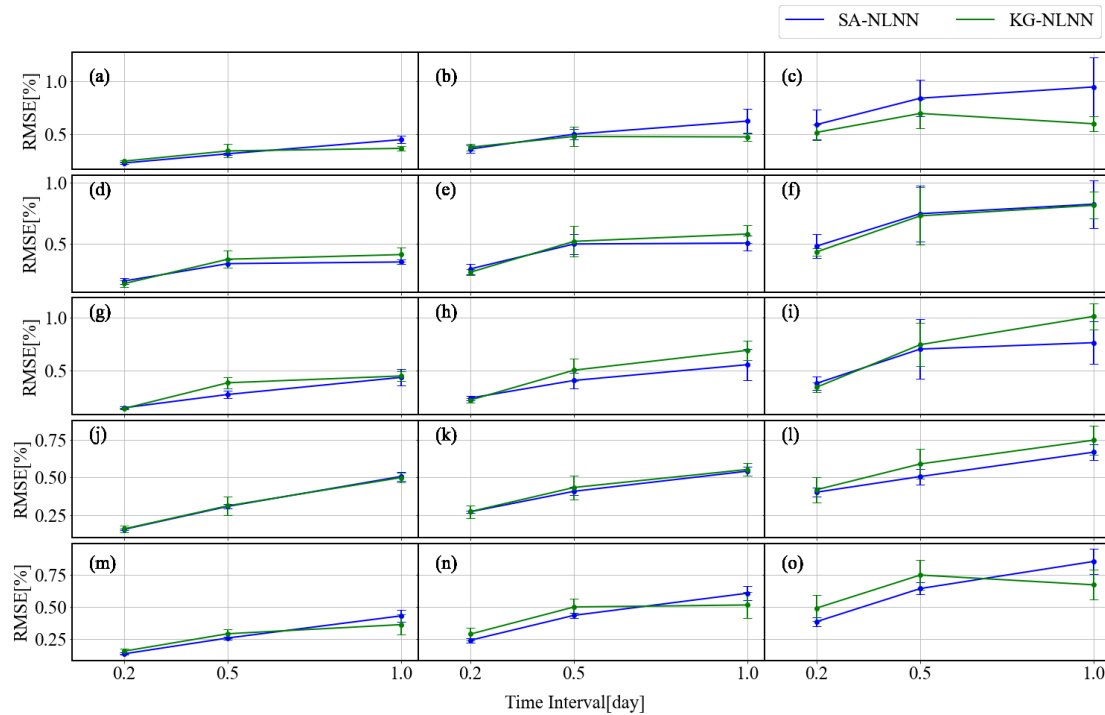


Figure R6. The RMSE results for 1, 3, 7-day at 0.05m(a-c), 0.10m(d-f), 0.20m(g-i), 0.50m(j-l) and 1.0m(m-o) in the homogenous soil under increasing time intervals. The error bar indicates the standard deviations of the RMSE, which are computed via ten training replicates.

Figure 13: Why were some red curves cut off?

Response52:

Thank you for your comment. The red curves appear cut off because we fixed the y-axis scale to better visualize the performance differences among the higher-performing models (LSTM_I, SA-NLNN and KG-NLNN). The error metrics for the baseline in those specific periods exceeded the upper limit of the y-axis, indicating significantly lower predictive accuracy. Expanding the y-axis to include the full red curves would have compressed the other models' results, making it difficult to distinguish their relative improvements. We have added a note to the figure caption to clarify this choice.

“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.”

Line 650: “What’s more, the proposed network framework is flexible and easily customizable to suit specific requirements,” is not consistent with the fact that the model is “site-specific”. Can you elaborate on your reasoning?

Response53:

Thank you for your comment. To address the reviewer's concern, we have We have clarified that the claimed flexibility and customizability refer to the NLNN framework itself, rather than to the direct transferability of a trained model across sites. The NLNN framework is designed to be flexible and

task-adaptive, allowing users to incorporate different structural designs or domain knowledge according to specific applications. Within this general framework, we developed the KG-NLNN architecture specifically for soil moisture prediction. We have revised the text to emphasize how this architectural design can be adapted for broader applications, moving beyond site-specific soil moisture tasks. We have revised these sentences here:

“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.”

Referee #3

I thank the authors for their thorough responses and the substantial revisions. Overall, my main concerns have been addressed, especially regarding the consistency of the benchmarking across synthetic and real-world experiments and the clarity of the experimental setup. Thus, I recommend minor revision, and the revised manuscript is substantially improved and, after the small adjustments below, should be suitable for publication.

Response:

We sincerely thank the reviewer for their time and careful evaluation of our manuscript. We have carefully considered the remaining minor comments and have made the corresponding adjustments accordingly.

Remaining issues (minor)

1. Interpretability claims vs. quantification: The authors provide convincing qualitative evidence that the inferred weight maps behave sensibly under controlled changes (e.g., different soil hydraulic properties). However, the manuscript still does not provide a quantitative link between interpretability outputs and physical properties. If the intent is not to infer parameters quantitatively, I recommend tightening the wording to clearly frame interpretability as qualitative (or add a simple quantitative association analysis if feasible).

Response1:

Thank you for your comment. Following your advice, we have carefully revised the manuscript to frame the interpretability of KG-NLNN as a qualitative diagnostic tool. We have thoroughly revised the manuscript to clarify this distinction. This clarification ensures that readers do not misinterpret the non-local weights as direct numerical substitutes for soil hydraulic parameters.

2. Please ensure that “physics-guided” phrasing (or equivalent) is fully removed throughout the manuscript (including abstract-like sections and conclusions), to avoid mixed messaging

Response2:

Thank you for your comment. We have systematically removed and replaced these phrases throughout the entire manuscript to ensure clarity and avoid mixed messaging.

Reference:

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. *Math. Control. signals Syst. 2*, 303–314.

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

Liu, Y., Mei, L., Ki, S.O., 2014. Prediction of soil moisture based on Extreme Learning Machine for an apple orchard. *CCIS 2014 - Proc. 2014 IEEE 3rd Int. Conf. Cloud Comput. Intell. Syst.* 400–404. <https://doi.org/10.1109/CCIS.2014.7175768>

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.

