This manuscript presents an interesting and exploratory study addressing the critical challenge of soil moisture prediction within the hydrological cycle. Specifically, the authors developed two variants of non-local neural networks (NLNN), and the method effectively models vertical heterogeneity and inter-layer connectivity. The authors validated their approach using synthetic data and in-situ observations from the International Soil Moisture Network, highlighting the models' interpretability and their robustness to noise. This approach, which attempts to learn soil water dynamics directly from data without relying on traditional physical assumptions, demonstrates the potential of integrating data-driven techniques with scientific knowledge, particularly in complex soil conditions such as wormholes and root water uptake.

Response:

We are very grateful to receive such valuable comments and suggestions provided by the reviewer regarding our manuscript. We have attempted to address the comments and revised our manuscript.

However, despite the innovative research direction, the manuscript suffers from several issues, particularly in terms of the clarity of the methods. My specific comments are as follows:

1. The claim of a "physics-guided" approach is a core theme, but the method lacks evidence of incorporating actual physical laws into the model structure or loss functions. The model constructed in this study essentially relies on data-driven feature interaction learning. The authors should either replace the term "physics-guided" with a more appropriate expression or explicitly incorporate more concrete physical information into the model.

Response1:

Thank you for your comments. We agree that our model is not driven by explicit physical equations or mechanisms. Instead, its design is guided by fundamental knowledge of soil water dynamics. Therefore, we have replaced the term "physics-guided" with the more accurate descriptor "knowledge-guided" throughout the manuscript to better reflect this approach.

2. The authors mention various deep learning models in the introduction, including CNN, LSTM, Transformer, etc., and state that "the complex coupling of actual physical processes and the presence of unknown governing equations pose substantial challenges in practical applications," based on which they propose the use of NLNNs in this paper. However, Graph Neural Networks (GNNs) can also incorporate spatial relationships into the model. The authors should expand the literature review to more clearly highlight the uniqueness of their approach.

Response2:

Thank you for your comments. We provided a detailed discussion of the differences between graph neural networks and our method NLNNs, highlighting NLNNs' relative advantages:

"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). GNNs rely on explicit, pre-defined graph structures, where neighbor nodes typically share the same transformation rules and the topological relationships remain fixed. In contrast, the Non-local Neural Networks (NLNNs) dynamically compute global dependencies (X.

Wang et al., 2018). Essentially, 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). The inherent design enables NLNNs to flexibly model relationships between variables according to the requirements, functioning like a more flexible GNN on a fully-connected graph. 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."

3. The authors include some conclusive statements in the introduction (e.g., "By integrating meteorological conditions and the spatial interactions of soil moisture within its four-part disentangled physics-guided operation framework, PG-NLNN demonstrates superior performance"). The introduction should mainly address the research background, motivation, scientific problems, and research content. Including such statements is useful for emphasizing the potential contribution of the paper, but it would be more appropriate to place them in the results or conclusion sections.

Response3:

Thank you for your comments. We have revised the introduction by removing these conclusive claims. The relevant content has been moved and integrated into the Discussion and Conclusion sections to provide a more appropriate interpretation of the findings.

4. The time dependency assumption presented in Section 2.1 needs further explanation. The sentence, "In our soil moisture forecasts at multiple depths, 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," suggests that the soil moisture at time $t1$ and the meteorological conditions at time $t2$ determine the soil moisture at time $t3$. However, it is unclear why the soil moisture at time $t2$ does not influence the soil moisture at time $t3$. Please clarify the time dependency and provide the theoretical foundation for this assumption.

Response4:

Thank you for your comments. To further explain why we design model this way, we have added the physical background and theoretical foundation of the soil water transport equation in the manuscript in section 2.1:

"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 under the assumptions that preferential flow, this equation takes the following form:

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

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."

5. The input symbol in Figure 2 should be consistent. The figure uses Xt, while the text uses smt, which creates an inconsistency. It is recommended to unify the notation for better understanding for readers. Additionally, elements like Wk, Wq, and Wv in the figure are not fully explained. It would be helpful to expand the caption to provide more detailed descriptions, aiding readers in better understanding the model structure.

Response5:

Thank you for your comments. We have updated Figure 2 by integrating the correspondence between X^t and sm^t to ensure uniformity. Besides, we have added more details about Wk, Wq, and Wg in the caption as follows:

"

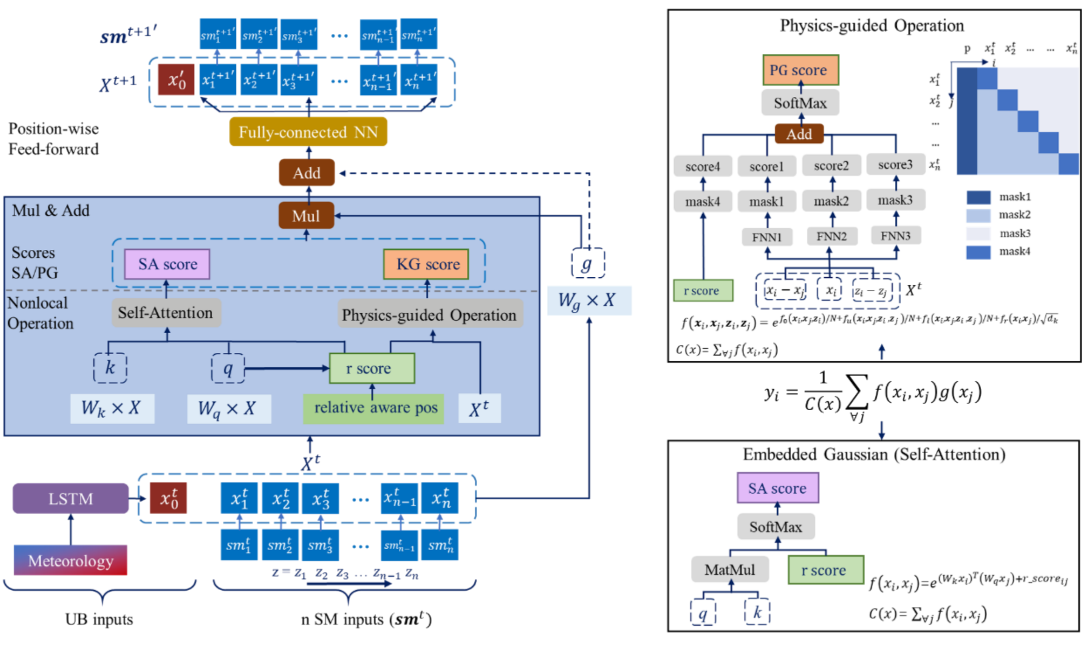

Figure 2. Left: non-local neural network structure for soil moisture forecasting. Right: embedded Gaussian operation and knowledge-guided non-local operation. RPE: relative position encoding. SA/ PG score: non-local weights computed through embedded Gaussian operation and knowledge-

guided operation. W_q, W_k and W_g are the weight matrixes to be learned for input embeddings."

6. The model's Physics-guided Operation section utilizes different mask layers to filter out specific data points while emphasizing useful information. This is similar to the self-attention mechanism in Transformers but with domain-specific adjustments. However, the authors have not provided the physical basis for this mechanism, nor have they discussed its advantages compared to the traditional Transformer mechanism. It is recommended to include the theoretical background of this mechanism and explain its specific advantages for soil moisture prediction to enhance the persuasiveness and originality of the work.

Response6:

Thank you for your comments.

The vertical movement of soil water exhibits directional divergence: downward movement is dominated by gravity, as a dissipation of potential energy, while upward movement relies on capillary forces and other mechanisms that work against gravity. On this physical basis, the soil structure (e.g., stratification and cracks) collectively regulate the distribution of moisture. All these moisture interactions are intricately related to the soil moisture content and its vertical depth.

In this specific context, we employ different masks to decouple the soil moisture interactions from different directions. The four masks in Figure 2 are designed for meteorological forcing, upper soil water interactions, same-depth soil moisture interactions, and lower soil water interactions, respectively. Each process is modeled by a fully connected network that takes soil moisture content and depth differences as inputs, ultimately forming the final model. Unlike the self-attention mechanism, this knowledge-guided design separates the significantly different moisture movement processes for independent learning, thereby better capturing the relationships between soil moisture variables. In the future, we will develop more suitable and refined solutions tailored to the specific characteristics of soil water movement.

We will incorporate this paragraph into section 2.2.2 of the manuscript:

"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 interactions, same-depth soil moisture interactions, and lower soil water interactions, respectively. Each of these components is modeled by a fully connected network, which takes soil moisture content and depth differences as inputs. 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."

7. In line 249 and in formulas (7), (8), and (9), $f(xi, xj)$ involves two variables, but the subsequent description mentions that the function is a mapping of three variables. The equation should either be modified or the description clarified to specify the actual number of input variables, ensuring consistency between the equation and the text.

Response7:

Thank you for your comments. We have modified the variables in formulas 7-9 and 12-13:

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

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

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

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

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

8. In lines 276–277, the authors mention that the LSTM input contains data from two time steps due to the delayed effect of meteorology on soil moisture. However, the authors do not explain why two time steps were selected or provide any physical or empirical justification. It is recommended to add the rationale and reasoning for this time window choice.

Response8:

Thank you for your comments. The ISMN soil moisture data in this study are derived from 24-hour averages, while precipitation data represent 24-hour cumulative totals. Hydrologically, meteorological conditions from the previous time step (t-1) do not cease their influence immediately at the end of the period. Instead, through processes such as infiltration, lateral flow, and redistribution, they continue to affect soil moisture at the following time step t. Thus, incorporating both time steps enables the model to capture cross-day causal relationships. We set the time step to 2, ensuring that the meteorological input remains both concise and informationally rich. Therefore, the task of learning the meteorological time dependencies is delegated to the LSTM network, which also justifies its application in processing boundary conditions. We have added the relevant content at Line of the manuscript:

"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.”

9. Formulas (6) and (14–16) both use the symbol a to represent the activation function, but the authors have not clarified whether the same function is used in both instances. If the activation functions differ, the notation should be distinguished to avoid confusion. Additionally, the explanation of LSTM variables lacks a description of the tanh activation function. It would be beneficial to add this explanation to maintain consistency between the notation and the equation definitions.

Response9:

Thank you for your comments. We have added corresponding subscripts to the activation function symbol "a" to differentiate between the tanh and sigmoid activation functions, represented as a_t and a_s. Besides, tanh in the equations for LSTM has been revised into a_t to maintain consistency:

“The operation of an LSTM can be summarized as follows:

$$\boldsymbol{i}^t = a_s(W_i \cdot [\boldsymbol{h}^{t-1}, \boldsymbol{ub}^t] + b_i) \tag{14}$$

$$\boldsymbol{f}^t = a_s\left(W_f \cdot [\boldsymbol{h}^{t-1}, \boldsymbol{ub}^t] + b_f\right) \tag{15}$$

$$\boldsymbol{o}^t = a_s(W_o \cdot [\boldsymbol{h}^{t-1}, \boldsymbol{ub}^t] + b_o) \tag{16}$$

$$\widetilde{\boldsymbol{C}}^t = a_t(W_c \cdot [\boldsymbol{h}^{t-1}, \boldsymbol{ub}^t] + b_c) \tag{17}$$

$$\boldsymbol{c}^t = \boldsymbol{f}^t \cdot \boldsymbol{c}^{t-1} + \boldsymbol{i}^t \cdot \widetilde{\boldsymbol{C}}^t \tag{18}$$

$$\boldsymbol{h}^t = \boldsymbol{o}^t \cdot a_t(\boldsymbol{c}^t) \tag{19}$$

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.”

10. In Section 3.1, the authors use reference evapotranspiration rather than actual evapotranspiration, which better reflects actual water consumption. It is recommended to clarify the reasoning behind this choice.

Response10:

Thank you for your comments. Actual evapotranspiration (AET) indeed more accurately represents the e soil water consumption. While actual evapotranspiration is intrinsically dependent on soil water content, the creation of virtual scenarios requires an independent and controllable evapotranspiration value, rather than one coupled with dynamic soil moisture conditions. As standardized in the FAO-56 guidelines (Allen et al., 1998), AET is calculated as AET = Kc × PET, where Kc 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 have added the corresponding descriptions in Lines:

“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.”

11. In line 346, the authors list meteorological input variables but do not specify their time and spatial resolutions. It is recommended to provide this information and explain whether the data have been resampled to ensure the completeness and reproducibility of the model input descriptions.

Response11:

Thank you for your comments. The meteorological data we applied in the model input were downloaded at a daily temporal resolution for each specific location based on its latitude and longitude coordinates. We have provided additional clarification in Line:

“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.”

12. It is recommended to adjust the placement of the legends in Figures 6, 7, 11, and 14, especially in Figures 7 and 14, where it would be better to position the legend at the top of the figure to avoid obscuring important elements of the figure.

Response12:

Thank you for your comments. We have modified the legends of Figures 6, 9, 11, and 14 to present the results more clearly. In accordance with another reviewer's comment, we have replaced Figure 7 with a matrix weight map for enhanced illustration of the model's interpretability in section 4.1. Besides, we have provided comparisons with the LSTM models in Figure 6 and Figure14. The figures with updated legends are presented below.

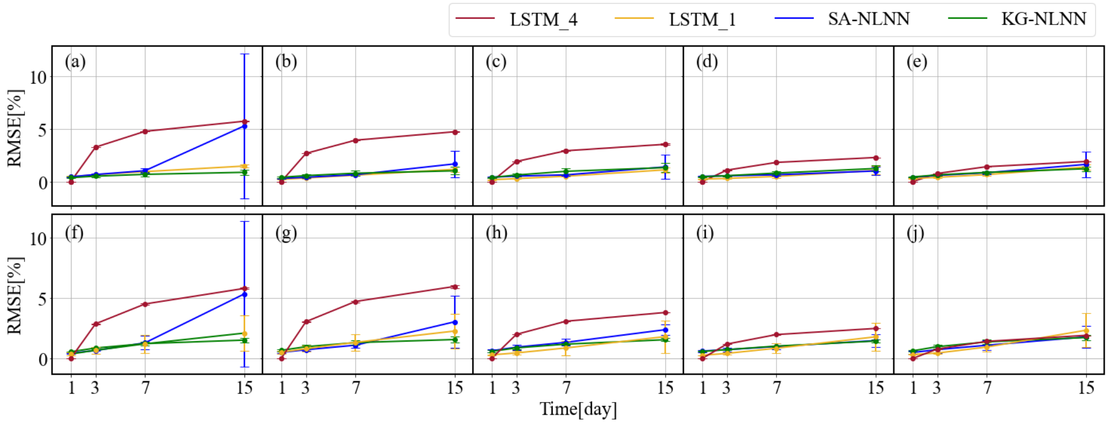

Figure 6. The RMSE results for 1, 3, 7, and 15-day for heterogeneous soil(a-e), and two-layered soil (f-j). The error bar indicates the standard deviations of the RMSE, which are computed via ten training replicates.

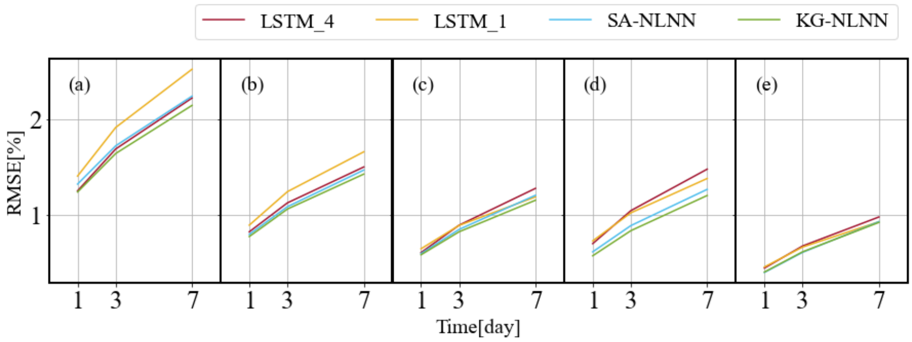

Figure 9. The average RMSE comparisons between LSTM_4, LSTM_1, SA-NLNN, and KG-NLNN across twenty research sites at 5 depths: 0.05m(a), 0.10m(b), 0.20m(c), 0.50m(d), 1.00m(e).

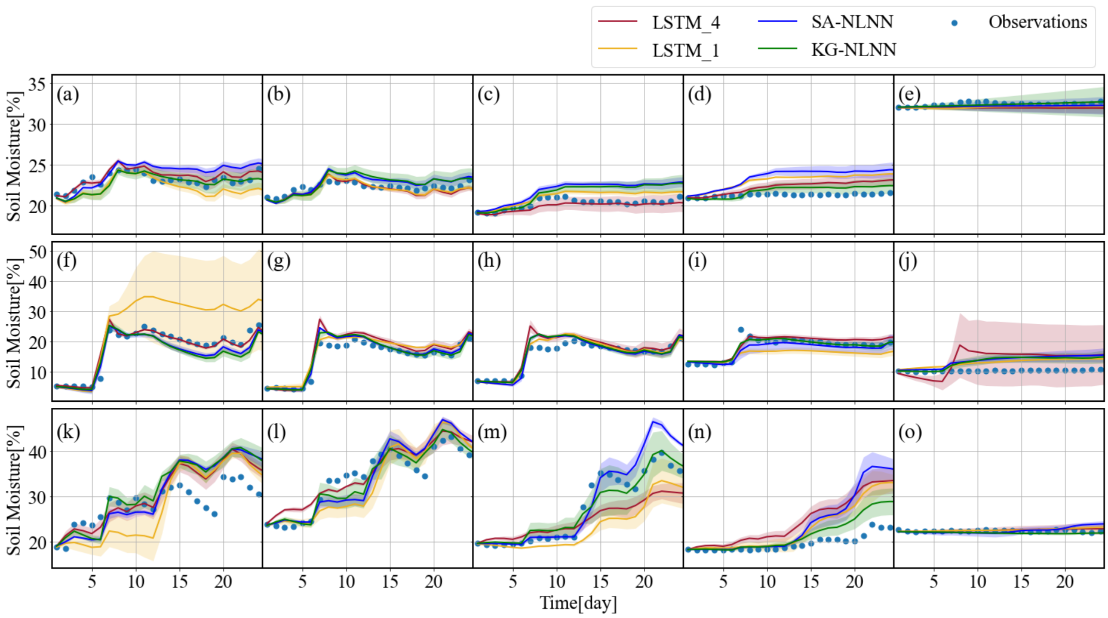

Figure 11. The autoregressive 24-day predicted soil moisture time series of 5 depths with LSTM_1, LSTM_4, KG-NLNN and SA-NLNN at Falkenberg (a-e), Cape-Charles (f-j), and Goodwell (k-o). The shaded region represents the confidence interval of the models, spanning 1 standard deviation.

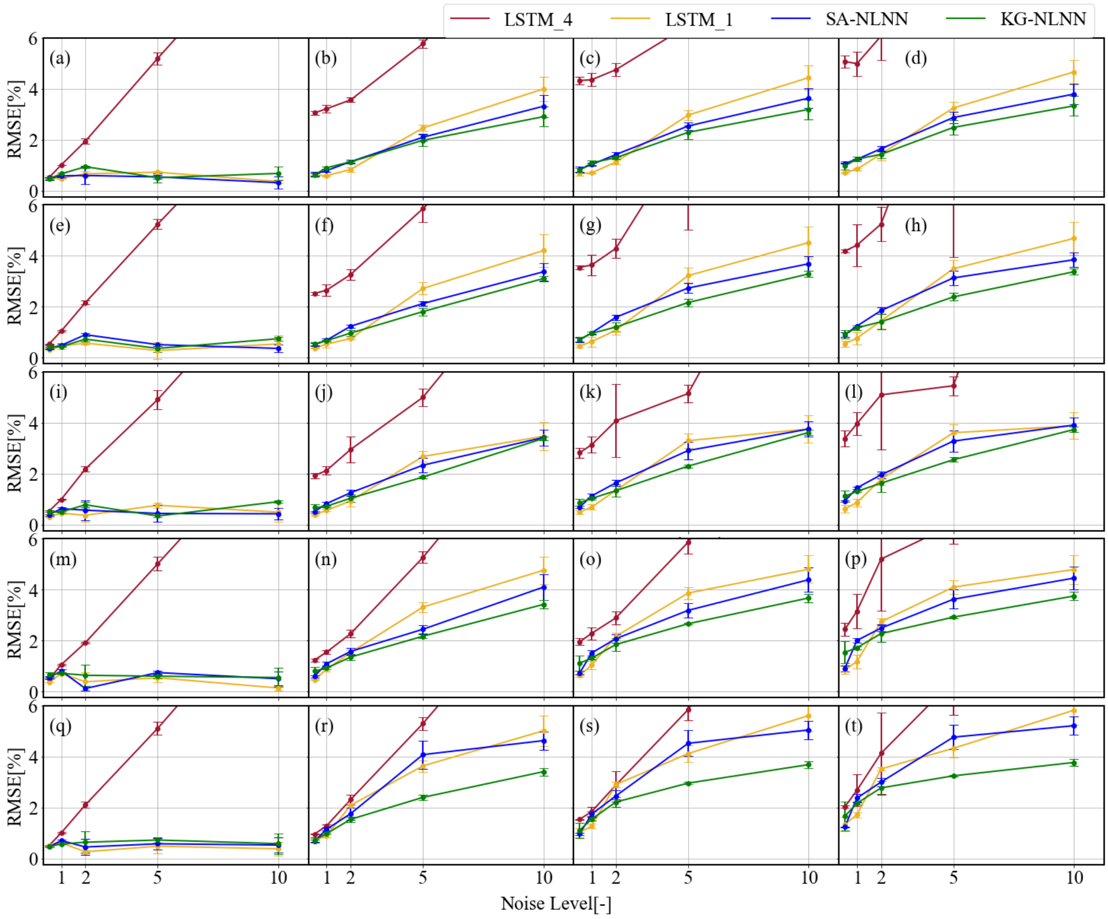

Figure 14. The RMSE results for 1, 3, 7, and 15-day at 0.05m(a-d), 0.10m(e-h), 0.20m(i-l), 0.50m(m-p) and 1.0m(q-t) in the homogenous soil under increasing noise levels through 4 models: SA-NLNN, KG-NLNN, LSTM_1 and LSTM_4. The error bar indicates the standard deviations of the RMSE, which are computed via ten training replicates.

13. In line 409, the term "significant errors" is used, but no statistical support is provided. It is recommended to either include p-values or replace “significant” with terms like “obvious” to avoid using the term without statistical backing.

Response13:

Thank you for your comments. As suggested, we have replaced "significant" with "obvious".

Reference:

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,

D05109.
Buckingham, E., 1907. Studies on the movement of soil moisture.
Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., 2017. Graph attention networks. arXiv Prepr. arXiv1710.10903.
Wang, W., Wei, Y., Hao, L., Wei, Z., Zhao, T., 2025. Soil moisture forecasting in wireless sensor networks via spatiotemporal graph convolutional networks 1–17. https://doi.org/10.1002/vzj2.70000