Computers and Electronics in Agriculture 237 (2025) 110614

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Decoding time: Unraveling the power of N-BEATS and N-HiTS vs. LSTM for accurate soil moisture prediction

Lisa Umutoni [a], Vidya Samadi [a,b,c,*], George Vellidis [d], Charles Privette III [a], Jose Payero [a], Bulent Koc [a]

[a] Agricultural Sciences Department, Clemson University, Clemson, SC, USA
[b] Artificial Intelligence Research Institute for Science and Engineering (AIRISE), School of Computing, Clemson University, Clemson, SC, USA
[c] Division of Civil Engineering, Department of Engineering, University of Cambridge, England, UK
[d] Crop & Soil Sciences Department, UGA, GA, USA

ARTICLE INFO

Keywords:
Soil water tension prediction
Deep neural networks
Uncertainty estimation
Irrigation decision making

ABSTRACT

Deep neural networks (DNNs) can be trained to predict soil moisture dynamics, which is crucial for effective irrigation scheduling. However, a lack of interpretability in these networks constrains their efficacy in grasping the nuanced patterns prevalent in soil moisture time series data. This study is the first attempt known to the authors that develop interpretable DNNs to predict soil moisture fluctuations expressed as soil water tension across three root zone depths and prediction horizons. The Neural Hierarchical Interpolation for Time Series (N-HiTS) and Neural Basis Expansion Analysis Time Series (N-BEATS) models were used in this research. Historical soil water tension data collected at the University of Georgia's C. M. Stripling Irrigation Research Park (SIRP) and in Blackville, South Carolina, were used to train and test the models. The results were benchmarked with the Long-Short-Term Memory (LSTM) to compare the models with a traditional, recurrent neural network. All the algorithms were coupled with a probabilistic multi-quantile loss function to quantify the uncertainty associated with predictions. Analysis suggested that the N-HiTS and N-BEATS models outperformed the LSTM across two testbeds by maintaining accuracy over the extended horizons and depths. The prediction uncertainty was more controlled for N-HiTS and N-BEATS with narrower uncertainty bands across horizons and soil depths, while LSTM exhibited widening intervals. We demonstrate how the proposed architecture can be augmented with uncertainty quantification to provide probabilistic soil water tension predictions that are interpretable without considerable loss in accuracy.

1. Introduction

The amount of water used in agriculture has increased significantly and continues to increase in comparison to a few decades ago (de Fraiture & Wichelns, 2010). The fact that over 70 % of anthropogenic water withdrawals are attributed to agriculture indicates the global scope of agricultural water usage (Sauer et al., 2010). Additionally, about 20 % of the global agricultural area is irrigated, contributing roughly 40 % of the world's agricultural yield (Alexandratos and Bruinsma, 2012). Producing enough food to feed a growing population in an environment where urbanization, population expansion, and increased food demands are competing for limited water resources is a primary issue in agriculture (Pereira, 2017). Consequently, the fundamental water resources that agriculture relies on are threatened by overexploitation and inadequate water management. Given the ongoing population growth and the limited area to increase suitable cropland, irrigation becomes increasingly vital to meet the expected global food demands (Wichelns & Oster, 2006).

Adequate irrigation scheduling is critical to attain good yields and profits, limits the environmental impacts of irrigation, and increases water use efficiency (Pereira, 2017; Zafarmomen et al., 2024). Irrigation scheduling is a process that determines the appropriate amount of crop water demand at the right time. The evapotranspiration and water balance-based approach, soil moisture status monitoring, quantification of plant water status, and modeling are among the commonly practiced irrigation scheduling approaches (Gu et al., 2020). Plant available water is the amount of water held within the root zone and is available for plant uptake (Jensen & Allen, 2016). Changes in soil moisture in the root

* Corresponding author at: Agricultural Sciences Department, Clemson University, Clemson, SC, USA.
E-mail address: samadi@clemson.edu (V. Samadi).

https://doi.org/10.1016/j.compag.2025.110614
Received 8 March 2025; Received in revised form 21 May 2025; Accepted 29 May 2025
Available online 6 June 2025

zone directly affect the plants' available water and regulate their water uptake and growth (Cai et al., 2019). Soil moisture is thus an important gauge for crop water stress and plays a significant role in irrigation decision-making (Feki et al., 2018). In addition to maintaining plant growth, soil moisture is a crucial element in the water cycle of soil–-plant-atmosphere continuum systems (Schlesinger & Jasechko, 2014). Promoting sustainable irrigation management practices to increase crop yield begins with an accurate prediction of soil water dynamics in the root zone. Consequently, accurate prediction of soil moisture is important to bridge the gap between crop irrigation demand and water availability.

Conventionally, the spatial–temporal variability of soil moisture is predicted by using physically based models. Models such as HYDRUS (Šimůnek et al., 1998), Root Zone Water Quality Model (RZWQM2;Ma et al., 2001) and Soil-Water-Atmosphere-Plant (SWAP; van Dam et al., 1997) are among the extensively used physically based models for root zone soil moisture prediction. Although physical models are recognized for their explainability, their performance is hampered by the numerous model parameters, introducing model uncertainty, defective representation of land-surface processes and their high computational power (Li et al., 2022). With the advancement of computing power, neural network models have gathered momentum in soil moisture prediction. Notable algorithms include Multilayer Perceptron (MLP; Rosenblatt, 1958), Support Vector Machine (SVM; Cortes & Vapnik, 1995), Random Forest (RF, Breiman, 2001) and Artificial Neural Networks (ANN; Rosenblatt, 1958). Studies that compared neural networks and physically based models demonstrated that the former can accurately estimate soil moisture without prior information on physical parameters such as soil texture and hydraulic characteristics (Gumiere et al. 2020; Li et al. 2020).

Deep Neural Network (DNN) is the state-of-the-art data-driven method that has made substantial strides in time series modeling of numerous study fields (LeCun et al., 2015; Umutoni and Samadi, 2024). Unlike conventional data-driven models, DNN possesses deeper networks that use different convolutions to represent the data in a hierarchical manner (Kamilaris & Prenafeta-Boldú, 2018). Furthermore, the backpropagation algorithm enables DNN to learn complex data structures, enabling algorithms to fine-tune model parameters of each layer in retrospect (LeCun et al., 2015). Due to DNN's superior prediction ability, these algorithms have been significantly applied in soil moisture prediction. Cai et al. (2019) were among the first scholars who used a deep learning regression network (DNNR) to predict soil moisture. Their results showed that DNNR can achieve better performance compared to MLP, outperforming the latter in terms of generalization and scalability capabilities. Shortly after, Yu et al. (2021) coupled a Convolutional Neural Network (CNN) with a Gated Recurrent Unit (GRU) to predict soil moisture at various depths of the maize root zone for effective irrigation planning. Recently, Zhao et al. (2023) used a hybrid Bidirectional Gated Recurrent Unit (BiGRU) and Long Short-Term Memory (LSTM) model to predict soil moisture focusing on learning the seasonality and trends in the time series data, resulting in a strong model generalization ability. Other research endeavors have attempted to predict soil moisture at a global or regional scale using remote sensing data, all exhibiting skillful performance and significant accuracy (see Zhang et al., 2017; Lee et al., 2019; Liu et al., 2022; Roberts et al., 2022, among others).

Although DNN has enabled substantial improvement in soil moisture forecasting, most DNN algorithms still struggle to provide accurate long-horizon forecasts. Furthermore, soil moisture dynamics are influenced by several factors such as precipitation, irrigation, evapotranspiration, and soil texture. The nonlinear interactions among these variables present distinctive challenges for accurate soil moisture prediction by traditional neural networks (Cai et al., 2019; Wang et al., 2023). Two common challenges faced are the predictions' variability and the increase in computing complexity as the forecasting horizon increases (Challu et al., 2023). For example, the memory and computational power of fully connected layers increase fourfold with respect to the forecasting horizon length. To address these challenges, Oreshkin et al. (2020) introduced the Neural Basis Expansion Analysis for interpretable Time Series forecasting (N-BEATS) based on the concept of backward and forward residual links to learn time series representations without relying on external feature engineering or domain knowledge. Building on the concept of N-BEATS, Challu et al. (2023) invented Neural Hierarchical Interpolation for Time Series Forecasting (N-HiTS), a new neural network architecture designed to reduce the memory and computational power of neural networks without affecting their ability to model long-term dependencies. Both algorithms were structured to generate interpretable predictions (Saberian et al., 2024).

This study is the first attempt known to the authors that implemented N-HiTS and N-BEATS to predict soil moisture fluctuations expressed as soil water tension across three root zone depths and prediction horizons, enabling both long-term forecasting and interpretability in irrigation contexts. The scope of this research is to develop fully connected N-BEATS and N-HiTS architectures by enhancing their input decomposition via multi-rate soil water tension data sampling and a synthesis of the outputs via multi-scale interpolation. Interpretability of predictions is addressed by leveraging the backward and forward residual links of N-HiTS and N-BEATS to explicitly decompose the soil water tension data into trends and seasonality. This enhanced the clarity of short-to-long-term variations and enabled the models to generate more explainable predictions. We benchmarked these advanced algorithms with LSTM to compare the results with a recurrent time series prediction model. In addition, the uncertainty associated with soil moisture predictions is quantified using a multi-quantile loss function to diminish the error associated with the prediction. The soil water tension prediction workflow has been emphasized with the multi-horizon and depth prediction strategy in all these approaches. We tested the models across two testbeds with different soil profiles using multiple years of sensor data. Our extensive experiments show the importance of the proposed algorithms and validate significant improvements in the accuracy and computational complexity of the proposed algorithms.

The goal of this research was to develop interpretable DNN models for soil water tension prediction across multiple soil depths and prediction horizons. The specific objectives to accomplish this goal were to:

(i) Improve sub-daily prediction of soil water tension across multiple root zone depths by leveraging the interpretability properties of the state-of-the-art N-HiTS and N-BEATS models,
(ii) Evaluate the performance of N-HiTS and N-BEATS with respect to LSTM in terms of predictive accuracy and performance,
(iii) Provide probabilistic soil water tension predictions at multiple depths and horizons to complement deterministic estimation for informed irrigation decisions-making.

Our contributions are summarized below:

1. **Interpretability:** We generated soil water tension predictions by lowering the dimensionality of DNN's prediction, extracting key low-dimensional features from the high-dimensional outputs, and harmonizing their time scale with the final output using multiple-scale hierarchical interpolation. This innovative approach is not exclusive to our model and can be applied to other algorithmic frameworks. The proposed algorithms can be also retrained and applied to other regions and climates.
2. **Benchmarking with LSTM:** We compared our approaches with LSTM as a benchmark model, using different error metrics to perform a thorough comparative analysis, examining the weaknesses and strengths of each approach. This comparison provided valuable insights into the strengths and weaknesses of our proposed algorithms and identified opportunities for improvement.
3. **Uncertainty Quantification:** We computed the loss based on predictions at multiple quantile levels using the multi-quantile loss

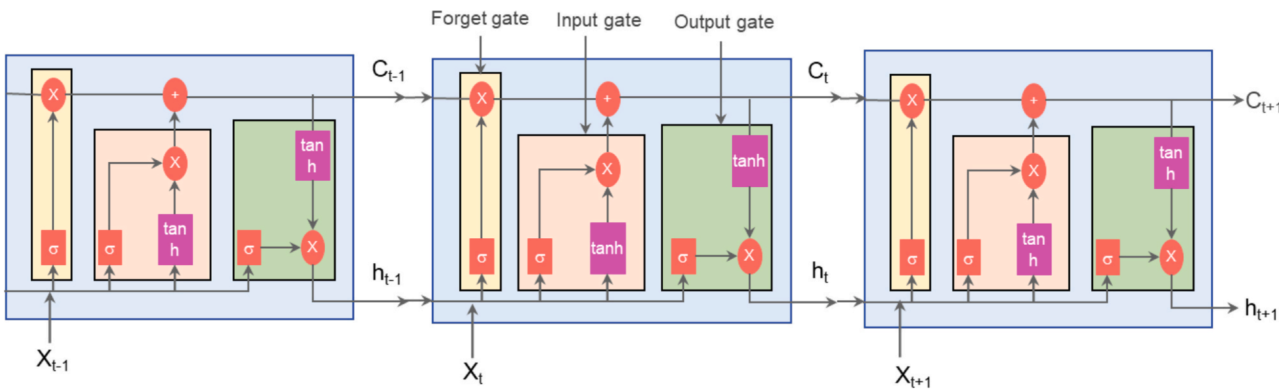


Fig. 1. The structure of LSTM model. Xt-$_1$, Xt and Xt$_{+1}$ are input soil water tension, Ct$_{-1}$, Ct, Ct$_{+1}$ are memory cells, and ht-1, ht, and ht$_{+1}$ are hidden layers. The tanh function is used to regulate the values in the cell state.

function (MQL) to capture the possible range of each model's predictions and provide a more comprehensive assessment of error and uncertainty across simulation horizons. Moreover, we employed several uncertainty quantification evaluation methods to improve the models' probabilistic predictions and support their viability in irrigation decision making.

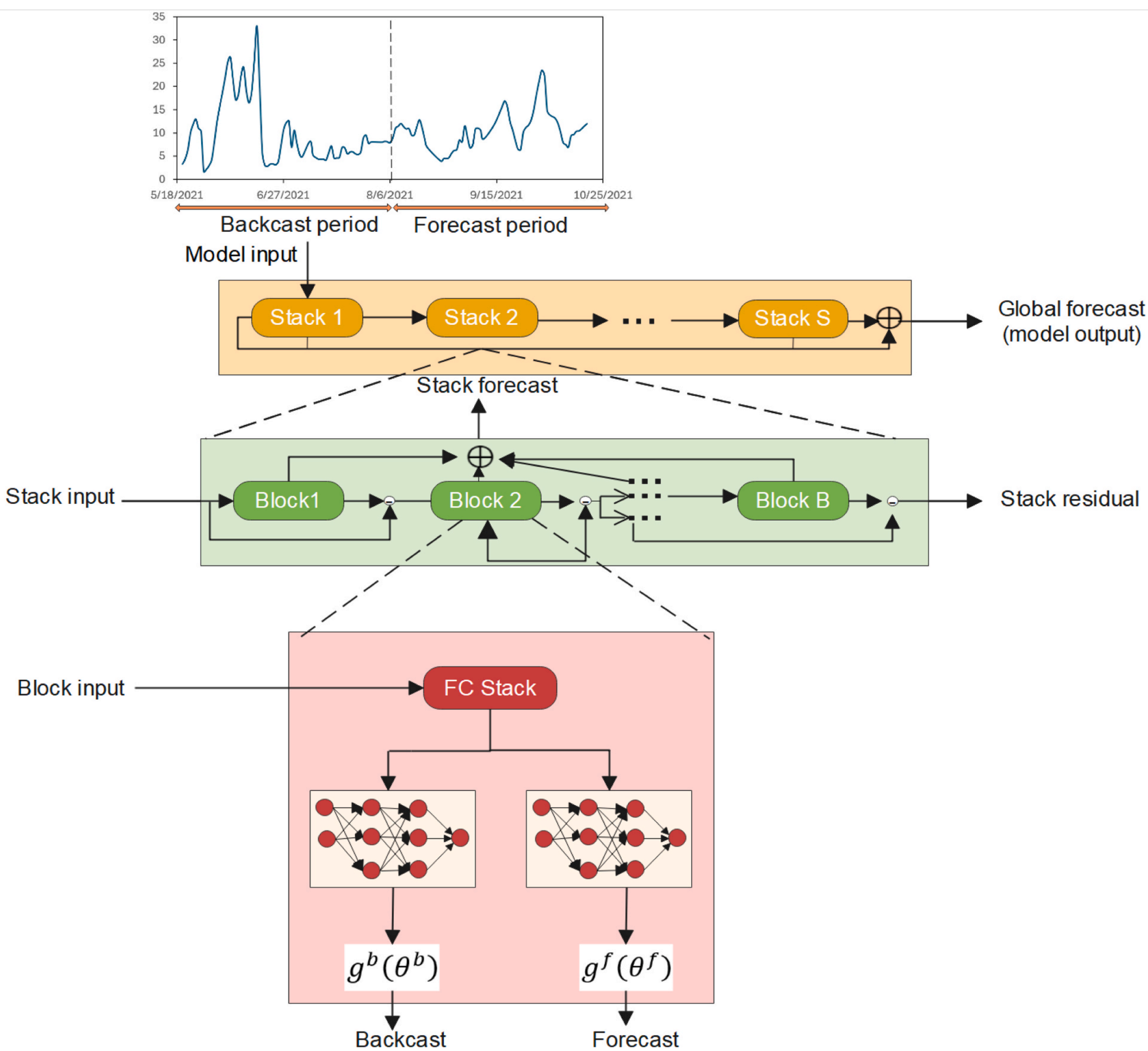


Fig. 2. The workflow of N-BEATS architecture includes stack input, block input, backcast and forecast components. Each block consists of layers of FC network with ReLu non-linearities. It uses the backward θ^b and forward θ^f expansion coefficients to generate the backcast and forecast. Multiple blocks form a stack, and forecasts from stacks are summed up hierarchically to produce the overall model forecast.

2. Materials and methods

2.1. Data collection and preprocessing

Following a rainfall or irrigation event, soil pores become saturated with water. As drainage begins, gravity causes water in macro-pores to drain quickly. Over time, the remaining soil moisture is contained only in smaller pores by capillary and surface tension forces. Soil water tension represents the force required to extract water from these micro soil pores and indicates the status of soil moisture content. The drier the soil, the higher the soil water tension and energy needed to extract soil moisture, making it difficult for plants to absorb water. Monitoring soil water tension helps determine when the soil moisture content has reached a set threshold that warrants irrigation. The magnitude of soil water tension with respect to the soil moisture content varies by soil type; their relationship is described by soil water characteristics curves originally developed by Buckingham (1907) and Gardner (1920). Soil water tension data from two irrigation sites located in Georgia (GA) and South Carolina (SC), collected for irrigation scheduling purposes, were used in this study. The description of each site is provided below, and the data is available upon request.

2.1.1. Irrigated field in GA

Data used for this site was collected in a study conducted at the University of Georgia's Stripling Irrigation Research Park (SIRP) located near Camilla, GA in a 4-ha research field. The field was divided into three blocks of 27 plots each. Each plot was 14.5 × 14. 5 m (48 × 48 ft or 48 ft long × 16 rows wide). Data was collected in eight middle rows in each plot, and the four rows on either side of the middle eight were buffers. The soil is classified as a Lucy Loamy Sand with an available water holding capacity of 0. 08 cm/cm with 0 to 5 % slope. Soil texture varies marginally across the field, with 83 % Sand, 10 % Silt and 7 % Clay in the South block to 86 % Sand, 8 % Silt, and 6 % Clay in the North block. The field was irrigated with a variable rate-enabled lateral irrigation system. Three Watermark soil moisture sensors were placed in each plot at 0.15, 0.3 and 0.46 m of soil depth, respectively, to monitor soil moisture using matric potential type soil moisture sensors during each growing season from 2019 to 2021.

2.1.2. Irrigated field in SC

The second study site is a center pivot irrigated farmer's field located in Blackville, Barnwell County, SC. The field's soil texture is a combination of loamy and sandy soil. A soil moisture probe consisting of Watermark sensors collecting soil water tension data every 30 min during the cotton growing season was placed in the field to monitor changes in soil matric potential at 0.15, 0.3, and 0.6 m of soil depth from 2020 to 2022.

2.1.3. Data preprocessing

The soil water tension data were collected hourly and half-hourly for the GA and SC fields, respectively. Data cleaning involved graphically inspecting the plots of soil water tension variations in time and removing outliers or data points recorded when sensors were not operating correctly. Furthermore, the data collected in SC were converted to an hourly scale by taking the average value of the data collected every 30 min.

2.2. DNN algorithms

We used two interpretable DNN algorithms (N-BEATS and N-HiTS) and benchmarked them with LSTM. Each of these algorithms is discussed in detail below.

2.2.1. LSTM

LSTM, introduced by Hochreiter and Schmidhuber (1997), is a recurrent neural network (RNN) designed to effectively handle sequential data by using gates to selectively remember or forget information across multiple time steps. Unlike traditional RNNs, LSTMs have a more advanced structure consisting of a hidden state that acts as short-term memory and an additional cell state that functions as long-term memory. Fig. 1 depicts the structure of an LSTM network.

2.2.2. N-BEATS

N-BEATS is a DNN architecture created for forecasting one-variable time series (Oreshkin et al., 2020). The architecture of this algorithm is based on a stack of fully connected layers arranged in blocks, each responsible for capturing different aspects of the time series data. Each block l takes its distinctive input X_l and produces two vector outputs. The backcast $\widehat{X}_l$ which is the best estimate of X_l following the boundaries on the functions that the block can use to estimate signals, and forward forecast the length of horizon $(H)\widehat{y}_l$. The network architecture is illustrated in Fig. 2. The first block receives the overall model input as X_l while the input X_l of the subsequent blocks are residual outputs of the previous blocks.

Within each block there are two separate parts. The first part is a fully connected network that yields the forward θ_l^f and the backward θ_l^b predictors of expansion coefficients. The second part includes the backward g_l^b and the forward g_l^f basis layers that take the corresponding forward θ_l^f and backward θ_l^b expansion coefficients, transform them internally based on the set of basic functions, and produce the backcast $\widehat{X}_l$ and the forecast $\widehat{y}_l$ outputs.

A fixed number of blocks are organized into sequential stacks interconnected using a novel double residual topology. Each connection consists of two residual branches, one operating on the backcast prediction and the other on the forecast. This structure allows blocks to focus on learning part of the input data that previous blocks have not yet learned. The forecast output of each block is hierarchically aggregated, from the stack to the overall network level, to obtain the global forecast as indicated in Fig. 2. The benefit of this architecture is that the partial backcast and forecast can be observed gradually during the simulation process, which facilitates the identification of each stack's contribution and enables the overall results to be interpreted.

In addition to the basic design of N-BEATS inherently allowing model interpretability, interpretability is further imposed by adding structure to the basis layers at the stack level. This is done by decomposing the input data into trends and seasons using some constraints on $g_{s,l}^b$ and $g_{s,l}^f$ to account for the slowly changing pattern in the data for trend, and the recurrent rise and fall fluctuations for seasons present in the input data.

The following group of equations describe the modeling processes. Equation (1) explains the process occurring in the first part of the l-th block.

$$h_{l,1} = FC_{l,1}(X_l), \quad h_{l,2} = FC_{l,2}(h_{l,1}), \quad h_{l,3} = FC_{l,3}(h_{l,2}), \quad h_{l,4} = FC_{l,4}(h_{l,3})$$

$$\theta_l^b = LINEAR_l^b(h_{l,4}), \quad \theta_l^f = LINEAR_l^f(h_{l,4}) \tag{1}$$

where FC is a fully connected network with ReLu non-linearity and LINEAR is a linear projection layer.

The second part of the network maps expansion coefficients θ_l^f and θ_l^b to outputs via basis layers as shown in the equations below (Equation 2).

$$\widehat{y}_l = g_l^f\left(\theta_l^f\right) \quad \text{and} \quad \widehat{X}_l = g_l^b\left(\theta_l^b\right) \tag{2}$$

2.2.3. N-HiTS

N-HiTS is another type of DNN developed specifically for long-term horizon forecasting (Challu et al., 2023). It improves N-BEATS's framework by adding hierarchical interpolation approaches to increase the forecast accuracy for long-term forecasting tasks. The algorithm uses multi-rate sampling of the input data and multi-scale synthesis of the forecast, giving a hierarchical construction of the forecast and greatly

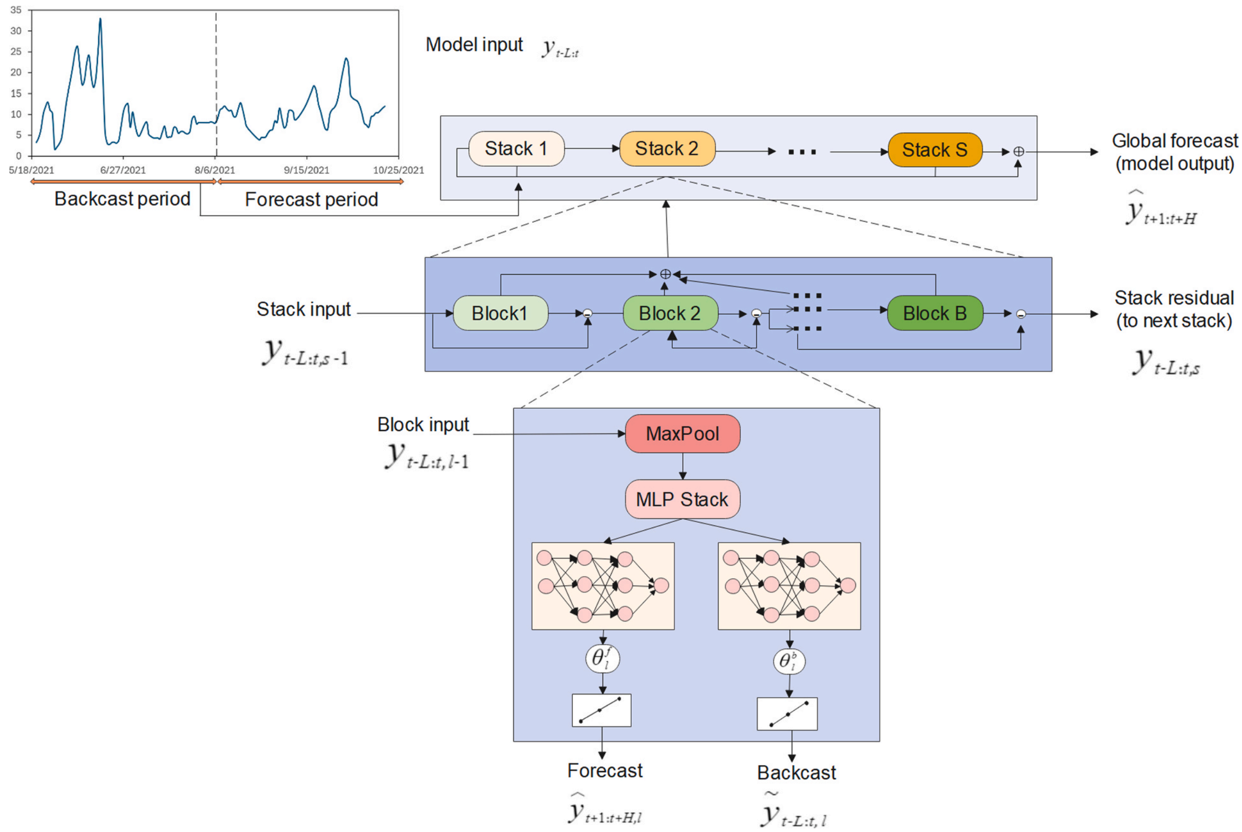


Fig. 3. The workflow of N-HiTS architecture consists of blocks composed of multiple MLPs with ReLu activation functions and a max-pooling layer. Blocks are connected using the double residual mechanism, allowing each block to output a backcast and a forecast. Several blocks form a stack, and the outputs of stacks are averaged to generate the overall model prediction.

lowering the computational power while improving the forecasting accuracy. Similar to N-BEATS, N-HiTS uses a modular block structure but with supplementary features for hierarchical interpolation. Each block has feed-forward neurons, known as MLP, predicting forward and backward coefficients of the basic functions, whereby each basis function learns specific data characteristics. The backcast output is used to process and clean the inputs of the following blocks, while the forecasts are added to determine the final prediction results. The blocks are organized into groups known as stacks that specialize in learning specific data characteristics using distinct basis functions to iteratively fine-tune the forecasts. The structure of the algorithm is depicted in Fig. 3.

A kernel size k_l is used on a MaxPool layer at the input of each block l to assist the algorithm in analyzing components of the input data with an explicit scale. Larger k_l omit highly frequent components of the block's input to impose it to concentrate on studying the less frequent content. This process, known as multi-rate sampling, allows each block to specialize in learning a specific input signal. Given block l input $y_{t-L:t,l}$ (where the input to the first block $l = 1$ is the overall model input, $y_{t-L:t,1} \equiv y_{t-L:t,l}$), this operation is expressed in Equation (3).

$$y^{(p)}_{y_{t-L:t,l}} = MaxPool\left(y_{t-L:t,l}, k_l\right) \quad (3)$$

After subsampling, block l takes the input and regresses nonlinearly the forward θ^f_l and backward θ^b_l interpolation coefficients of MLP that learns the hidden vector $h_l \in \mathbb{R}^{N_h}$, which is then transformed linearly as shown by Equation (4):

$$h_l = MLP_l\left(y^{(p)}_{y_{t-L:t,l}}\right)$$

$$\theta^f_l = LINEAR^f(h_l)\theta^b_l = LINEAR^b(h_l) \quad (4)$$

The coefficients are then used to generate the backcast $\tilde{y}_{t-L:t,l}$ and forecast $\hat{y}_{t+1:t+H,l}$ outputs of the block.

Furthermore, N-HiTS uses a temporal interpolation technique to avoid a drastic increase in the computational cost in multi-horizon forecasting tasks. This approach uses an expressiveness ratio r_l to express the dimensionality of the interpolation coefficients and regulate the number of parameters for each output time unit. Hierarchical interpolation is achieved by distributing these expressiveness ratios across blocks in a way that matches the multi-rate sampling. This explicit control over prediction granularity using the r_l ensures that forecasts are smooth and interpretable. The overall forecast is obtained by aggregating the interpolations generated at different time-scale hierarchy levels. Multi-rate signal sampling and temporal interpolation for forecast construction are unique techniques that make N-HiTS more transparent and interpretable.

2.3. Model performance measures

In this study, the simulation results were evaluated using the Nash-Sutcliffe Efficiency coefficient (NSE; Nash & Sutcliffe, 1970), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), and graphically visualized to assess their performance across depths and horizons. An NSE value equivalent to 1 shows a perfect agreement between observed and modeled soil water tension. On the other hand, an NSE value of 0 indicates that the predictive ability of the model is as good as the mean of the time series data (Nash & Sutcliffe, 1970). RMSE is also commonly used when evaluating the quality of a model's prediction. RMSE values vary from 0 to infinity; the lower the RMSE value, the better the model performance is (Jamil & Akhtar, 2017). MAE is another error estimation measure that is calculated as the arithmetic average of the difference between predictions and observations

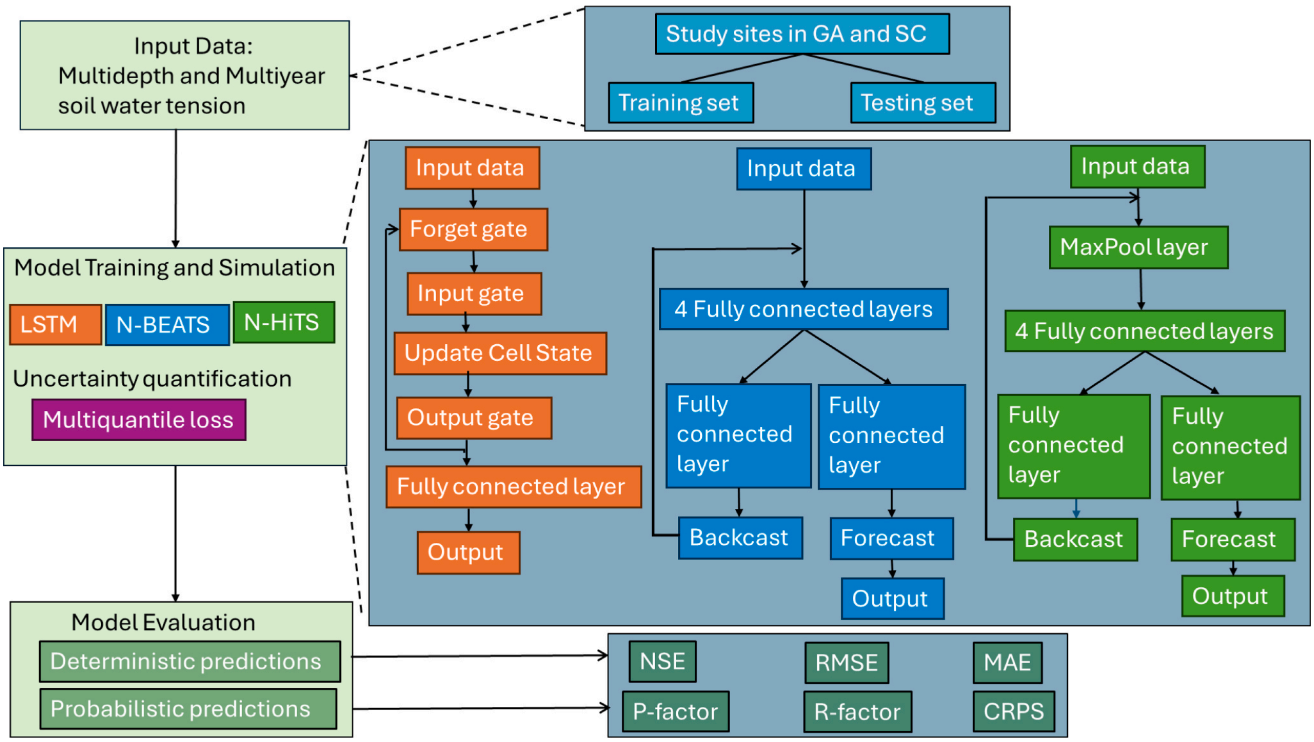


Fig. 4. The workflow of the proposed methodologies.

(Hyndman & Koehler, 2006). Like RMSE, MAE ranges between 0 and infinity, lower values show good model performance.

To estimate the uncertainty associated with the deterministic predictions, the MQL function was used. MQL computes the mean quantile loss (QL) across a specific set of quantiles based on the residue value by introducing distinct penalties for under-predicted and over-predicted values depending on the quantile level considered. The probability levels included in our analysis were 50, 75, 90, 95, and 99, corresponding to the 0.5, 0.75, 0.90, 0.95, and 0.99 quantiles, respectively. For each quantile, the models produced lower, median, and upper predictions of soil water tension values, iteratively learning to improve the prediction performance at each quantile level (Equations (5) and (6)).

$$QL\left(y_T, \hat{y}_T^{(q)}\right) = \frac{1}{H}\sum_{T=t+1}^{t+H}\left[(1-q)\left(\hat{y}_T^{(q)} - y_T\right) + q\left(y_T - \hat{y}_T^{(q)}\right)\right] \tag{5}$$

$$MQL\left(y_T, \left[\hat{y}_T^{(q_1)}, \cdots, \hat{y}_T^{(q_n)}\right]\right) = \frac{1}{n}\sum QL\left(y_T, \hat{y}_T^{(q_i)}\right) \tag{6}$$

where q refers to the predicted quantile, y_T is the observed value at time T, $\hat{y}_T^{(q)}$ is the corresponding prediction for a given quantile, and H is the forecast horizon.

Without applying any smoothing or post-processing of probabilistic results, the performance of each model's uncertainty estimates was assessed using the P- and R-factors. The P-factor measures the percentage of observed data points that fall within the 95 percent prediction uncertainty (95 PPU), Equation (7), revealing how effective a model is in providing predictions that align with the observed value and measuring the uncertainty associated with those predictions. The P-factor ranges between 0 and 100 %; the higher it is, the better the model fits and the lower the uncertainty. The R-factor evaluates the width of the uncertainty band around the predictions relative to the variability of the observed data. A lower R-factor reflects a narrower uncertainty band, indicating a higher model precision, while a higher R-factor indicates a wider uncertainty band, implying greater uncertainty in model predictions. Furthermore, we employed these metrics to determine the difference in models' predictions convergence and accuracy at the 50, 75, 90, and 99 % uncertainty bands. The P- and R-factors are computed using Equations (8) and (9).

$$95\,PPU = \left[\hat{y}_T^{(0.05)}, \hat{y}_T^{(0.95)}\right] \tag{7}$$

$$P-factor = \frac{\text{Observations bracketed by 95 PPU}}{\text{Number of observations}} \times 100 \tag{8}$$

$$R-factor = \frac{\frac{1}{k}\sum_{i=1}^{k}(X_U - X_L)}{\sigma_x} \tag{9}$$

where k is the number of observations, X_U and X_L are the upper and lower limits of the uncertainty band, respectively, and σ_x is the standard deviation of observed data.

Additionally, we used the Continuous Ranked Probability Score (CRPS) metric (Equation (10)) to assess the accuracy of the probabilistic predictions of the 95 PPU against measured soil water tension. The closer the score is to zero, the more accurate the probabilistic predictions are.

$$CRPS(F, x) = \int_{-\infty}^{x} F(y)^2 dy + \int_{x}^{+\infty} (F(y) - 1)^2 dy \tag{10}$$

$F(\ y)$ is the Cumulative Distribution Function (CDF) of the predictions' distribution, while x is the actual soil water tension value at a specific time and depth

CDFs were used to assess the cumulative probability of the residuals of each model and evaluate the variability and spread. The steepness of the CDF curve shows the variability of the residuals; a steeper CDF curve implies that the errors are clustered around zero, while the spread indicates the extent of the errors. Mathematically, CDF is expressed in Equation (11).

$$F(x) = P(X \leq x) \tag{11}$$

where, $F(x)$ is the probability that X (residual) is less than or equal to x (any possible value of X)

The methodological framework followed in this study is summarized

Table 1
Optimized hyperparameter values.

Hyperparameter	LSTM	N-HiTS	N-BEATS
Scaler type	minmax	minmax	minmax
Epochs	10,000	3000	3000
Batch size	32	32	32
Lookback window (h)	7	7	7
Number of layers	1	1	1
Number of stacks	Not applicable	3	3
Number of blocks per stack	Not applicable	1	1
Number of units in each hidden layer	64	512	512
Dropout rate	0.2	0.2	0.0
Activation function	tanh	ReLU	ReLU
Optimizer	Adam	Adam	Adam

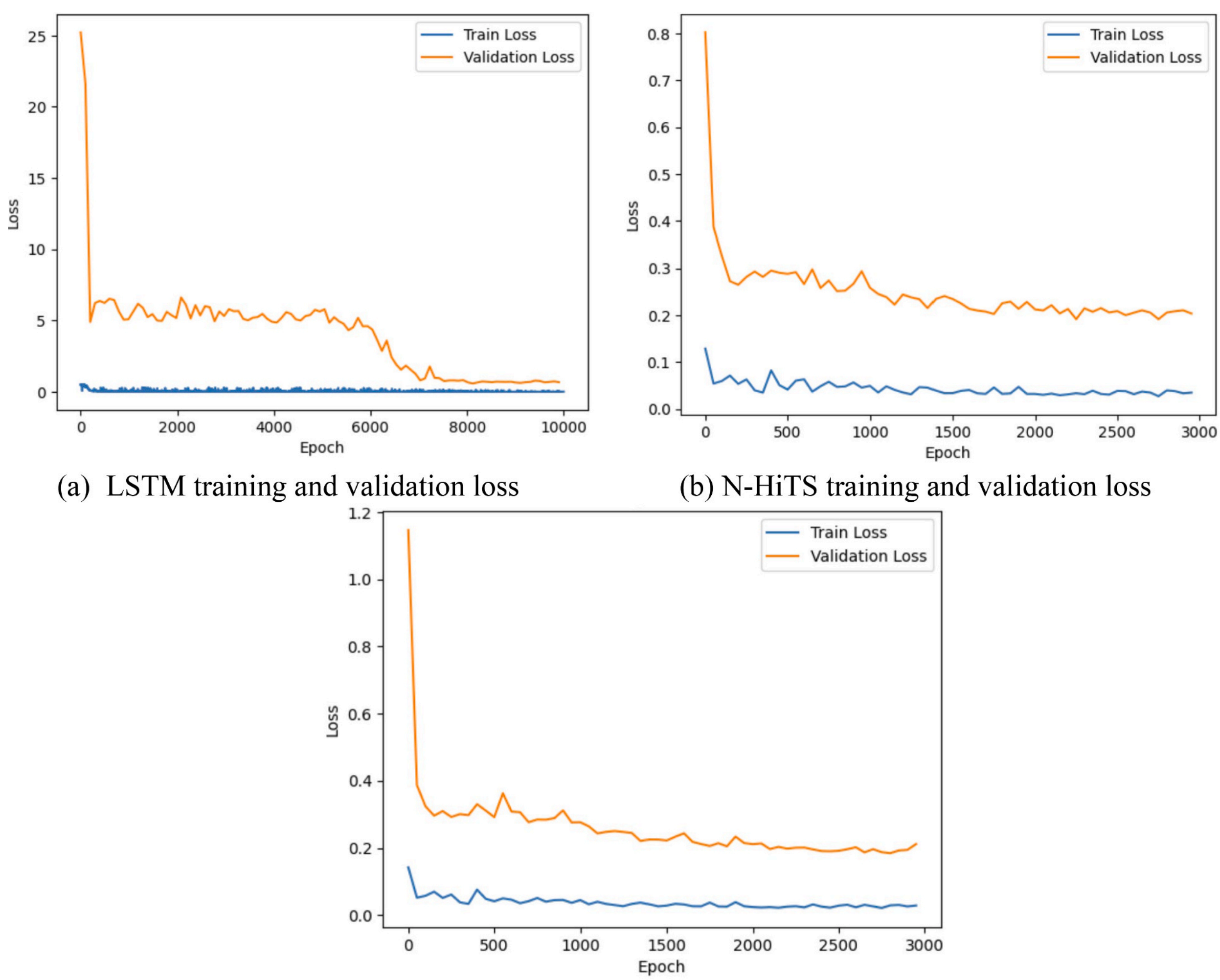
in Fig. 4. The soil water tension prediction models were executed on a NVIDIA V100 GPU and implemented using the NeuralForecast open-source library (Olivares et al., 2022). As illustrated in Fig. 4, we first split the three-year soil water tension data collected in both sites into the training (80 %) and testing (20 %) sets, which were then used as input for each model. Based on this ratio, the data from the two test beds were incorporated into LSTM, N-BEATS and N-HiTS. All algorithms were coupled with the MQL function to estimate the uncertainty associated with the predictions. In this research, deterministic results were evaluated using the NSE, RMSE and MAE metrics, while probabilistic predictions were assessed using the P and R-factors and CRPS.

3. Results and discussion

Three years of sub-daily soil water tension data were collected during the growing seasons. The datasets were split chronologically into training and testing sets, with two years of data used for training and the remaining year for testing the developed models. The training sets, years 2019 and 2020 for GA and 2020 and 2021 for SC, comprised 5576 and 5589 data points, respectively. The testing sets, year 2021 and 2022 for GA and SC, consisted of 3369 and 2677 hourly data points, respectively. The training sets were used to train the models to find the best parameter set that minimizes the error between observations and predictions, while a validation set consisting of the last seven prediction windows for each horizon was used to monitor the validation loss during training. The testing sets were used to evaluate the models' performance on unseen data. Furthermore, we used LSTM as a benchmark model to compare the accuracy of N-HiTS and N-BEATS models against a recurrent algorithm with a mechanism to avoid the vanishing gradient problem and learn temporal dependencies in data. Below, we discussed the hyperparameter tuning, simulation results and uncertainty quantification for each testbed.

3.1. Hyperparameter tuning

We employed a systematic trial-and-error method to find the best parameter set for each model. The tuned hyperparameters include batch size, number of blocks per stack, number of stacks, and the optimizer,



(a) LSTM training and validation loss (b) N-HiTS training and validation loss

(c) N-BEATS training and validation loss

Fig. 5. The training against the validation loss of each model for the 12 h horizon at 0.15 m depth.

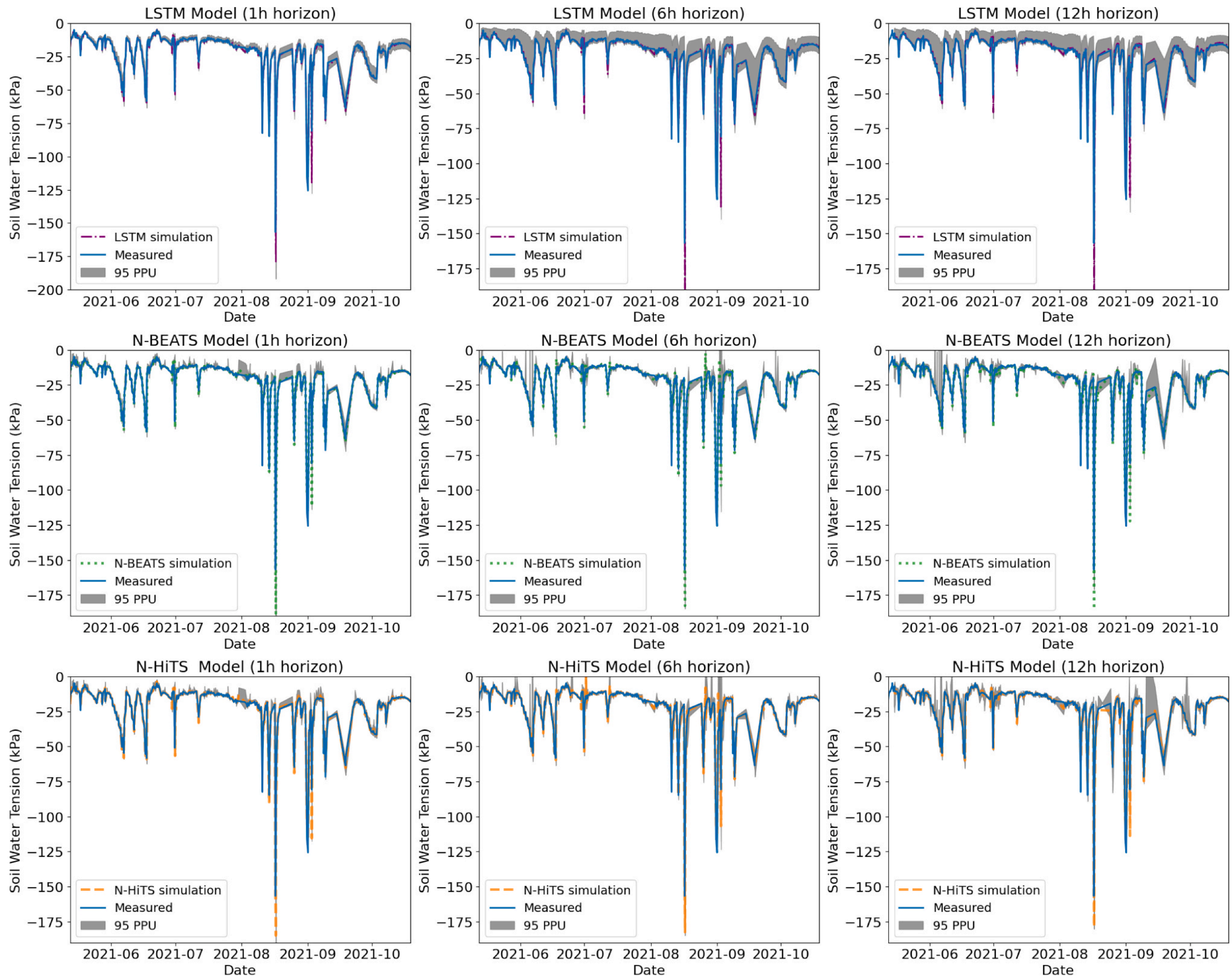


Fig. 6. Soil water tension simulations across different time horizons at 0.15 m soil depth. The grey shaded area indicates 95 PPU.

along with other parameters listed in Table 1. The network weights were continuously updated to reduce errors and bias in the simulation. Initial weights of the network were set to random values, and a gradient-based Adam optimizer was used to adjust the network weights. Early stopping with a patience of 100 epochs was applied for N-HiTS and N-BEATS to avoid overfitting, while a patience of 500 epochs was used for LSTM. After the model training reached a certain limit, increasing the number of iterations was no longer significant for improving the performance, as illustrated in Fig. 5. The influence of hidden layers on model precision has always been a key problem in DNN performance because they can influence error on the nodes to which their output is connected. The minimal error reflects better network stability, while a higher error reflects poor stability. The optimal number of hidden layers was estimated by a trial-and-error method, which cost time but was more efficient in reaching high accuracy. The time used to train the LSTM, N-HiTS, and N-BEATS models was on average 82, 43, and 28 s, respectively.

3.2. GA irrigated field case study

3.2.1. Soil water tension simulation at 0.15 m depth

The soil water tension prediction at 0.15 m soil depth with 1 h increments is graphically depicted in Fig. 6. The simulation results indicated that the trained N-BEATS and N-HiTS models closely followed the general trend of the measured data, which was indicated by their ability to capture both the seasonality and trend effectively. Both models provided skillful predictions compared to LSTM as the benchmark model. In addition, both N-HiTS and N-BEATS successfully predicted water stress conditions, particularly in August and September, when irrigation demands were high. The performance metrics of each model (see Table 2), further supported the prediction capabilities of N-BEATS and N-HiTS compared to the benchmark LSTM. N-BEATS and N-HiTS relatively better performance is attributed to their structural design, which is equipped with mechanisms to explicitly structure the input data into distinct temporal patterns to account for seasonality and trends. Based on the performance error evaluation, N-HiTS presented the least error between the measured and predicted values. N-HiTS slightly outperformed N-BEATS because of its multi-rate sampling and hierarchical interpolation properties that enhance the model's ability to tackle multi-scale temporal patterns efficiently.

Results obtained after increasing the horizon to 6 h showed that the tested models were still robust with N-HiTS providing the best performance across all three metrics, as indicated in Table 2. The NSE value associated with N-HiTS prediction was very good, indicating the ability of this algorithm to predict soil water tension variabilities over time. N-BEATS results suggested that the prediction horizon minimally affected the model as the error metrics slightly altered. The benchmark model, on the other hand, is substantially affected by the increase in time horizon as indicated by a 5.44 % drop in NSE value. The decline in performance is mainly observed for peak values where the model overestimated the magnitude of soil water tension. The reduced performance of the benchmark model exemplifies its limited ability to capture long-term dependencies in time series like N-HiTS and N-BEATS despite being

Table 2
Evaluation metrics of prediction at 1, 6, and 12 h horizons across three soil depths in GA. Best performances are shown in bold.

Model	Soil depth	Horizon	NSE (%)	RMSE	MAE	P-factor (%)	R-factor	CRPS
N-HiTS	0.15 m	1 h	**92.15**	**4.3**	**1.22**	82.72	**0.21**	**1.58**
N-BEATS			90.85	4.79	1.32	80.58	**0.2**	1.64
LSTM			91.41	4.64	1.35	**90.87**	0.37	1.77
N-HiTS		6 h	**91.59**	**4.60**	**1.58**	78.67	0.39	1.61
N-BEATS			91.41	4.65	1.64	79.26	**0.34**	**1.48**
LSTM			85.97	5.94	1.59	**92.37**	0.91	2.41
N-HiTS		12 h	**92.05**	**4.5**	**1.52**	86.55	0.41	**1.56**
N-BEATS			89.77	5.1	1.9	83.73	**0.38**	1.67
LSTM			82.13	6.75	1.92	**91.37**	0.9	3.00
N-HiTS	0.3 m	1 h	**98.57**	**0.79**	**0.42**	**72.23**	0.29	0.59
N-BEATS			98.53	0.8	0.43	57.28	**0.18**	**0.45**
LSTM			98.52	0.80	0.47	72.04	**0.19**	0.45
N-HiTS		6 h	98.61	0.78	0.44	84.54	0.28	0.42
N-BEATS			**98.7**	**0.75**	**0.42**	81.21	**0.29**	**0.41**
LSTM			98.33	0.85	0.52	**85.71**	0.36	0.54
N-HiTS		12 h	**98.57**	**0.79**	**0.44**	84.34	**0.30**	0.45
N-BEATS			98.32	0.85	0.49	81.93	0.32	**0.43**
LSTM			98.11	0.90	0.56	**86.14**	0.96	1.12
N-HiTS	0.46 m	1 h	97.3	0.70	0.24	72.04	0.17	0.2
N-BEATS			**98.35**	**0.55**	**0.22**	58.64	**0.12**	**0.18**
LSTM			97.75	0.64	0.30	**93.01**	0.27	0.27
N-HiTS		6 h	**98.27**	**0.20**	**0.55**	83.37	0.19	**0.18**
N-BEATS			**98.27**	**0.20**	**0.55**	80.04	**0.18**	**0.18**
LSTM			96.02	0.46	0.83	**96.09**	1.28	0.77
N-HiTS		12 h	**98.40**	**0.51**	**0.19**	89.16	0.28	**0.21**
N-BEATS			98.15	0.55	0.21	84.74	**0.25**	**0.21**
LSTM			97.07	0.69	0.28	**93.17**	3.00	1.47

powerful for sequential data.

At the 12 h horizon, LSTM results showed a sharp decline in NSE value and significant increase in RMSE and MAE error metrics (see Table 2). This reflects the fact that LSTM struggled to explain variability in the measured data and showed increased deviation from the actual values over longer horizons. Compared to the benchmark model, the accuracy of N-HiTS and N-BEATS decreased slightly at the 12 h horizon, whereby the NSE declined by 0.1 % and 1.08 % for N-HiTS and N-BEATS, respectively. Both models followed the measured data pattern closely despite some noticeable over- or underestimation values. While N-HiTS and N-BEATS performed better than LSTM, they also struggled to accurately capture soil water tension values that rose to extremes. This is arguably due to the limited occurrence of such peak values, which does not provide the models with enough information to be trained on. Nevertheless, both N-BEATS and N-HiTS exhibited compelling prediction results, demonstrating their ability to predict different ranges of soil water tension values at 0.15 m. This underscores their ability to signal water depletion in the root zone, indicating that irrigation is needed.

In addition, 95 PPUs were quantified for each model. The shaded gray area in Fig. 6 illustrates 95 PPU and delineates the boundary within which the measured values are expected to fall 95 % of the time. The goal is to have 95 PPU as narrow as possible while capturing most observational data. A wider interval indicates more significant uncertainty in predictions, while the narrower uncertainty band implies that the model is more certain about the accuracy of its prediction. During periods of instant change, such as around peak values, all models presented wider prediction intervals, showing greater uncertainty in predicting extreme soil water tension values. This can be expected as accurately predicting rapidly changing conditions is challenging. LSTM exhibited good predictions of measured values while intermittently providing a wider 95 PPU, indicating satisfactory responsiveness with some sensitivity to data fluctuations. N-HiTS, on the other hand, maintained a closer fit to the measured data and had narrower intervals, suggesting that it was more robust to sudden changes in measured data. 95 PPU associated with the N-BEATS simulation showed a similar convergence to that of N-HiTS. Table 2 presents the P-factor and R-factor indices used to quantify the confidence level in predictions. Overall, N-HiTS and N-BEATS models captured 82.72 % and 80.58 % of observational data, respectively, while the benchmark model performed better regarding bracketing observations with a higher P-factor (90.87 %). The obtained R-factor values that compare the average width of the prediction interval to the variability of the observed data indicate that LSTM has a higher R-factor than N-HiTS and N-BEATS, as noted in the broader intervals around peak values illustrated in Fig. 6. On the contrary, N-HiTS and N-BEATS, with their narrower R-factor, showed more confidence in soil water tension prediction despite failing to capture some extreme values like LSTM. The comparatively better probabilistic performance of N-HiTS and N-BEATS was further demonstrated by their lower CRPS values across all horizons compared to those of LSTM (Fig. 7).

The uncertainty band for the 6 h horizon was significantly broader than that of 1 h for LSTM throughout the simulation period. The higher P-factor for LSTM compared to N-HiTS and N-BEATS suggests (see Table 2) that the uncertainty band enclosed most of the observed values, as supported by Fig. 6. Despite presenting a less smooth uncertainty band, N-HiTS and N-BEATS appeared more confident in their predictions as they maintained a narrow uncertainty band throughout the simulation period. Even though both N-HiTS and N-BEATS provided similar uncertainty estimates, N-HiTS provided the best overall performance, exhibited by its ability to capture variability in the data.

The 95 PPU estimation at the 12 h horizon indicated that the benchmark model scored the highest P-factor bracketing more measured data points in the confidence interval compared to N-HiTS and N-BEATS (Fig. 6). However, LSTM's highest R-factor indicated a wider interval and less prediction confidence. It is important to note that N-HiTS and N-BEATS uncertainty bands were quite similar; both models exhibited a high uncertainty, although the extent of their uncertainty bands slightly differs. Overall, N-BEATS appeared more confident in their predictions, as indicated by the narrower uncertainty band and a lower R-factor. N-HiTS, however, remained the best performer as it generally showed a better performance across the other error metrics.

A comparison of the prediction performance of three models using the CDF of residuals (Fig. 8) indicated that the benchmark LSTM model presented a wider spread of residuals when the horizon increased, as evidenced by the gradual increase in the spread of the curves. For the 1 h horizon, all models showed slightly similar results when the residuals

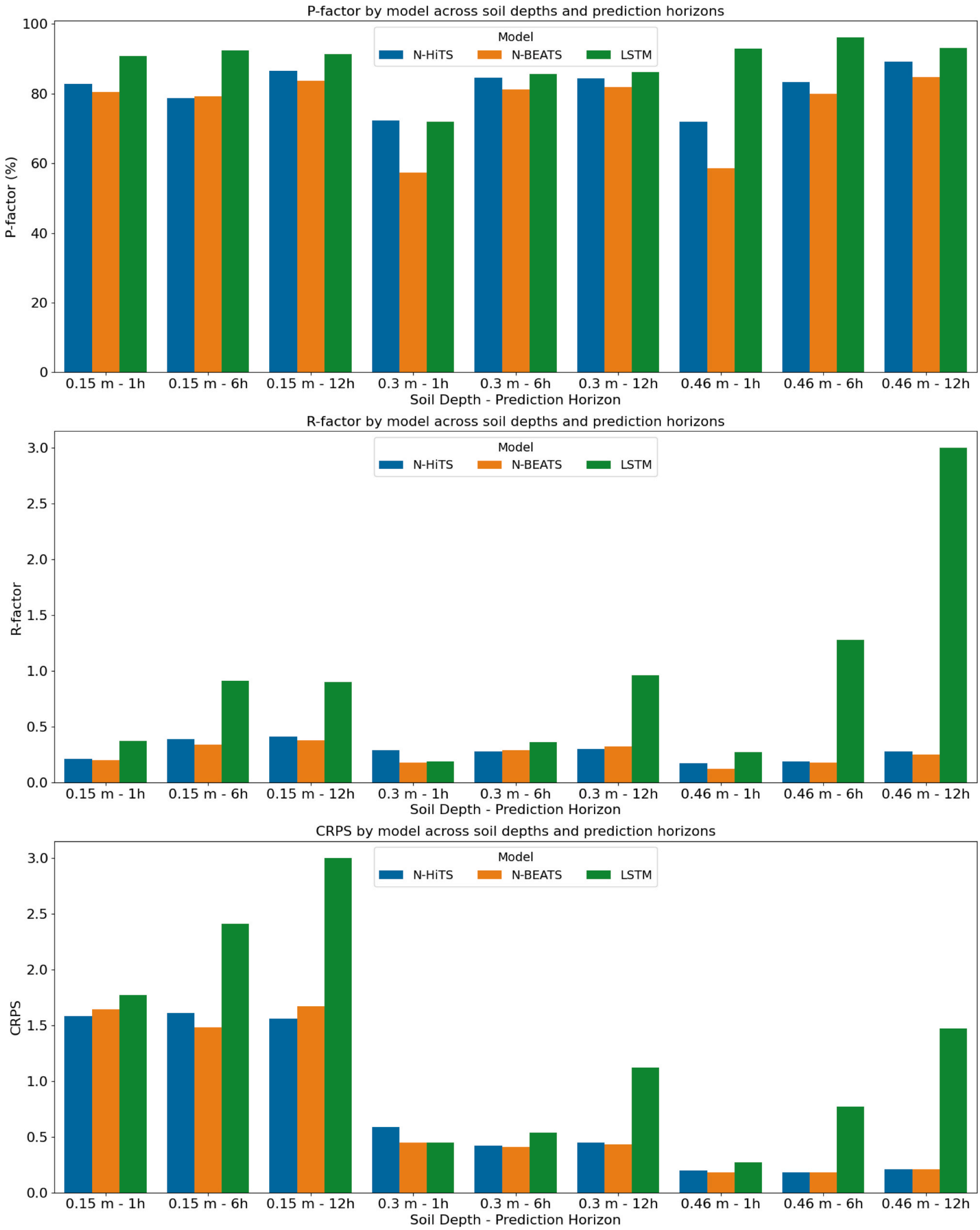


Fig. 7. Uncertainty quantification performance across depths and horizons.

fell within the same range. As the prediction horizon extended to 6 h, the LSTM's residuals' spread increased further, indicating a deterioration in performance, while N-HiTS and N-BEATS displayed a moderate spread of residuals.

At the 12 h horizon, the CDF of LSTM was notably wider where the residuals span from –70 to 75 kPa. N-HiTS and N-BEATS were less affected by the increase in horizons. However, at this stage, the difference between the two is more noticeable as N-BEATS overestimated

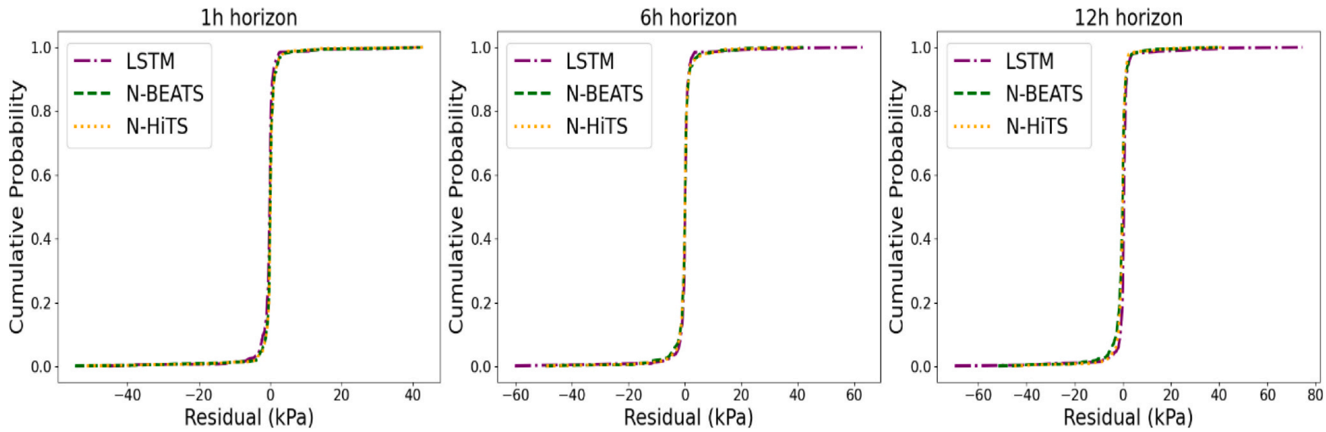


Fig. 8. The CDF curves of residuals across different time horizons at 0.15 m soil depth. The x-axis represents the residual of soil water tension prediction versus the cumulative probability of observing a value less than or equal to soil water tension data on the y-axis.

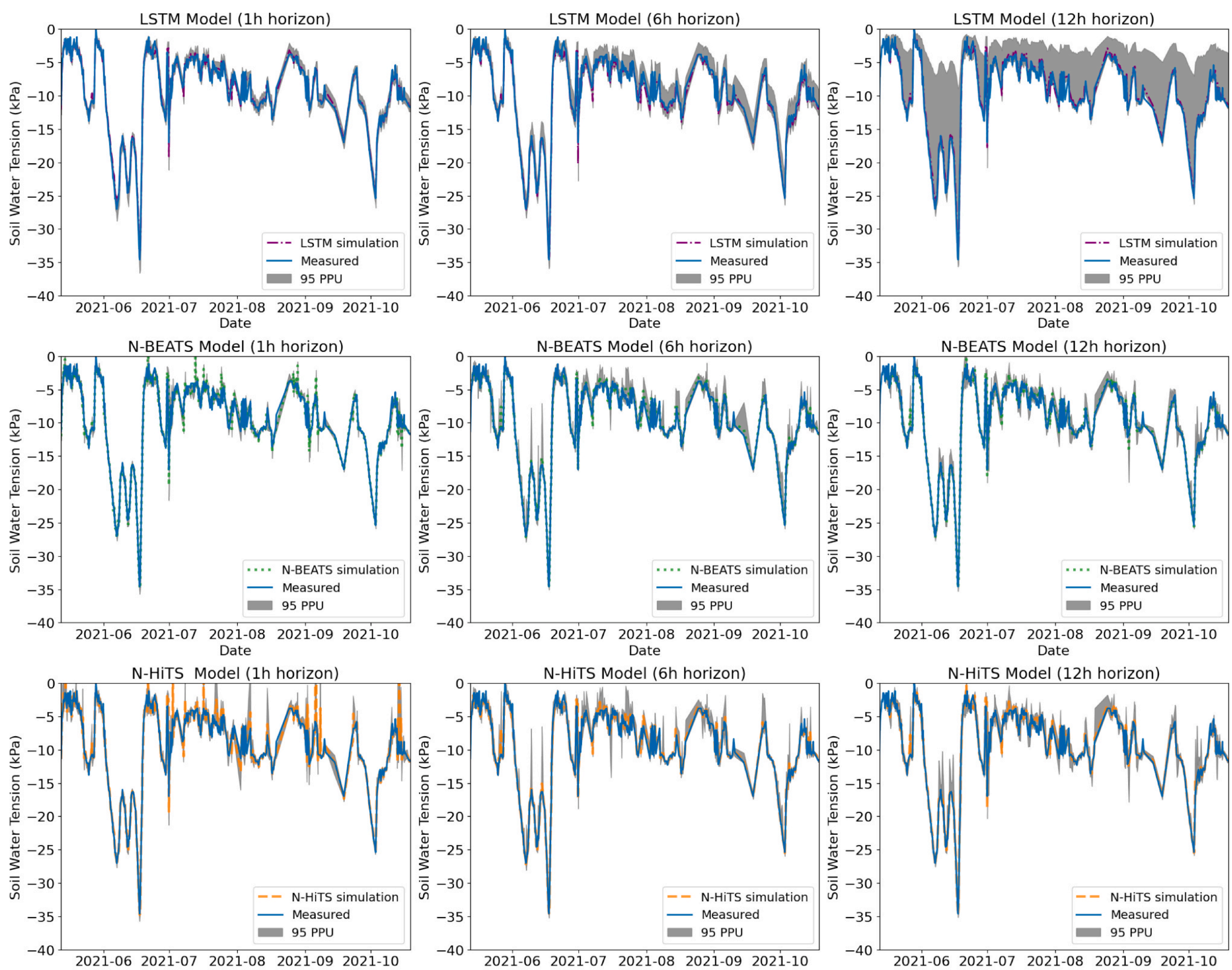


Fig. 9. Soil water tension simulations across different time horizons at 0.3 m soil depth. The grey shaded area indicates 95 PPU. As shown, 95 PPU associated with LSTM is large, indicating that this model did not converge to a stationary simulation over time.

more observed values. Overall, the benchmark model showed significant performance deterioration as the horizon increased; N-BEATS performed reasonably well, but its accuracy decreased with the increasing horizon as evidenced by wider CDF spreads, particularly at 12 h of the horizon. N-HiTS is the least affected model, demonstrating a constant 95 PPU performance across all horizons as characterized by low residuals.

3.2.2. Soil water tension simulation at 0.3 m depth

In addition to predicting soil water tension at 0.15 m soil depth, predictions were also performed for 0.3 m to evaluate how the models

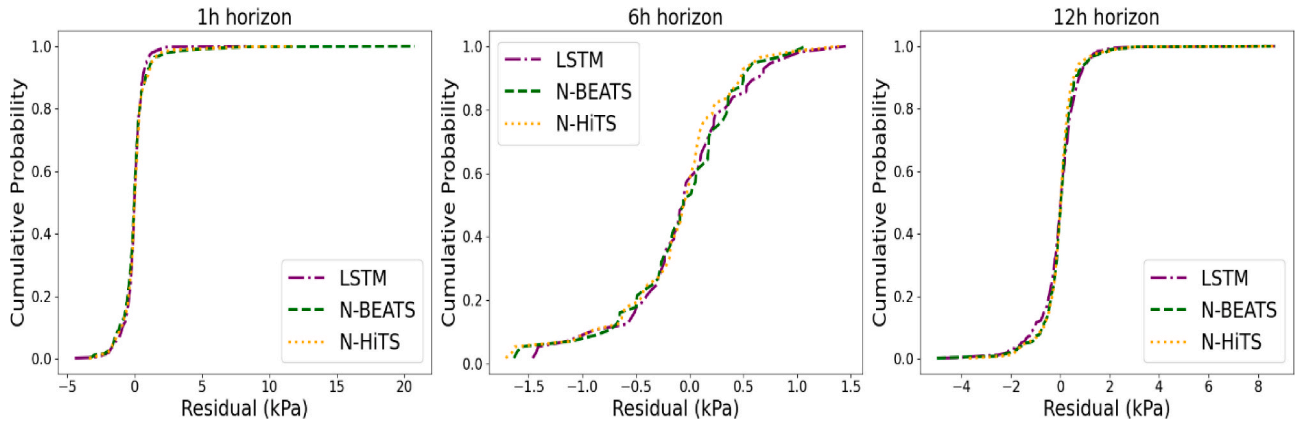


Fig. 10. The CDF curves of residuals across different time horizons at 0.3 m soil depth. LSTM demonstrated an increased residual spread at longer horizons.

perform for different horizons across the root zone. Results illustrated in Fig. 9 revealed that all models were barely affected by this increase, whereby they effectively captured variations from peak to low and medium values. At 1 h horizon, the NSE values (see Table 2) were above 98 %, demonstrating a skillful performance of all models in capturing the variability in the measured data. As the horizon increased, all models maintained good NSE values. Furthermore, noticeable changes were observed in the prediction of peak values, where the models slightly overestimated the peak values, especially at the 12 h horizon. However, the model's performance was more skillful at 0.3 m compared to 0.15 m. This can be explained by the fact that soil water tension at greater depths usually experiences less variability compared to shallower depths. This

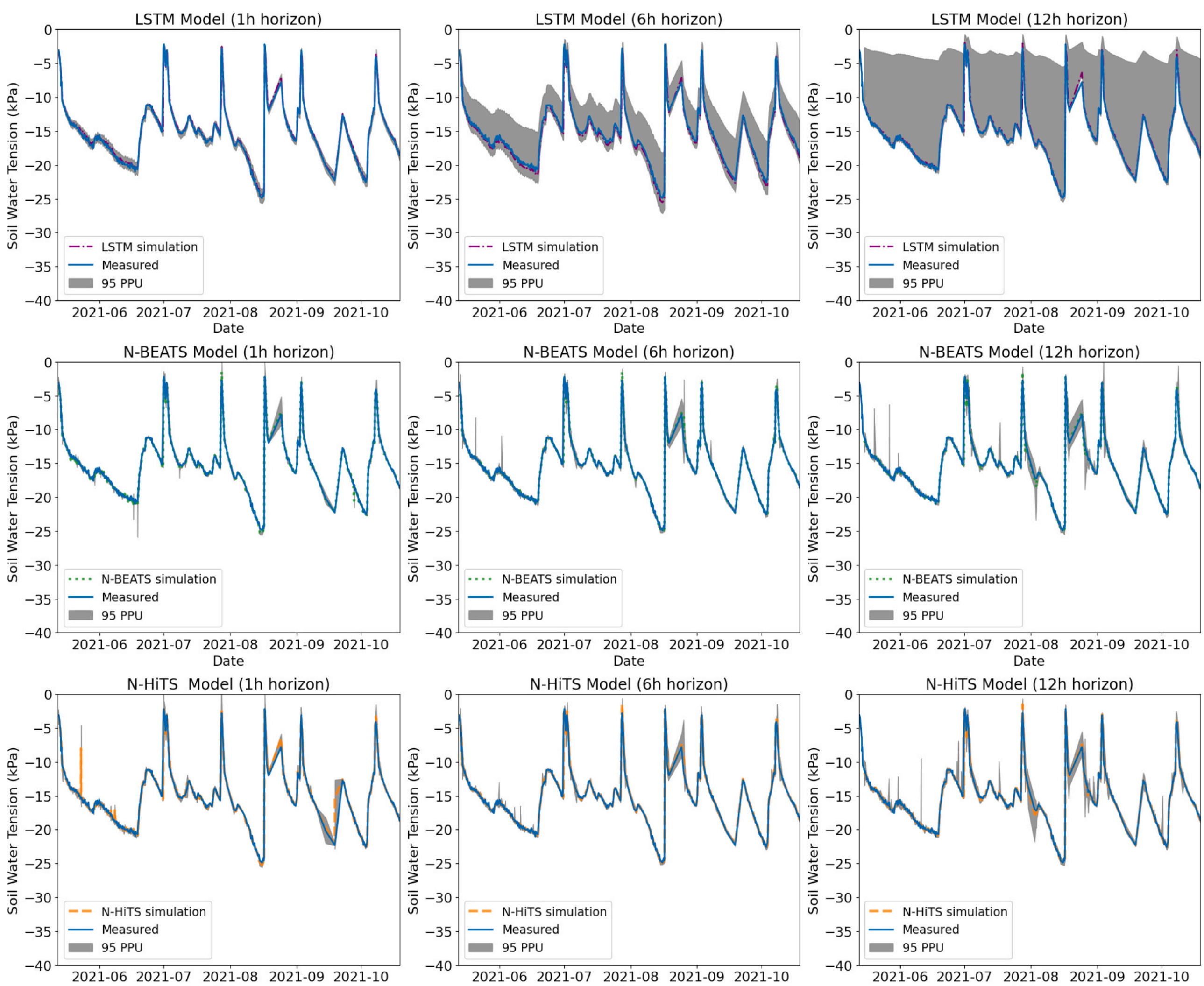


Fig. 11. Soil water tension simulations across different time horizons at 0.46 m soil depth. The grey shaded area indicates 95 PPU. As illustrated, the 95 PPU associated with LSTM simulation didn't converge with a stationary prediction across prediction horizons.

resulted in smoother trends in soil water tension that the models can easily predict, compared to the drastic changes observed at 0.15 m. These outcomes explain the higher performance exhibited by all the models at 0.3 m depth.

Analysis of uncertainty results at 1 h horizon showed that LSTM and N-HiTS performed a 72% P-factor, which is a reasonably high performance compared to N-BEATS. The 0.19 R-factor value for LSTM indicated that this recurrent network was quite confident in its predictions, although LSTM confidence was slightly reduced when there was a rapid fluctuation in soil water tension values. N-HiTS, on the other hand, presented a marginally wider prediction interval than the benchmark model (R-factor equal to 0.29. The wider intervals are particularly apparent around the sudden increase or decrease in soil water tension. N-BEATS presented the narrowest prediction intervals with an R-factor of 0.18, showing higher confidence in its predictions, but to the detriment of not capturing all variability, given its low P-factor. Therefore, although N-BEATS was precise, its lower convergence caused challenges for this algorithm. Arguably, N-HiTS was the most reliable model, harmonizing both prediction accuracy and uncertainty convergence effectively.

At the 6 h horizon, the P-factor values of N-HiTS, N-BEATS and the benchmark model increased considerably, and the models captured more observations using 95 PPU than the 1 h horizon. Typically, the uncertainty bands become more expansive as the horizon rises because the model has a longer time frame to provide predictions, thus introducing more uncertainties. Therefore, the broader bands obtained at the 6 h horizon indicated that the models are less confident in their predictive capabilities due to the accumulation of significant errors over time, especially around sudden transitions from low to high values or vice versa.

Uncertainty estimation results at the 12 h horizon indicated a wider uncertainty band for LSTM compared to the 1 h and 6 h horizons. Although the uncertainty band still encapsulated most of the observational data, a wide 95 PPU indicated that LSTM showed less certainty in soil moisture predictions as the horizon increased. On the other hand, the 95 PPU bands of N-HiTS and N-BEATS models broadened moderately. N-HiTS bracketed more observed data than N-BEATS, as supported by a higher P-factor, although both models exhibited a relatively equal R-factor. Conversely, the benchmark model showed a greater R-factor (0.96) compared to the other two models. This implies that although this algorithm encapsulates most of the measured data it is less capable of providing accurate predictions at higher horizons. Overall, for uncertainty estimation across multiple horizons, N-HiTS provided a more consistent and reliable 95 PPU with a balance between the R-factor and P-factor, encapsulating most of the observed data. Similarly, N-BEATS offered a robust 95 PPU, although this model showed slightly less convergence than N-HiTS.

An analysis of the distribution of errors by each model for each horizon was undertaken based on CDF curves, illustrated in Fig. 10. As shown, significant errors fell around 0 in the 1 h horizon, suggesting good modeling accuracy with relatively small errors. On the x-axis, the CDF spread to about 8 kPa, indicating a slight underprediction of observed values. At 6 h, the curve reduced its steepness, extending more broadly around 0. This indicates the fact that the residuals and variability in data amplified as the horizon increased. At 12 h, the curve slightly shifts to the right and shows a consistent increase in residuals extending to about 8 kPa. At the 1 h horizon, the N-HiTS curve behaved similarly to LSTM, with a slight tendency towards higher residuals. This suggests the fact that a neural hierarchical interpolation approach presented slightly more variability compared to LSTM at this prediction horizon. The N-BEATS curve also indicated a similar trend but with a broader range of errors at 1 h horizon compared to N-HiTS and LSTM. At the 6 h horizon, the N-HiTS model exhibited a more pronounced shift, indicating a significant increase in errors compared to the shorter horizon. A similar pattern is observed for the N-BEATS model, with a rightward shift in the CDF at 6 h compared to the 1 h horizon. In the 12 h horizon, N-HiTS residual errors spread to about 9, similar to the N-BEATS model.

3.2.3. Soil water tension simulation at 0.46 m depth

In addition to predicting soil water tension at 0.15 m and 0.3 m, this variable was also predicted at 0.46 m, taking the 1 h, 6 h and 12 h as prediction horizons. Fig. 11 shows the predicted and observed soil water tension on each horizon. The N-BEATS model performed well across the three horizons, indicating insignificant depreciation in its predictive capabilities as the horizons increased. Similarly, the N-HiTS model yielded good results for all three horizons, displaying a slight decrease in its prediction accuracy with an increase in the horizon but showcasing compelling prediction capabilities as the prediction was closely aligned with the observation. The performance metrics in Table 2, indicate that N-BEATS and N-HiTS performed similarly at the 6 h horizon where both models produced equivalent NSE, RMSE, and MSE values. In addition, these algorithms also performed equally in the 1 h and 12 h horizons. Conversely, the performance of the benchmark model gradually decreased, as underscored by the discrepancies between the measured and predicted soil water tension with increasing horizon (Fig. 11).

The uncertainty in models' predictions was comprehensively assessed across all horizons as illustrated in Fig. 11. For the 1 h horizon, the uncertainty band for N-HiTS adequately bracketed about 72 % of the observed. Among all models, N-BEATS encapsulated the lowest percentage of the measured data (58.64 %) but presented the narrowest uncertainty band. On the contrary, LSTM provided the highest P-factor

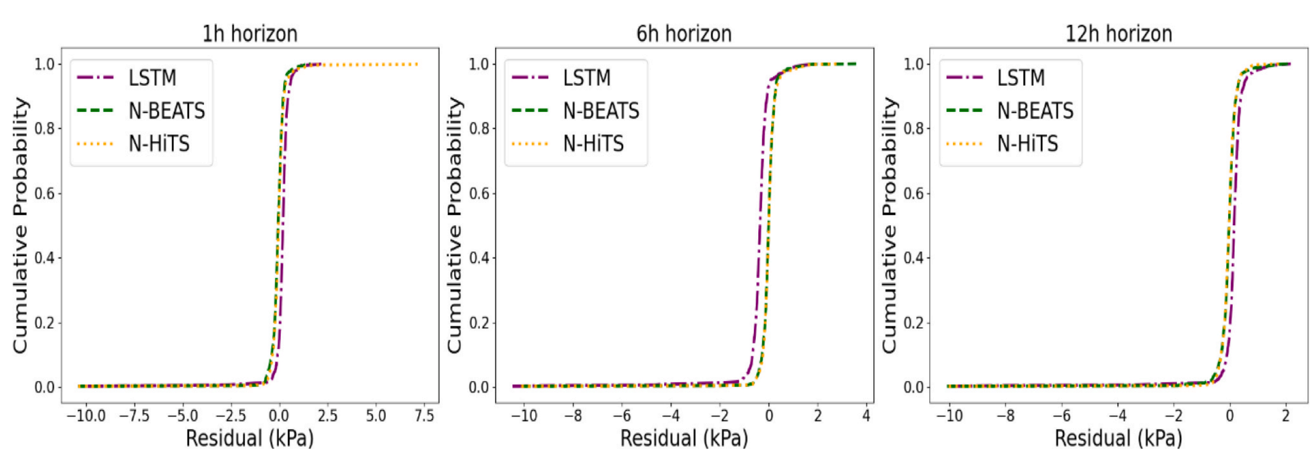


Fig. 12. The CDF curves show residuals across different time horizons at 0.46 m soil depth. As shown, the residual error curve of LSTM is slightly different than the other two models.

Table 3
Uncertainty estimation metrics across different depths and horizons for different confidence intervals. The best performances are shown in bold.

Model	Soil depth	Horizon	P-factor	R-factor	P-factor	R-factor	P-factor	R-factor	P-factor	R-factor
Confidence interval			**50**		**75**		**90**		**99**	
N-HiTS	0.15 m	1 h	26.41	**0.07**	42.14	**0.11**	57.28	0.17	80.58	0.34
N-BEATS			26.99	0.09	40.00	**0.12**	50.29	**0.16**	56.12	**0.31**
LSTM			**59.61**	0.1	**73.59**	0.17	**77.28**	0.2	**98.06**	1.35
N-HiTS		6 h	36.79	0.1	56.36	0.22	71.62	0.33	90.61	0.57
N-BEATS			35.81	**0.08**	55.77	0.18	73.39	**0.27**	91.98	**0.54**
LSTM			**56.56**	**0.08**	**69.28**	**0.16**	**87.87**	0.48	**96.09**	1.62
N-HiTS		12 h	**43.37**	**0.09**	67.87	**0.17**	80.52	**0.32**	93.37	0.65
N-BEATS			35.94	0.1	57.83	0.19	79.12	**0.32**	93.98	**0.63**
LSTM			31.53	0.11	**70.48**	0.20	**89.16**	0.67	**97.99**	1.8
N-HiTS	0.3 m	1 h	**29.71**	0.13	42.72	0.18	58.06	0.25	79.42	0.47
N-BEATS			28.16	**0.07**	38.06	0.1	48.74	**0.14**	70.10	**0.23**
LSTM			20.19	0.04	**52.43**	**0.01**	**65.05**	**0.14**	**95.73**	1.28
N-HiTS		6 h	30.14	**0.07**	**55.58**	**0.12**	73.19	0.21	89.04	0.49
N-BEATS			**34.05**	**0.06**	54.99	0.13	72.21	**0.2**	90.22	**0.42**
LSTM			29.75	0.07	52.05	**0.12**	**82.97**	0.27	**97.06**	1.56
N-HiTS		12 h	**43.98**	**0.07**	**61.45**	0.13	**75.10**	0.24	91.16	0.5
N-BEATS			38.55	**0.07**	60.84	0.13	74.30	0.22	89.56	**0.46**
LSTM			22.89	0.06	51.81	**0.12**	64.86	**0.21**	**92.17**	1.83
N-HiTS	0.46 m	1 h	25.83	**0.04**	42.33	**0.07**	**62.14**	0.13	87.57	0.33
N-BEATS			25.83	0.05	39.42	**0.07**	51.07	**0.1**	73.79	**0.2**
LSTM			**64.85**	0.11	**89.71**	0.21	92.43	0.29	**96.50**	3.46
N-HiTS		6 h	**40.31**	0.05	59.69	0.1	76.71	0.15	94.72	0.34
N-BEATS			35.03	**0.04**	56.16	**0.09**	**72.80**	**0.14**	93.15	**0.31**
LSTM			20.55	0.09	**74.36**	0.22	**92.56**	0.36	**99.22**	3.72
N-HiTS		12 h	36.14	**0.05**	59.64	**0.1**	79.52	**0.2**	94.98	**0.47**
N-BEATS			**36.95**	**0.05**	57.03	**0.1**	76.10	**0.2**	**95.38**	0.48
LSTM			30.12	0.04	**74.90**	0.12	84.14	0.3	**85.74**	4.01

(93.13 %), bracketing more observed values within 95 PPU compared to N-HiTS and N-BEATS. By achieving a considerably higher P-factor and a slightly higher R-factor, the benchmark model outperformed N-HiTS and N-BEATS at the 1 h horizon. The recurrent network skillfully maintained a balance between capturing more than 90 % of the measurements while maintaining a relatively good R-factor.

For the 6 h horizon, a moderate increase in the uncertainty bands of N-HiTS and N-BEATS was observed. On the contrary, LSTM's uncertainty band increased significantly compared to the 1 h horizon; this reveals less certainty of the benchmark model for the 6 h horizon. Overall, N-HiTS and N-BEATS were less affected by the transition from shorter to longer horizons as indicated by R-factor, while LSTM allowed more variability with a rise in R-factor from 0.27 to 1.28.

At the 12 h horizon, LSTM's uncertainty band increased considerably compared to N-HiTS and N-BEATS, particularly during high soil water tension values. For N-HiTS, the uncertainty band widened during rapidly changing soil water tension values. Unlike N-HiTS, N-BEATS maintained a closer fit to the measured values, as reflected by its lowest R-factor. Regarding data convergence, N-BEATS bracketed the smallest percentage of observed data and showed slightly less convergence than N-HiTS and the benchmark model. The benchmark model proved excellent in bracketing most observations by maintaining a good to excellent P-factor. However, LSTM's R-factor increased from 1.28 at 6 h to 3.00 at 12 h, reducing confidence in the predictions as horizon size increased. A noticeable rise in R-factor suggested that LSTM might be less reliable for long-term predictions due to the rapid depreciation of its accuracy across longer horizons. On the contrary, N-HiTS and N-BEATS offered robust and reliable predictions with a lower R-factor across all horizons. In other words, their predictions were closely clustered around the observed data, and their high P-factor reflected their ability to capture a considerable percentage of data points while maintaining a narrow uncertainty band.

The CDF curves of all models were plotted to explicitly determine changes in the performance of each model at different horizons. Fig. 12 indicates that at the 1 h horizon, the simulation results of N-HiTS, N-BEATS and the benchmark models were closely aligned while their residuals clustered around zero. However, the benchmark model presented a somewhat steeper slope, implying a minor variance in error

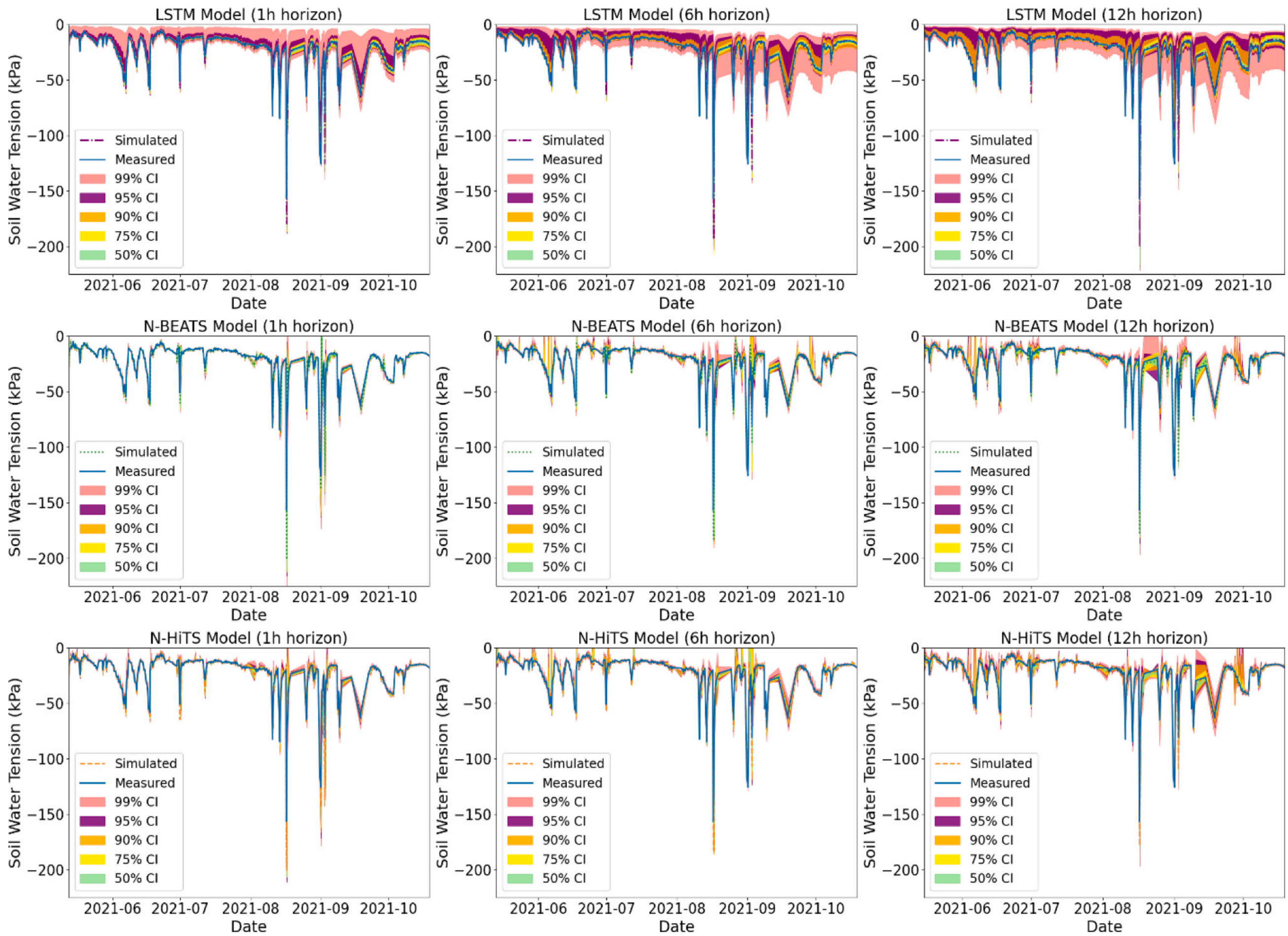


Fig. 13. The 50 %, 75 %, 90 %, 95 %, and 99 % uncertainty bands across different time horizons at 0.15 m soil depth. As shown, lower soil water tension values are better bracketed by lower uncertainty bands (i.e., 50 % and 75 %) while higher values are mostly bracketed by > 90 % uncertainty bands.

estimation compared to the other two models. At the 6 h horizon, the CDF curves maintained a steep ascent but spread slightly wider than the 1 h horizon. LSTM's curve differed somewhat from those of N-HiTS and N-BEATS, indicating that this algorithm was unable to capture trends as effectively as the other two models. As the horizon extended to 12 h, the benchmark model showed a more significant decline in performance compared to N-HiTS and N-BEATS. N-BEATS and N-HiTS performed equally, with minimal difference between the spread of their respective errors.

Furthermore, analysis of the uncertainty at various 95PPU intervals revealed key insights into the capabilities of all models to balance convergence and accuracy. According to the results, N-HiTS and N-BEATS maintained lower R-factor values than LSTM across all the 95PPU (see Table 3). LSTM, on the other hand, presented better convergence across most depths and horizons, encompassing most of the observations within its uncertainty boundaries. However, this resulted in a drastic rise in R-factor values, leading to considerably wide intervals that reduced accuracy. While N-HiTS and N-BEATS increased their respective uncertainty bands to improve convergence, they sustained a better steadiness between convergence and accuracy by exhibiting R-factor values significantly lower than LSTM, particularly at the 12 h horizon. N-BEATS and N-HiTS demonstrated their capabilities to maintain accurate predictions across depths and prediction horizons. This led to effective simulation of low and high soil water tension values (see Fig. 13, Fig. 14, and Fig. 15). On the contrary, the broad uncertainty bands indicated that LSTM is conservative in its predictions as the timeframe increases by overestimating the soil water tension.

The differences in the exhibited performance stem from the distinctive architectural design of N-HiTS and N-BEATS. N-HiTS generally performed better than N-BEATS at longer horizons, mainly due to its unique multi-rate sampling technique. This approach breaks down complex temporal patterns in the data into fine-grained, more predictable patterns that capture fluctuations more efficiently. Depending on the kernel size, each stack in N-HiTS is tailored to learn either high or low frequency data components. These structural differences make N-HiTS particularly effective when soil water tension shows high variability. On the other hand, N-BEATS is better suited for modeling soil water tension of repetitive patterns such as trends and seasonality.

Furthermore, both N-HiTS and N-BEATS consist of multiple layers of stacks and blocks, producing a backcast and forecast at each timestep. The backcast indicates parts of the input data that are effectively learned by a block and uncovers remaining features that the following blocks should emphasize to minimize residuals. Therefore, only a portion of the data that antecedent blocks have not effectively captured is passed on to the next. This process enhances both models' prediction capability, allowing them to adapt to fluctuations in data across different depths and horizons by limiting the aggregation of errors. In contrast, LSTM uses a recursive prediction approach, where each prediction is fed as input to the next time-step. As a result, errors propagate through the network across timesteps, culminating in compounded inaccuracies and reduced prediction accuracy as horizon increases. This error accumulation contributes to the wider uncertainty bands observed for LSTM, particularly in long-term predictions.

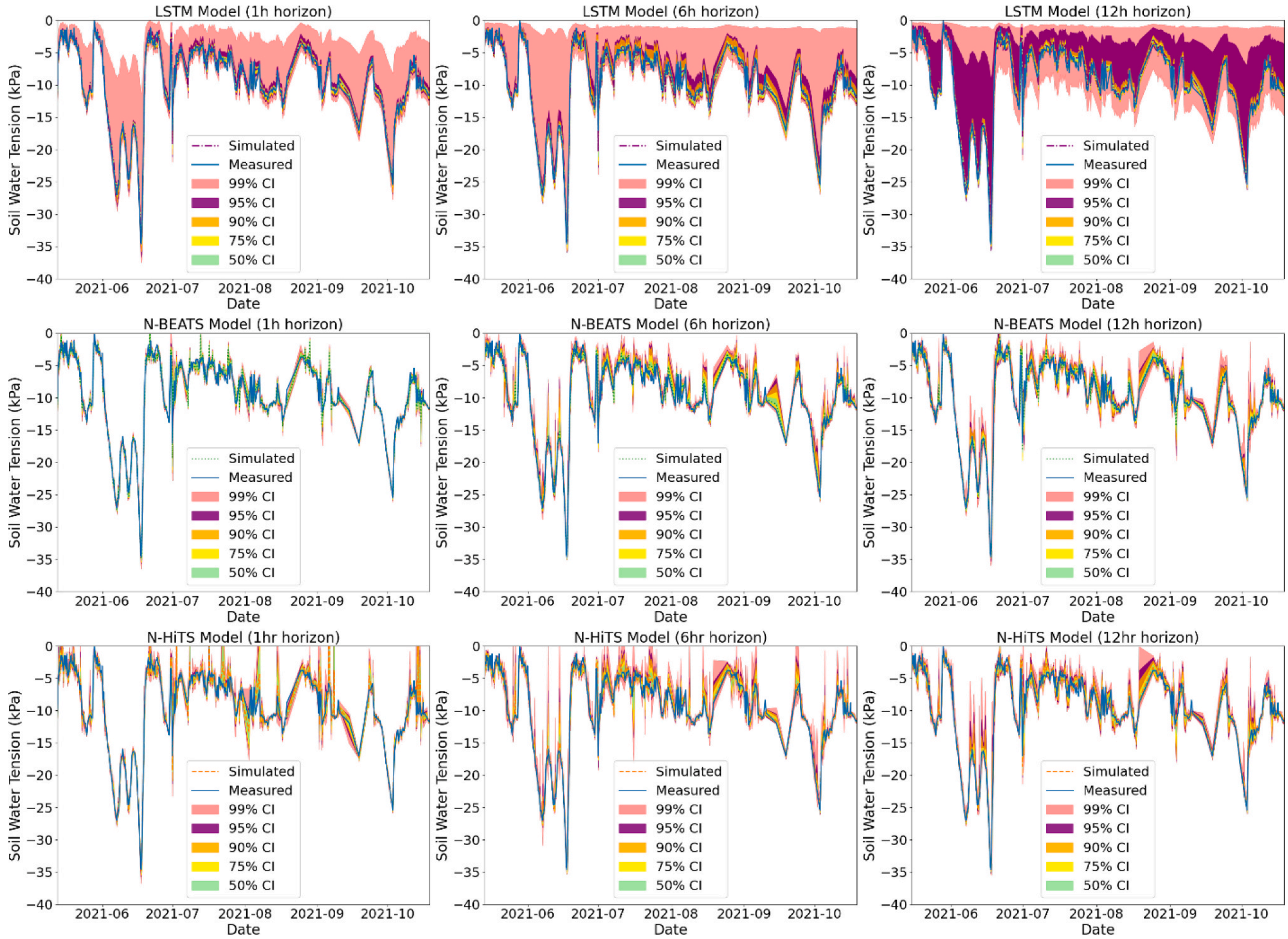


Fig. 14. The 50 %, 75 %, 90 %, 95 %, and 99 % uncertainty bands across different time horizons at 0.3 m soil depth. As illustrated, all uncertainty bands showed significant fluctuations, revealing high variability in observational data.

3.3. SC irrigated field case study

3.3.1. Soil water tension simulation at 0.15 m depth

In addition to testing N-HiTS and N-BEATS on their capabilities to predict soil water tension at different time horizons for a field in GA, they were also tested using data from an irrigated field in SC. Fig. 16 presents the simulated against predicted soil water tension for the 1 h, 6 h, and 12 h horizons at 0.15 m of soil depth. At the 1 h horizon, N-HiTS and N-BEATS performed equally, as indicated by their respective performance metrics. Both models outperformed the benchmark model as shown in Table 4. At the 6 h horizon, the benchmark model was the most affected by the increase in the horizon as reflected by increasing RMSE and MAE values from 2.68 to 3.85 and 1.23 to 1.98, respectively. On the other hand, the RMSE and MAE values of the N-HiTS model for the 6 h horizon increased by 0.47 and 0.14, respectively, while those of N-BEATS increased by 0.72 and 0.19. Similarly to the 6 h horizon, N-HiTS and N-BEATS outperformed LSTM at the 12 h horizon, maintaining their NSE values at 99 % while LSTM's NSE value reduced to 96 %. This result indicates a notably higher performance of N-HiTS and N-BEATS compared to the benchmark model. The accurate prediction of various soil water tension ranges, particularly extreme values, demonstrates the capabilities of N-HiTS and N-BEATS models in supporting irrigation decisions based on soil moisture dynamics.

The 95 PPU at 0.15 m soil depth for the 1 h time horizon indicated that the LSTM model produced a broader uncertainty band compared to N-HiTS and N-BEATS. This was most noticeable around periods of sudden changes such as transitions from low to high soil water tension in September. N-HiTS and N-BEATS, on the other hand, presented a very narrow uncertainty band during the same period. They also provided the highest P-factor of 92 %, and a narrower R-factor, outperforming the benchmark model. At the 6 h horizon, both N-HiTS and N-BEATS models maintained their accuracy despite the increase of time horizon, as underscored by the extent of their respective uncertainty bands and metrics (Fig. 16). On the other hand, LSTM's uncertainty band further widened. This suggested that the benchmark model was less confident in soil water tension prediction as the horizon increased. At the 12 h horizon, there was a noticeable increase in the uncertainty bands of N-HiTS and N-BEATS. These two models showed less confidence in their predictions on this horizon due to more uncertainty being introduced over time. However, both models maintained a closer fit to the observed data. Conversely, the accuracy of the benchmark model declined significantly compared to N-HiTS and N-BEATS. The narrow confidence bands around N-HiTS and N-BEATS predictions showcased their reliability for long-term prediction compared to the benchmark model, particularly in the events of rapid soil–water tension changes, where LSTM appeared to fail at maintaining a narrower uncertainty band. Furthermore, the comparison of CRPS metrics demonstrated that LSTM's predictions differ more significantly from measured values than those of N-HiTS and N-BEATS (see Fig. 17).

The CDF plots of residuals across the 1 h, 6 h, and 12 h horizons (Fig. 18) demonstrated the reliability of N-BEATS and N-HiTS models in predicting soil water tension at 0.15 m of soil depth. As the horizon increased from 1 h to 12 h, a slight expansion of the residual spread was observed for both models. The benchmark model displayed a less concentrated residual distribution, evidenced by a more gradual curve that spreads across a more extensive residual range. This implies that

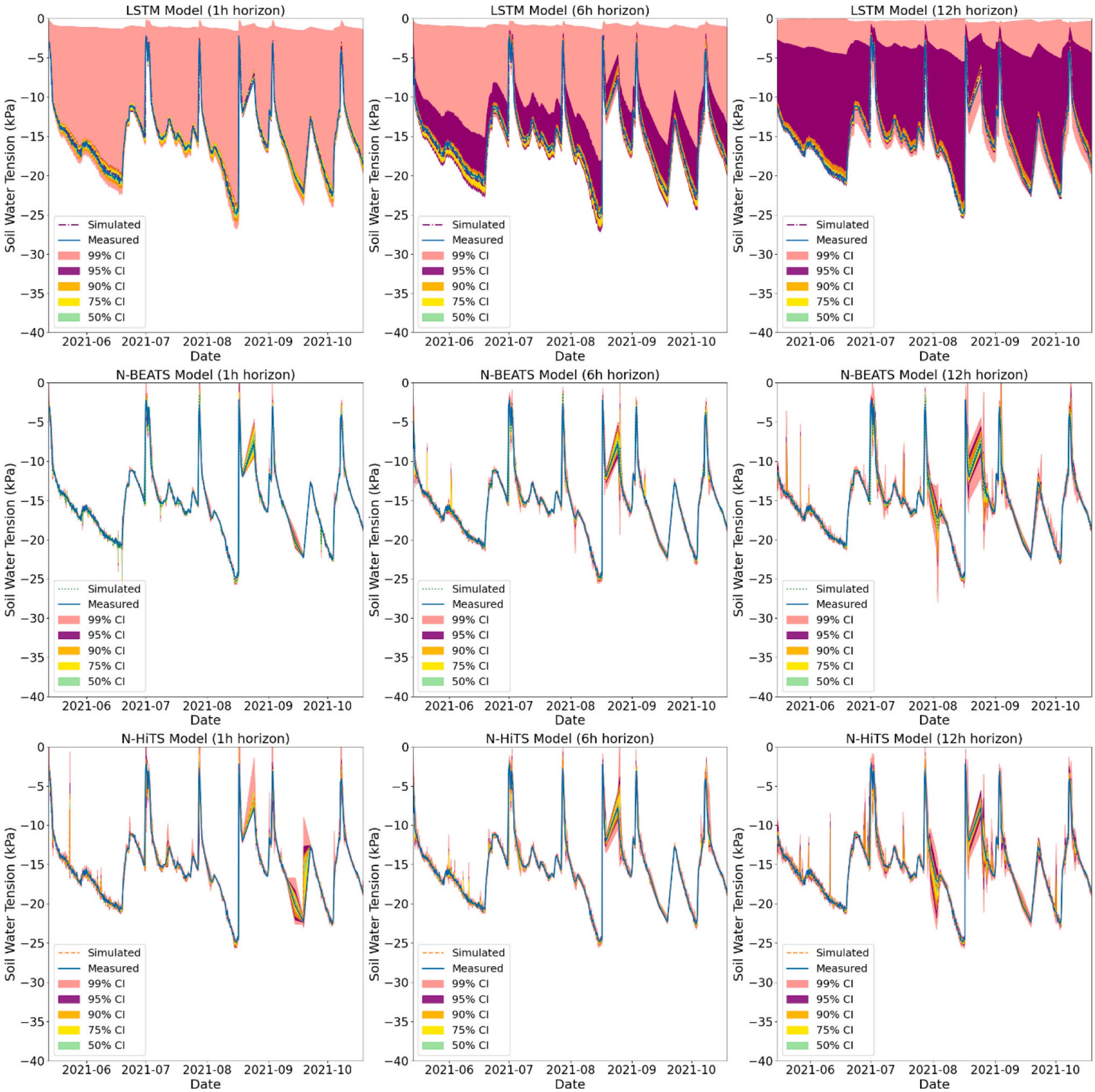


Fig. 15. The uncertainty bands of 50 %, 75 %, 90 %, 95 % and 99 % across different time horizons at 0.46 m soil depth. As shown, there is more agreement among different uncertainty bands, particularly during peak soil water tension predictions.

LSTM tended to produce more significant errors than N-HiTS and N-BEATS. As a result, the benchmark model exhibited lower reliability, as indicated by a wider spread of residuals around zero. N-HiTS and N-BEATS models exhibited a steep curve closely centered around zero residuals, indicating that the prediction errors were insignificant. This sharp ascent around zero suggested that N-HiTS and N-BEATS provided highly accurate predictions with minimal error spread, maintaining consistency across all horizons.

3.3.2. *Soil water tension simulation at 0.3 m depth*

Soil water tension predictions at 0.3 m depth across different models indicated that all algorithms produced predictions that align well with measured values. For the 1 h horizon, N-HiTS and N-BEATS presented near-optimum NSE scores of 99.71 % and 99.73 %, respectively, and maintained a low RMSE and MAE. High P-factor and low R-factors reflect a narrow uncertainty band, implying high confidence of N-HiTS and N-BEATS in predicting soil water tension (see Fig. 19). In comparison, the benchmark model achieved a lower NSE with higher RMSE and MAE metrics, indicating more significant prediction errors.

In addition, N-HiTS and N-BEATS performed well as the time horizon increased to 6 h, as shown in Fig. 19. Both models maintained a strong performance, with NSE scores of 99.36 % and 99.41 % and minor increases in RMSE and MAE, indicating that these models retained accuracy over extended horizons. Their P-factors also remained high, although slightly lower than the 1 h horizon. The benchmark model, on the other hand, experienced a notable decline in accuracy with an NSE of 92.61 % and much higher RMSE and MAE.

For the 12 h horizon, N-HiTS and N-BEATS continued to show strong predictive power with NSE scores of 99.28 % and 99.00 %, respectively. Although, their uncertainty bands widened slightly, as indicated by a

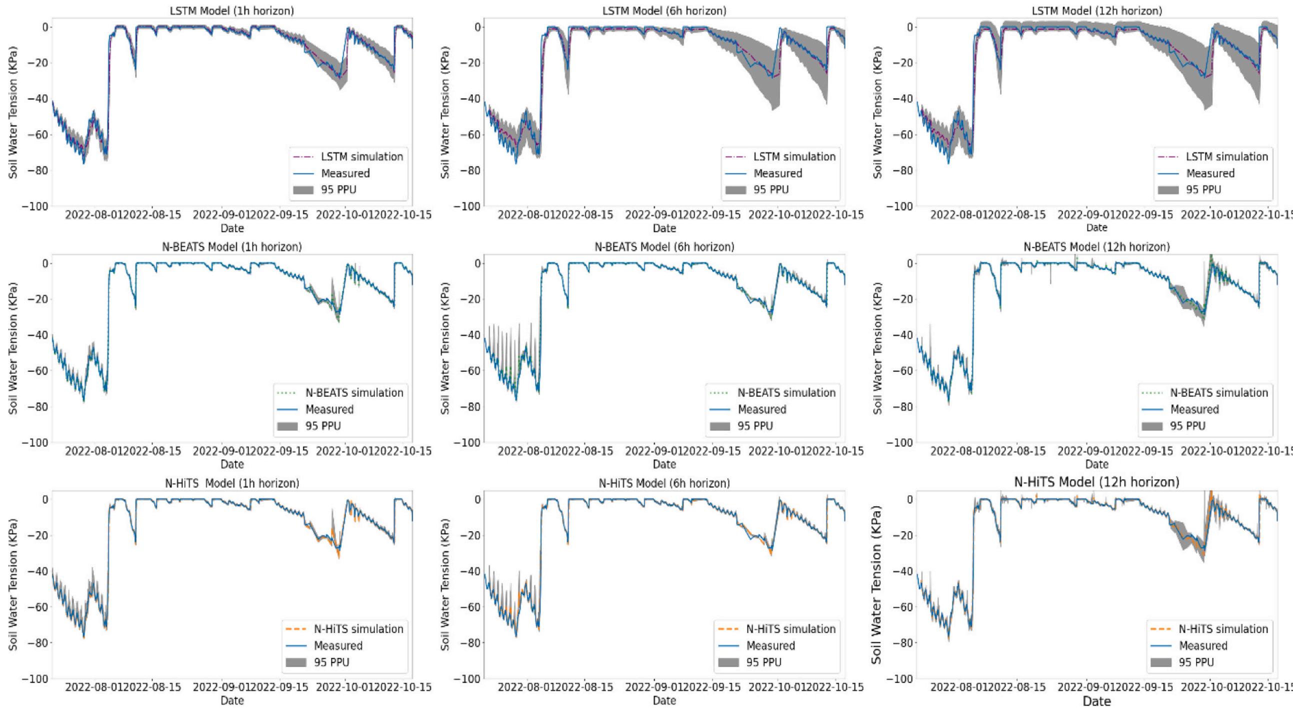


Fig. 16. Soil water tension simulations across different time horizons at 0.15 m soil depth. The grey shaded area indicates 95 PPU. As shown, the 95 PPU of N-HiTS and N-BEATS are skillfully optimized, while the 95 PPU associated with LSTM did not converge with a stationary state, particularly for the 6 h and 12 h horizons.

Table 4
Performance metrics for soil water tension prediction for irrigated field in SC. The best performances are shown in bold.

Model	Soil depth	Horizon	NSE (%)	RMSE	MAE	P-factor (%)	R-factor	CRPS
N-HiTS	0.15 m	1 h	**99.90**	**0.69**	**0.26**	**92.22**	**0.075**	0.3
N-BEATS			**99.91**	**0.68**	**0.26**	92.00	**0.05**	**0.25**
LSTM			98.54	2.68	1.23	90.29	0.21	1.1
N-HiTS		6 h	**99.72**	**1.16**	**0.40**	82.46	**0.06**	**0.35**
N-BEATS			99.60	1.40	0.45	84.19	**0.09**	0.39
LSTM			96.95	3.85	1.98	**86.04**	0.40	1.51
N-HiTS		12 h	**99.85**	**0.85**	**0.35**	89.80	**0.08**	0.34
N-BEATS			99.83	0.91	0.37	90.30	**0.07**	**0.33**
LSTM			96.95	3.84	2.02	91.70	0.5	1.51
N-HiTS	0.3 m	1 h	99.71	1.45	0.279	91.72	**0.06**	**0.28**
N-BEATS			**99.73**	**1.41**	**0.274**	93.43	**0.06**	**0.28**
LSTM			96.57	5.03	2.44	82.89	0.56	2.18
N-HiTS		6 h	99.36	2.17	**0.41**	**87.44**	**0.07**	**0.35**
N-BEATS			**99.41**	**2.1**	**0.41**	82.46	**0.07**	0.39
LSTM			92.61	7.4	3.38	57.46	0.46	1.51
N-HiTS		12 h	**99.28**	**2.3**	**0.49**	84.50	0.14	0.66
N-BEATS			99.00	2.71	0.64	78.42	0.10	**0.63**
LSTM			93.09	7.13	3.56	**86.51**	0.65	3.98
N-HiTS	0.6 m	1 h	**99.72**	**0.69**	**0.30**	89.02	**0.07**	**0.25**
N-BEATS			**99.71**	**0.70**	**0.31**	**90.07**	**0.08**	0.26
LSTM			95.30	2.81	1.83	87.42	0.36	1.33
N-HiTS		6 h	**99.67**	**0.7**	**0.32**	79.82	**0.05**	**0.27**
N-BEATS			**99.66**	**0.7**	**0.32**	**84.30**	**0.06**	**0.27**
LSTM			93.24	2.81	1.89	89.52	0.45	1.48
N-HiTS		12 h	**99.64**	**0.79**	**0.35**	85.65	**0.09**	0.3
N-BEATS			99.59	0.85	**0.36**	84.79	**0.08**	**0.28**
LSTM			91.56	3.84	1.69	**90.76**	0.64	1.73

moderate R-factor. On the contrary, the benchmark model's performance decreased significantly, with an NSE of 93.09 % and much higher RMSE and MAE values, indicating significant deviation from observed values. Furthermore, N-HiTS and N-BEATS predictions aligned closely with the observed values across all horizons (see Fig. 19) while the benchmark model showed a noticeable lag and a wider uncertainty band, particularly at 6 and 12 h horizons. Overall, N-HiTS and N-BEATS demonstrated improved performance at 0.3 m soil depth, making these models more suitable for long-term soil water tension predictions than LSTM.

The CDF plots of the residuals produced by the models are illustrated in Fig. 20. These results indicated that the performance and accuracy of models differed as the time horizon increased. At the 1 h horizon, the CDF curves of N-HiTS and N-BEATS models approached zero rapidly and centered around zero residuals, indicating that these two models provided a limited number of underpredicted values and that a high percentage of errors was close to zero. The models, however, showed a wide spread of positive residuals. For example, at 6 h and 12 h horizons, the

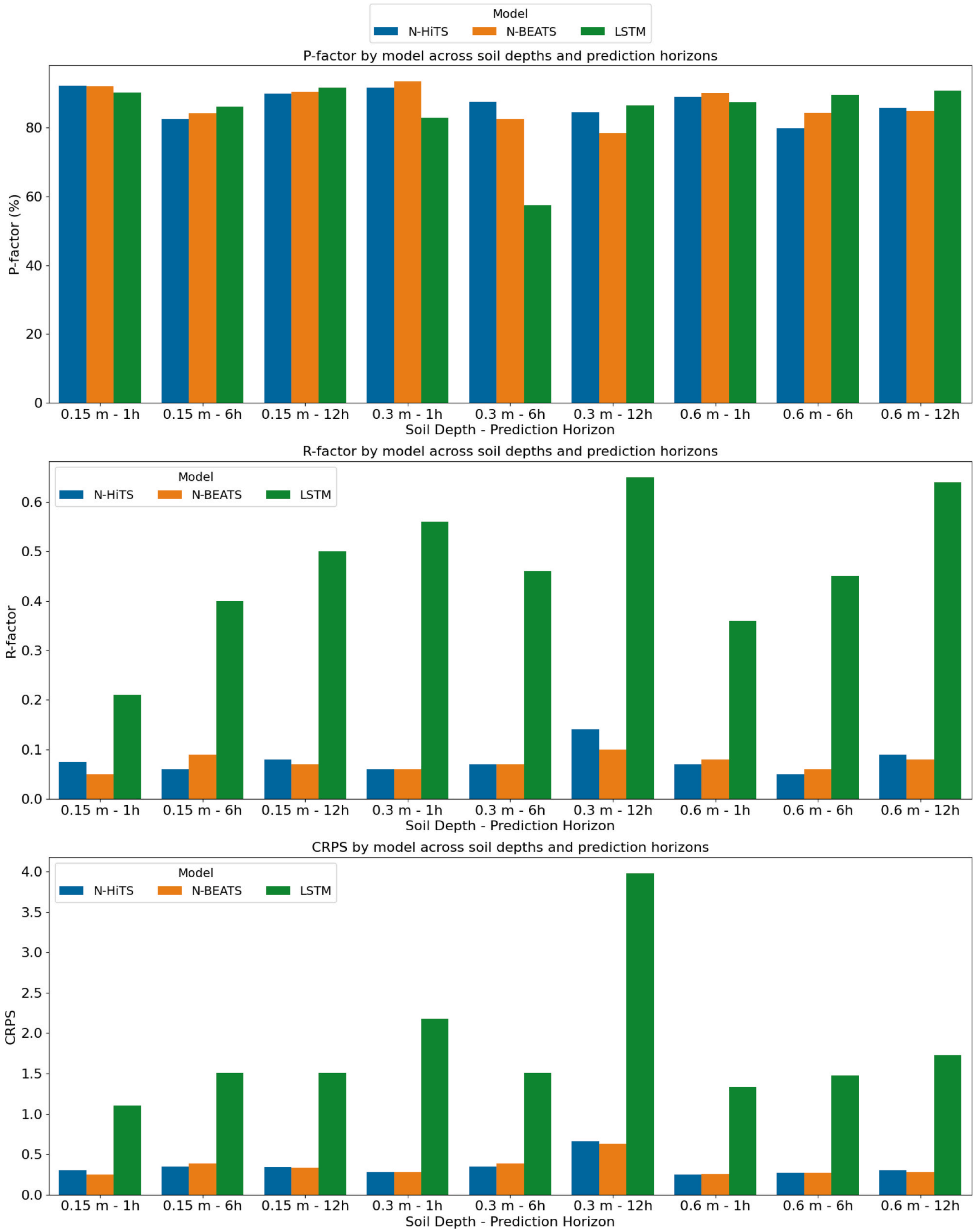


Fig. 17. Uncertainty quantification performance across different depths and horizons.

curves spread further, particularly on the negative side of residuals. This highlights that as the horizon increased, the accuracy reduced, thus the observed rise in the percentage of underpredicted values. However, the curves maintained their steepness. This consistency implies that N-HiTS and N-BEATS provided acceptable accuracy and reliability despite the extension of the prediction time frame. On the contrary, the benchmark model showed a gentle curve with errors extending over an extensive range of values. This implies that LSTM tended to generate predictions

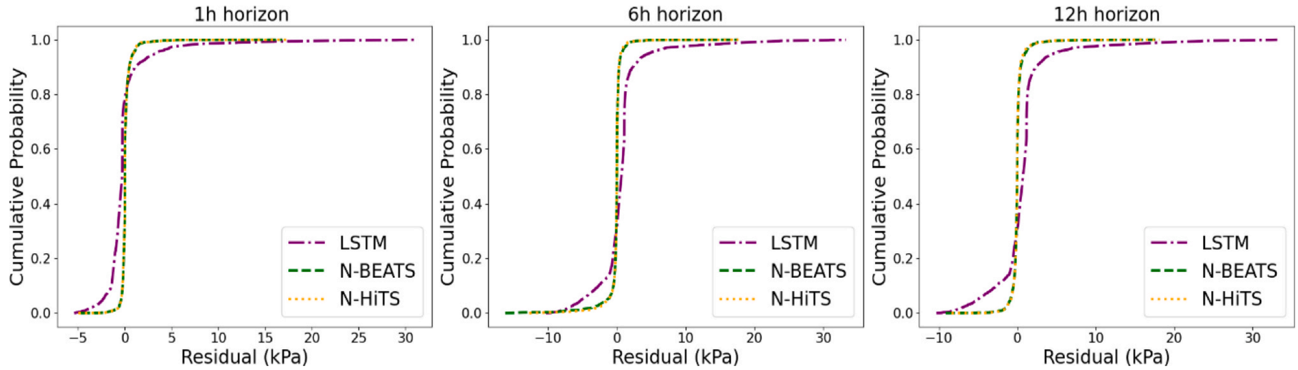


Fig. 18. The CDF curves of residuals across different time horizons at 0.15 m soil depth.

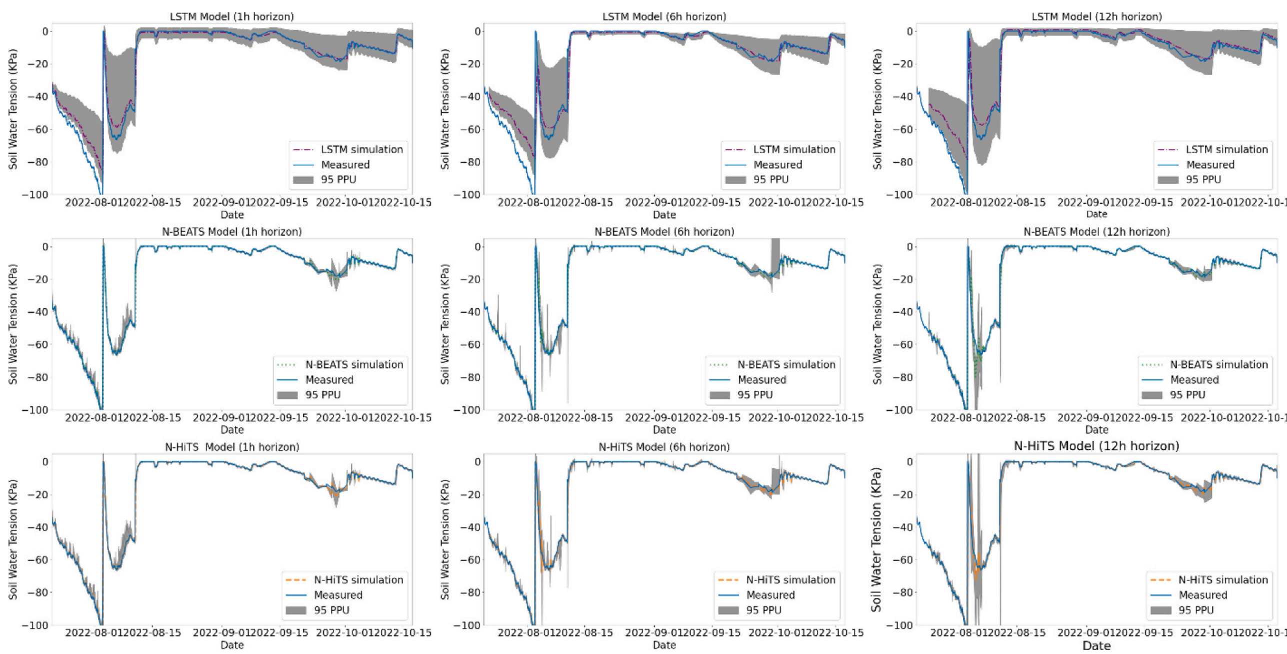


Fig. 19. Soil water tension simulations across different time horizons at 0.3 m soil depth. The grey shaded area indicates 95PPU.

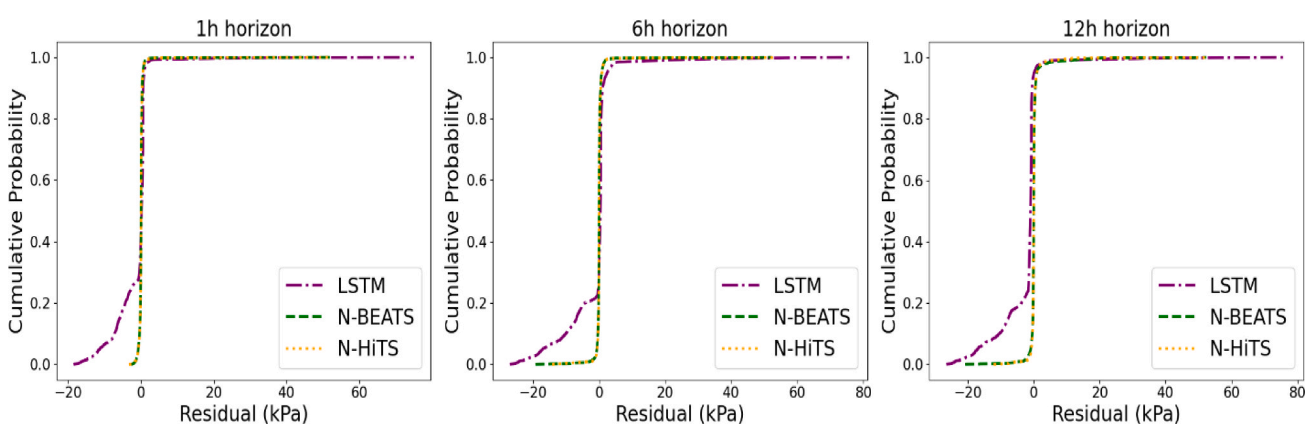


Fig. 20. The CDF curves of residuals across different time horizons at 0.3 m soil depth.

that significantly overpredict or underpredict the actual values.

3.3.3. Soil water tension simulation at 0.6 m depth

The result of soil water tension prediction at the 0.6 m soil depth revealed different performance trends for the N-HiTS and N-BEATS compared to LSTM across three horizons (see Fig. 21). At the 1 h horizon, both N-HiTS and N-BEATS demonstrated very high predictive accuracy, with NSE values above 99.7 % and low RMSE and MAE values. These results indicated excellent short-term performance, where both models accurately tracked the measured data. The LSTM model, in

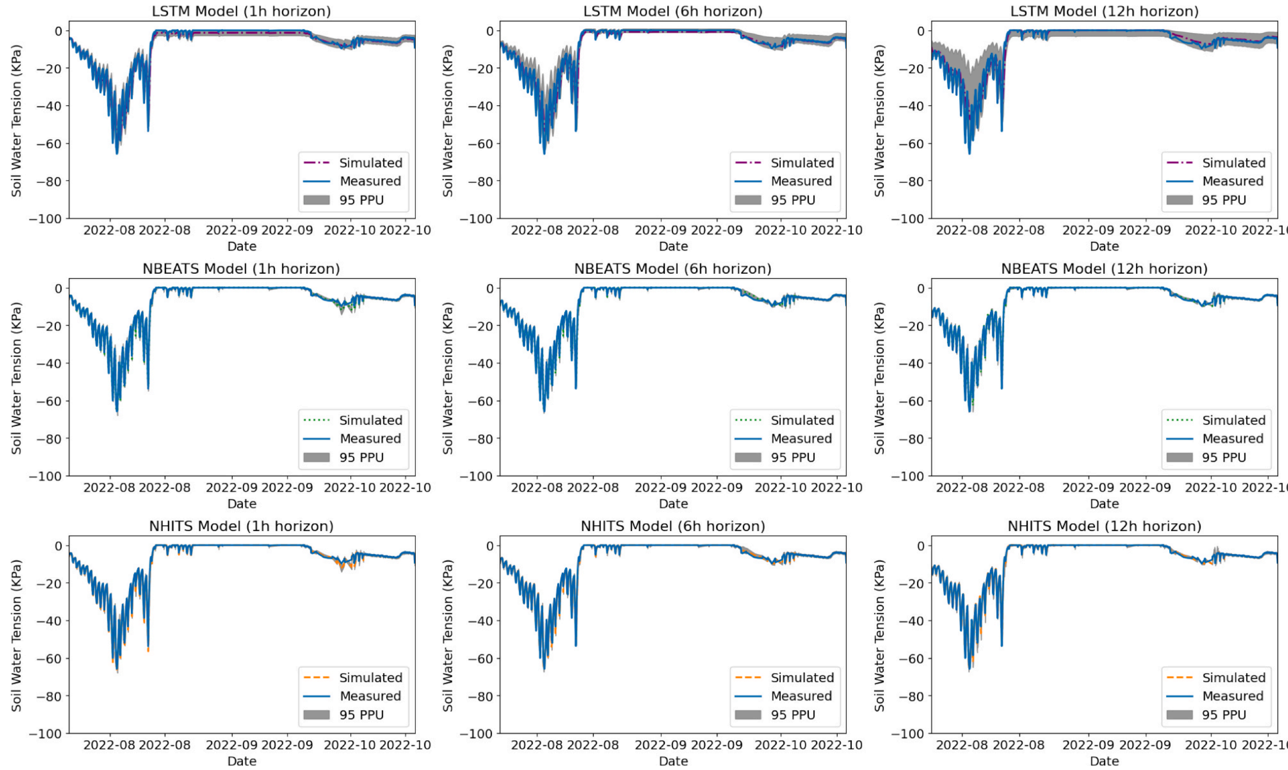


Fig. 21. Soil water tension simulations across different time horizons at 0.3 m soil depth. The grey shaded area indicates 95PPU. As shown, the 95PPU associated with LSTM widened as the prediction horizon increased.

contrast, showed comparatively lower accuracy with an NSE of 95.3 % and higher error metrics (RMSE of 2.81 and MAE of 1.83), suggesting a slightly lower precision of the benchmark model in capturing short-term fluctuations at this soil depth. At the 6 h horizon, N-HiTS and N-BEATS maintained strong performance with NSE values slightly lower than the 1 h horizon (around 99.6 %) and minor increases in RMSE and MAE, indicating robustness over extended horizons. LSTM's performance further declined, with a lower NSE of 93.24 % and higher RMSE and MAE, highlighting relatively reduced predictive reliability over time.

At the 12 h horizon, N-HiTS and N-BEATS continued to outperform LSTM, although with a slight increase in RMSE and MAE, reflecting a decreased precision as the prediction horizon progressed. At this horizon, N-HiTS achieved the highest accuracy with NSE of 99.64 %. Meanwhile, the benchmark model exhibited the highest RMSE, MAE, and the lowest NSE (91.56 %) values. Overall, the N-HiTS and N-BEATS models demonstrated improved predictive accuracy and stability across all horizons, particularly in short-term horizons. At the same time, LSTM's performance diminished rapidly across the 6 h and 12 h horizons.

The 95PPU plots of N-BEATS and N-HiTS models indicated that the accuracy and reliability of predictions were inclined to vary significantly at 0.6 m soil depth across 1 h, 6 h, and 12 h horizons. This is underscored by both the closeness of the predictions to measured values and the extent of the uncertainty bands shown in Fig. 21. The 1 h horizon generally showed narrower uncertainty bands for all models, suggesting higher reliability and confidence in short-term predictions. This aligns with high NSE values (around 99 %) and low RMSE and MAE, particularly for N-BEATS and N-HiTS models.

As the prediction horizon extended to 6 h and 12 h, the 95PPU widened slightly, reflecting an expected reduction in predictive accuracy and confidence of the models over time. This trend was most noticeable in the LSTM model, particularly at the 12 h horizon, and was marked by the increased confidence interval width. Despite the decline in predictive power, both N-HiTS and N-BEATS models maintained relatively high accuracy and reliability even at the 12 h horizon, with narrower bands and lower RMSE and MAE values than the benchmark model. This performance stability over longer horizons made N-HiTS and N-BEATS favorable for soil water tension prediction, as accuracy over an increasing prediction time window is critical to avoid over- or under-irrigation.

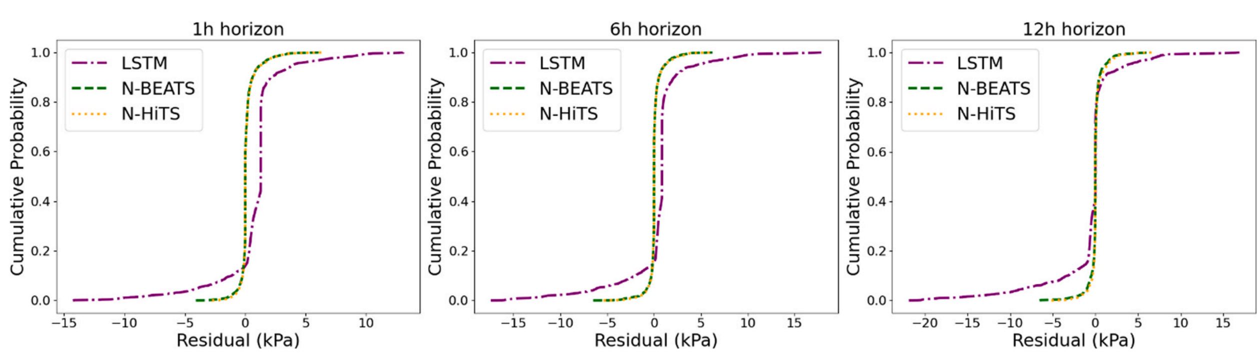


Fig. 22. The CDF curves of residuals across different time horizons at 0.6 m soil depth.

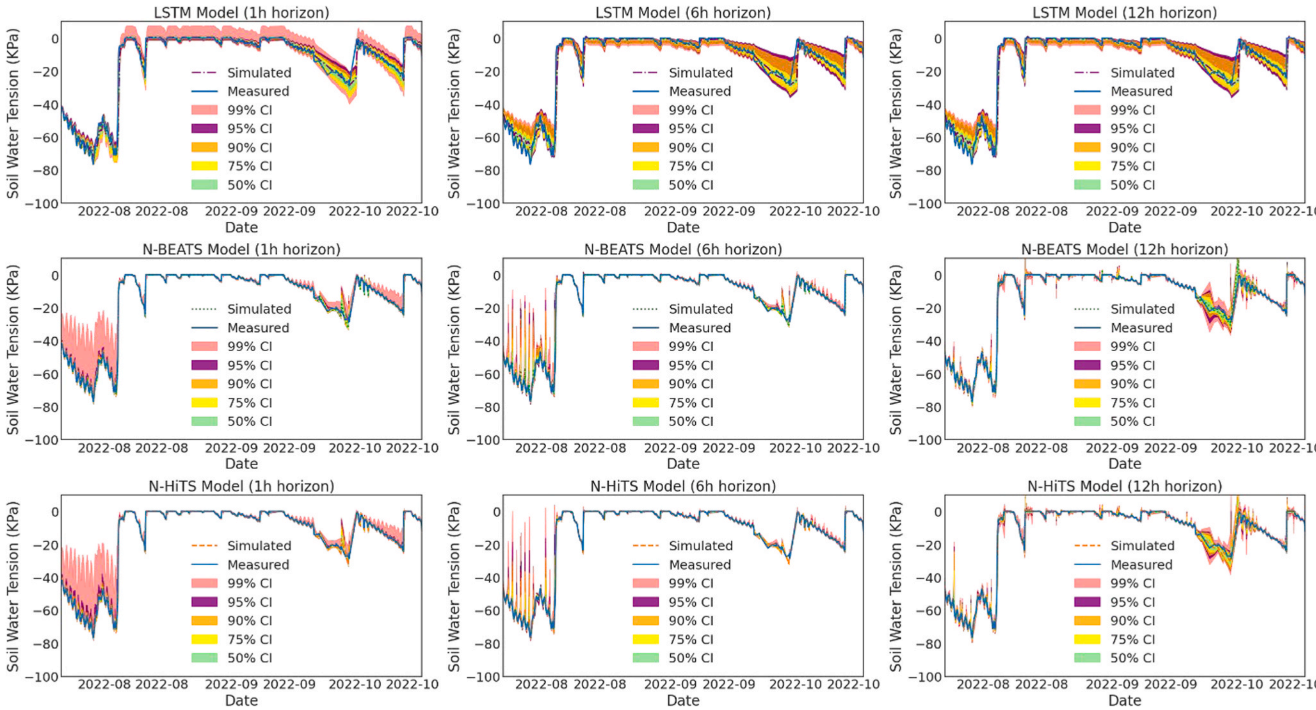


Fig. 23. The 50 %, 75 %, 90 %, 95 % and 99 % uncertainty bands across different time horizons at 0.15 m soil depth.

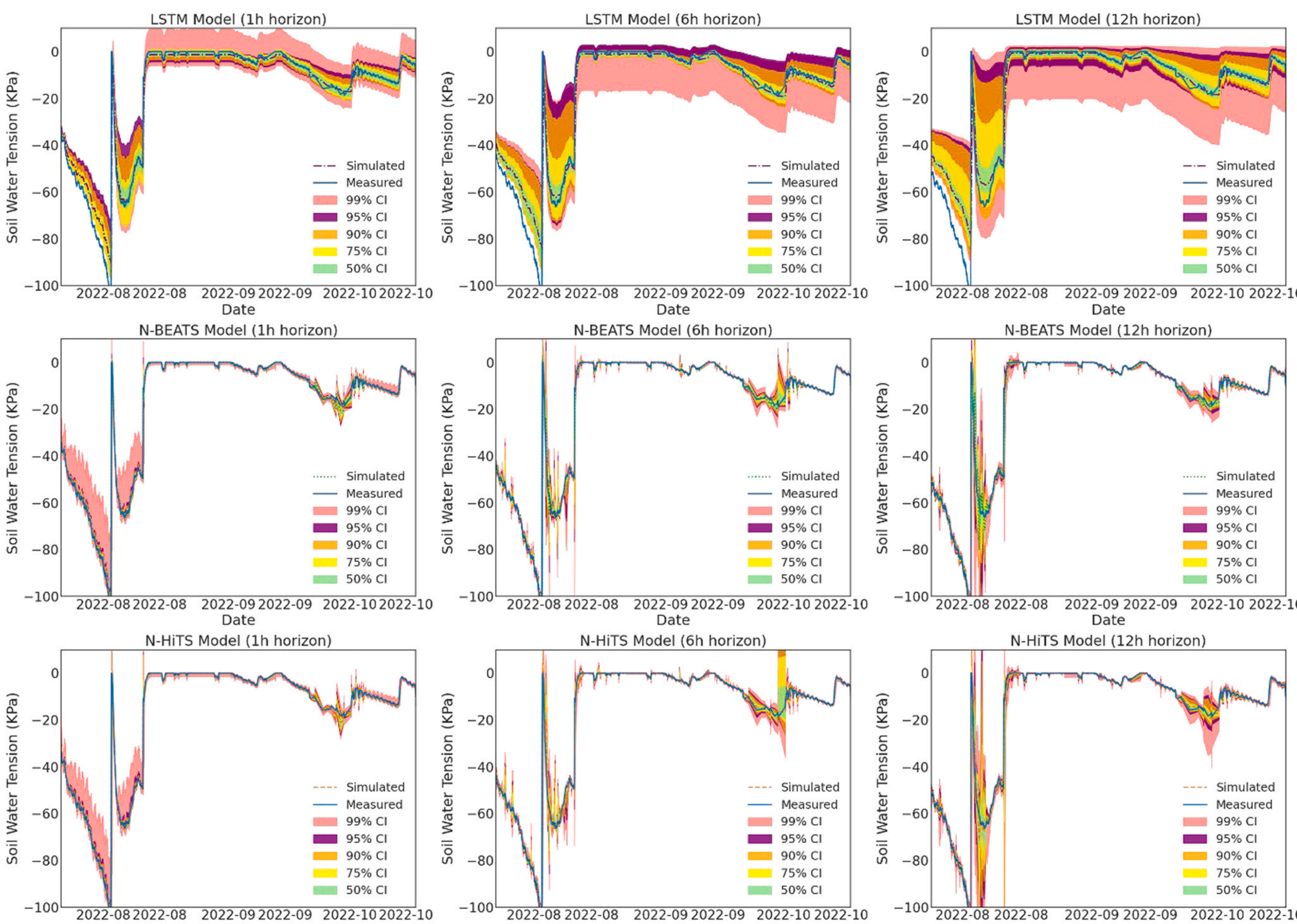


Fig. 24. The 50 %, 75 %, 90 %, 95 %, and 99 % uncertainty bands across different time horizons at 0.3 m soil depth. The uncertainty bands of N-HiTS and N-BEATS models showed more fluctuations for the 12 h horizon while LSTM uncertainty bands showed poor convergence.

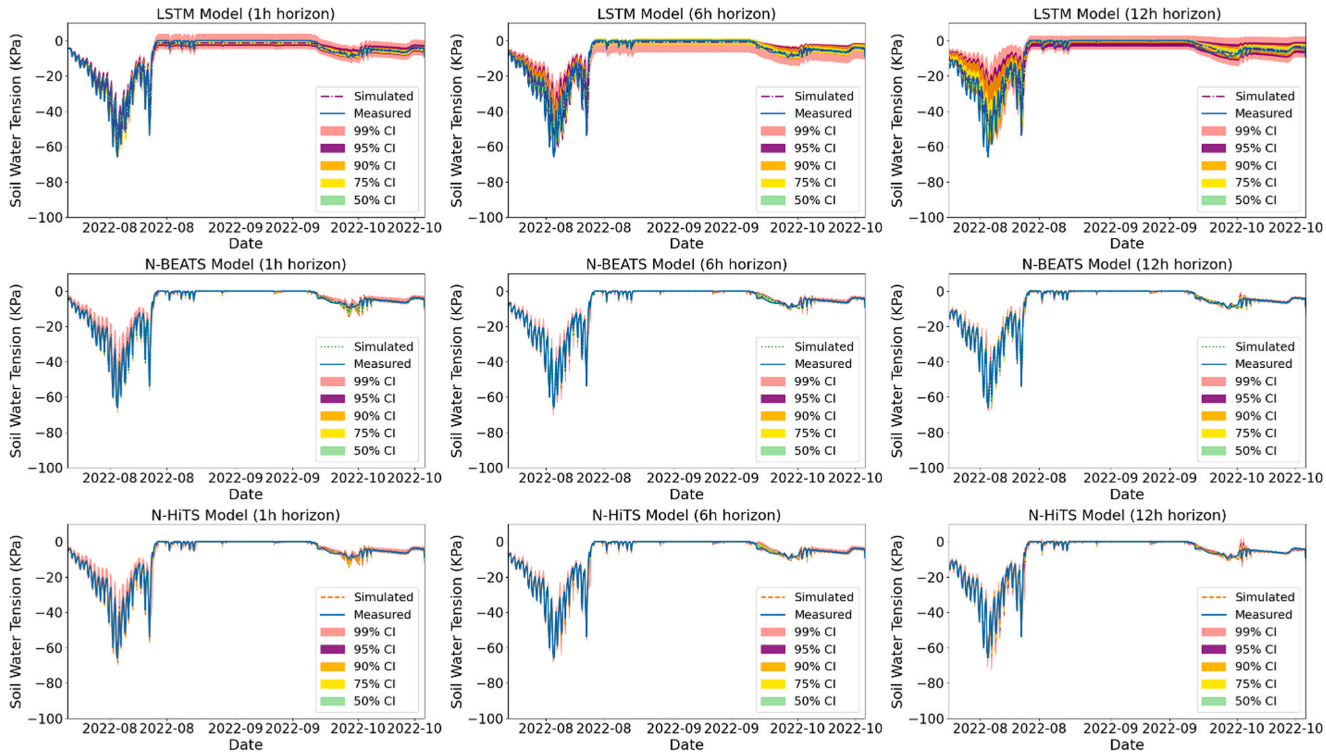


Fig. 25. The 50 %, 75 %, 90 %, 95 % and 99 % uncertainty bands across different time horizons at 0.6 m soil depth. The uncertainty bands are more consistent on this horizon.

The CDF plots of residuals for soil water tension prediction at 0.6 m show how the errors accumulated as the prediction horizon increased. As shown in Fig. 22, the CDF curves indicate a steeper ascent near zero for both N-HiTS and N-BEATS models, implying that these models tended to have a high probability of producing minor errors. The error of the two models was also tight, spanning from about −5 to 5 kPa across all horizons. This indicates better accuracy, confidence in predictions and the robustness achieved by the two models over extended prediction horizons, as their increase had a minor effect on the performance. Additionally, the CDF curves of N-HiTS and N-BEATS models were closely aligned, indicating similar performance across all horizons. The benchmark model, on the other hand, showed a broader spread of residual spanning from about −15 to 14 kPa for 1 h to −21 to 17 kPa for the 12 h horizons. The CDF curve also gradually reached one cumulative probability, implying that LSTM produced more variable and less precise predictions than N-HiTS and N-BEATS.

In addition to the 95 PPU uncertainty analysis, the performance of N-HiTS and N-BEATS at 50, 75, 90 and 99 % were assessed across depths and horizons (see Fig. 23, Fig. 24 and Fig. 25). P-factor and R-factor values showed that both models provided compelling predictions. Although LSTM provided comparative results, its probabilistic predictions across depth and horizons were less accurate.

Table 5 indicated that N-HiTS and N-BEATS outperformed LSTM regarding prediction accuracy, exhibiting narrow uncertainty bands. However, as we observed in GA's results, the two models showed lower convergence than LSTM.

4. Limitations of this research

The study emphasizes the use of two advanced neural forecasting models, N-HiTS and N-BEATS, and compares them with the traditional LSTM model based on the learned trends and seasonality embedded in the observed data. We acknowledge that this research, while thorough, is not without limitations. Here we highlight potential areas for improvement in future studies.

First, noisy sensor data can introduce erratic fluctuations, resulting in inaccurate predictions and poor model performance. Similarly, significant data gaps can limit the models' capabilities to capture relevant temporal features, consequently affecting modeling predictive potential. Second, if models are not well-calibrated, uncertainty results may lead to overestimation or underestimation of confidence intervals in the probabilistic fashion. Ensuring proper model calibration is, therefore, crucial to obtain more accurate and reliable predictions. Third, while weather and crop-specific variables regulate soil–water dynamics, this study focused on exploring the univariate capabilities of N-HiTS and N-BEATS. However, evaluating their sensitivity to different additional input variables such as precipitation and evapotranspiration could provide insights into the most influential features and support model interpretability.

Fourth, soil water tension varies across climatic regions and soil and crop types. In this study, the developed algorithms were trained and tested on data collected during cotton growing seasons under relatively homogeneous climatic conditions. Their generalization capabilities across distinct climates and cropping systems are yet to be determined. Therefore, their deployment in new agricultural environments would necessitate model retraining to adapt to site-specific climatic and agronomic conditions. Fifth, the deep learning architecture of N-HiTS and N-BEATS fundamentally demands significant computational resources. The training and testing typically require access to powerful CPUs and GPUs to achieve reasonable processing times. Therefore, N-HiTS and N-BEATS deployment in real time agricultural settings with computationally constrained systems may hinder timely predictions and irrigation decisions.

5. Conclusion and future works

This study employed two modern DNN models, N-HiTS and N-BEATS, to predict soil water tension across multiple depths at 1 h, 6 h and 12 h horizons and benchmarked them with LSTM. Additionally, we assessed the uncertainties associated with soil water tension predictions

Table 5
Uncertainty estimation metrics across different depths and horizons for different confidence intervals. The best performances are shown in bold.

Model	Soil depth	Horizon	P-factor	R-factor	P-factor	R-factor	P-factor	R-factor	P-factor	R-factor
Confidence interval			**50**		**75**		**90**		**99**	
N-HiTS	0.15 m	1 h	**40.01**	**0.01**	60.10	0.03	71.14	0.05	**97.02**	0.32
N-BEATS			36.15	0.015	**75.11**	**0.025**	**88.52**	**0.04**	96.85	**0.30**
LSTM			23.12	0.07	40.12	0.12	84.60	0.16	92.60	0.45
N-HiTS		6 h	44.73	**0.01**	57.79	**0.03**	79.26	**0.05**	92.43	**0.13**
N-BEATS			**52.13**	0.02	**64.91**	0.04	**82.96**	0.07	**93.78**	0.15
LSTM			15.02	0.04	36.94	0.12	57.79	0.27	88.85	0.38
N-HiTS		12 h	49.61	**0.02**	**73.88**	**0.04**	85.20	**0.06**	**96.47**	0.14
N-BEATS			**50.45**	**0.02**	68.72	**0.04**	**85.71**	**0.06**	94.73	**0.12**
LSTM			15.02	0.04	36.94	0.12	57.79	0.27	88.85	0.38
N-HiTS	0.3 m	1 h	55.52	**0.01**	55.19	**0.018**	**89.07**	**0.03**	**99.06**	**0.24**
N-BEATS			60.32	**0.01**	75.39	0.02	88.91	**0.03**	98.68	0.25
LSTM			**76.55**	0.096	**80.74**	0.19	81.24	0.28	82.67	0.6
N-HiTS		6 h	49.94	**0.02**	**67.66**	**0.03**	**81.17**	**0.06**	**96.19**	0.14
N-BEATS			45.12	**0.02**	63.73	0.04	65.30	**0.06**	94.90	**0.13**
LSTM			**52.30**	0.12	**72.87**	0.22	55.16	0.37	84.42	0.9
N-HiTS		12 h	43.05	**0.03**	63.09	0.06	78.42	0.10	**94.89**	0.28
N-BEATS			34.50	**0.03**	55.86	**0.05**	77.44	0.08	93.28	**0.22**
LSTM			**73.42**	0.14	**80.37**	0.33	**85.99**	0.54	86.62	1.18
N-HiTS	0.6 m	1 h	57.78	**0.02**	74.67	**0.038**	84.49	**0.05**	**97.30**	**0.26**
N-BEATS			**59.38**	0.024	**75.39**	0.04	**86.70**	0.06	97.13	0.27
LSTM			13.36	0.05	28.15	0.17	32.84	0.2	91.00	0.65
N-HiTS		6 h	28.20	**0.019**	42.26	**0.033**	79.37	**0.05**	**92.49**	0.19
N-BEATS			30.94	**0.018**	45.24	**0.033**	79.88	**0.05**	91.48	**0.16**
LSTM			**69.56**	0.16	**78.59**	0.28	**87.50**	0.37	89.41	0.76
N-HiTS		12 h	28.20	**0.019**	42.26	**0.03**	79.37	**0.05**	**92.49**	0.19
N-BEATS			30.94	0.018	45.24	**0.03**	79.88	**0.05**	91.48	**0.16**
LSTM			**69.56**	0.16	**78.59**	0.28	**87.50**	0.37	89.41	0.76

at multiple uncertainty levels. N-HiTS and N-BEATS models provided comparable results and emerged as more accurate than the benchmark model, particularly for long-term predictions. Our findings demonstrated that N-HiTS and N-BEATS' outstanding performances were driven by their state-of-the-art modular structures and advanced modeling mechanisms. N-HiTS leveraged multi-rate data sampling and hierarchical interpolation approaches to account for the seasonality, trends and variability in data, while N-BEATS used distinct blocks that explicitly modeled the trend and seasonality in the data. This decomposition of input data into trends and seasons and the intermediate predictions at the stack level provided insights into the models' prediction process. Additionally, both models employed their double residual mechanism to significantly minimize the prediction errors, especially for long-term probabilistic predictions. Overall, the results indicated that N-BEATS performed better in short-term prediction. This can be further attributed in part to N-BEATS' ability to carry out a form of *meta*-optimization, which shows efficient fine-tuning of hyperparameters over prediction horizons.

Accurate soil water tension predictions are crucial in optimizing irrigation water use, particularly in water-scarce regions. By providing accurate predictions, N-HiTS and N-BEATS can assist growers and irrigation specialists in making more informed irrigation decisions. The results can also help with irrigation planning, leading to more efficient and sustainable farming practices. In addition, this study lays the groundwork for data-driven irrigation decision making. However, additional studies are needed to examine the individual algorithms in different soil water dynamics (e.g., dry vs wet conditions), climatic regions, and cropping conditions. Furthermore, including additional input features that affect soil water tension, such as weather and crop-specific variables, can provide a better understanding of the sensitivity of the models to different input variables. This will improve our perception of the key neural network shortcomings and limitations.

To summarize, the presented results are promising for both test beds across multiple horizons. As we continue to make progress in DNN applications, we expect to advance our understanding of modeling behaviors across different spatio-temporal scales, enhance our ability to improve DNN efficiency for short-term to long-term soil water tension predictions, and leverage advanced approaches such as Bayesian frameworks (e.g., Samadi et al., 2020; Tabas and Samadi, 2022) to understand how uncertainty propagates across space and time and modeling structures. The research presented herein is intended to provide a basis for neural applications in soil water tension prediction across growing seasons. However, additional studies are needed to examine the individual algorithms in different soil water tension conditions (e.g., dry vs wet) and check the behavior of the algorithms and the uncertainty propagation over time to understand key DNN shortcomings and limitations. The present work tends to contribute substantially to the field of neural irrigation hydrology, particularly to accelerate the design and optimization of new DNN algorithms, offering a substantial improvement in irrigation demand calculation. Future work could expand on this research by assessing how N-HiTS and N-BEATS models would perform against physical soil moisture models such as HYDRUS or SWAP. Furthermore, one could develop a hybrid physics-informed model that leverages the strengths of physically based models with DNNs that can help understand how water flow in the soil, governed by the soil hydraulic functions, can be computed in a data-driven fashion.

Acknowledging a growing enthusiasm for DNN applications in precision agriculture, we expect progress on multiple fronts: (i) a better uncertainty algorithm to quantify both data and modeling uncertainties in soil water tension prediction, (ii) a more sophisticated DNN model such as those that leverage self-attention mechanism to identify relevant patterns across different periods in the time series, and (iii) a better benchmarking model including traditional soil moisture models to enhance robustness and promote transferability of neural network results to new climatic and soil conditions. The present work tends to contribute substantially to the field of neural irrigation hydrology, particularly to accelerate the design and optimization of new DNN algorithms, offering a substantial improvement in irrigation demand calculation.

CRediT authorship contribution statement

Lisa Umutoni: Writing – original draft, Visualization, Validation, Formal analysis, Data curation. **Vidya Samadi:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **George Vellidis:** Writing – review & editing, Supervision, Methodology, Data curation, Conceptualization. **Charles Privette III:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Jose Payero:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Bulent Koc:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors appreciate the funding support from the Southern Sustainable Agriculture Research and Education (SARE; grant # 2022001722) graduate fellowship program as well as USDA-NIFA (grant # 2023000603). Any opinions, findings, and discussions expressed in this study are those of the authors and do not necessarily reflect the views of the SARE and USDA-NIFA. Clemson University is acknowledged for generously allocating computing time on the Palmetto cluster.

Code and data availability

The data and codes are available upon request after publication.

Data availability

Data will be made available on request.

References

Alexandratos, N., & Bruinsma, J., 2012. World agriculture towards 2015/2030: the 2012 revision. ESA Working paper (Vol. 12 No. 3). In *ESA Working paper: Vol. 12 No.3*.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.
Buckingham, E., 1907. Studies on the movement of soil moisture. Bureau of Soils--Bulletin No. 38.
Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L., Xue, X., 2019. Research on soil moisture prediction model based on deep learning. PLoS One 14 (4). https://doi.org/10.1371/journal.pone.0214508.
Challu, C., Olivares, K.G., Oreshkin, B.N., Ramirez, F.G., Mergenthaler-Canseco, M., Dubrawski, A., 2023. NHITS: Neural Hierarchical interpolation for time series forecasting. In: *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023, 37*. https://doi.org/10.1609/aaai.v37i6.25854.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3). https://doi.org/10.1023/A:1022627411411.
de Fraiture, C., Wichelns, D., 2010. Satisfying future water demands for agriculture. Agric. Water Manag. 97 (4). https://doi.org/10.1016/j.agwat.2009.08.008.
Feki, M., Ravazzani, G., Ceppi, A., Milleo, G., Mancini, M., 2018. Impact of infiltration process modeling on soil water content simulations for irrigation management. Water (Switzerland) 10 (7). https://doi.org/10.3390/w10070850.
Gardner, W., 1920. The capillary potential and its relation to soil-moisture constants. Soil Sci. 10 (5). https://doi.org/10.1097/00010694-192011000-00003.
Gu, Z., Qi, Z., Burghate, R., Yuan, S., Jiao, X., Xu, J., 2020. Irrigation scheduling approaches and applications: a review. J. Irrig. Drain. Eng. 146 (6). https://doi.org/10.1061/(asce)ir.1943-4774.0001464.
Gumiere, S.J., Camporese, M., Botto, A., Lafond, J.A., Paniconi, C., Gallichand, J., Rousseau, A.N., 2020. Machine learning vs. physics-based modeling for real-time irrigation management. Front. Water 2. https://doi.org/10.3389/frwa.2020.00008.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9 (8), 1735–1780.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4). https://doi.org/10.1016/j.ijforecast.2006.03.001.
Jamil, B., Akhtar, N., 2017. Comparative analysis of diffuse solar radiation models based on sky-clearness index and sunshine period for humid-subtropical climatic region of India: a case study. Renew. Sustain. Energy Rev. 78. https://doi.org/10.1016/j.rser.2017.04.073.
Jensen, M.E., Allen, R.G., 2016. Evaporation, evapotranspiration, and irrigation water requirements. In Evaporation, Evapotranspiration, and Irrigation Water Requirements. https://doi.org/10.1061/9780784414057.
Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. In *Computers and Electronics in Agriculture* (Vol. 147). https://doi.org/10.1016/j.compag.2018.02.016.
LeCun, Y., Hinton, G., & Bengio, Y. (2015). Deep learning (2015), Y. LeCun, Y. Bengio and G. Hinton. *Nature, 521*.
Lee, C. suk, Sohn, E., Park, J.D., Jang, J.D., 2019. Estimation of soil moisture using deep learning based on satellite data: a case study of South Korea. Giscience and Remote Sensing 56 (1). https://doi.org/10.1080/15481603.2018.1489943.
Li, P., Zha, Y., Shi, L., Tso, C.H.M., Zhang, Y., Zeng, W., 2020. Comparison of the use of a physical-based model with data assimilation and machine learning methods for simulating soil water dynamics. J. Hydrol. 584. https://doi.org/10.1016/j.jhydrol.2020.124692.
Li, Q., Li, Z., Shangguan, W., Wang, X., Li, L., Yu, F., 2022. Improving soil moisture prediction using a novel encoder-decoder model with residual learning. Comput. Electron. Agric. 195. https://doi.org/10.1016/j.compag.2022.106816.
Liu, J., Rahmani, F., Lawson, K., Shen, C., 2022. A multiscale deep learning model for soil moisture integrating satellite and in situ data. Geophys. Res. Lett. 49 (7). https://doi.org/10.1029/2021GL096847.
Ma, L., Ahuja, L.R., Ascough, J.C., Shaffer, M.J., Rojas, K.W., Malone, R.W., Cameira, M. R., 2001. Integrating system modeling with field research in agriculture: applications of the root zone water quality model (RZWQM). Adv. Agron. Vol. 71. https://doi.org/10.1016/s0065-2113(01)71016-4.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - a discussion of principles. J. Hydrol. 10 (3). https://doi.org/10.1016/0022-1694(70)90255-6.
Olivares, K.G., Challú, C., Garza, A., Canseco, M. M., & Dubrawski, A. (2022). *NeuralForecast: User friendly state-of-the-art neural forecasting models.*
Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y., 2020. N-beats: neural basis expansion analysis for interpretable time series forecasting. 8th International Conference on Learning Representations, ICLR 2020.

Pereira, L.S., 2017. Water, agriculture and food: challenges and issues. Water Resour. Manag. 31 (10). https://doi.org/10.1007/s11269-017-1664-z.

Roberts, T.M., Colwell, I., Chew, C., Lowe, S., Shah, R., 2022. A deep-learning approach to soil moisture estimation with GNSS-R. Remote Sens. (Basel) 14 (14). https://doi.org/10.3390/rs14143299.

Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. 65 (6). https://doi.org/10.1037/h0042519.

Saberian, M., Samadi, V., Popescu, I., 2024. Probabilistic Hierarchical Interpolation and Interpretable Configuration for Flood Prediction. https://doi.org/10.5194/hess-2024-261.

Samadi, S., Pourreza-Bilondi, M., Wilson, C.A.M.E., Hitchcock, D.B., 2020. Bayesian model averaging with fixed and flexible priors: Theory, concepts, and calibration experiments for rainfall-runoff modeling. Journal of Advances in Modeling Earth Systems 12 (7), e2019MS001924.

Sauer, T., Havlík, P., Schneider, U.A., Schmid, E., Kindermann, G., Obersteiner, M., 2010. Agriculture and resource availability in a changing world: the role of irrigation. Water Resour. Res. 46 (6). https://doi.org/10.1029/2009WR007729.

Schlesinger, W.H., Jasechko, S., 2014. Transpiration in the global water cycle. Agric. For. Meteorol. 189–190. https://doi.org/10.1016/j.agrformet.2014.01.011.

Tabas, S.S., Samadi, S., 2022. Variational bayesian dropout with a Gaussian prior for recurrent neural networks application in rainfall-runoff modeling. Environ. Res. Lett. 17 (6).

Šimůnek, J., Huang, K., Šejna, M., & van Genuchten, M. Th., 1998. The Hydrus-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably saturated media. Version 1.0, IGWMC - TPS - 70. In *International Ground Water Modeling Center, Colorado School of Mines, Golden, Colorado* (Issue July).

van Dam, J. C., Huygen, J., Wesseling, J. G., Feddes, R. A., Kabat, P., van Walsum, P. E. V., Groenendijk, P., & van Diepen, C. A., 1997. Theory of SWAP version 2.0. In *Wageningen UR*.

Umutoni, L., Samadi, V., 2024. Application of machine learning approaches in supporting irrigation decision making: A review. Agricultural Water Management 294, 108710.

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.

Wichelns, D., Oster, J.D., 2006. Sustainable irrigation is necessary and achievable, but direct costs and environmental impacts can be substantial. Agric. Water Manage. 86 (1–2). https://doi.org/10.1016/j.agwat.2006.07.014.

Yu, J., Zhang, X., Xu, L., Dong, J., Zhangzhong, L., 2021. A hybrid CNN-GRU model for predicting soil moisture in maize root zone. Agricultural Water Management 245, 106649.

Zafarmomen, N., Alizadeh, H., Bayat, M., Ehtiat, M., Moradkhani, H., 2024. Assimilation of sentinel-based leaf area index for modeling surface-ground water interactions in irrigation districts. Water Resour. Res. 60 (10), e2023WR036080. https://doi.org/10.1029/2023WR036080.

Zhang, D., Zhang, W., Huang, W., Hong, Z., Meng, L., 2017. Upscaling of surface soil moisture using a deep learning model with VIIRS RDR. ISPRS Int. J. Geo Inf. 6 (5). https://doi.org/10.3390/ijgi6050130.

Zhao, L., Luo, T., Jiang, X., Zhang, B., 2023. Prediction of soil moisture using BiGRU-LSTM model with STL decomposition in Qinghai–Tibet Plateau. PeerJ 11. https://doi.org/10.7717/peerj.15851.