

A hybrid ConvLSTM-Nudging model for predicting surface soil moisture in the Qilian Mountains, China (Postprint)

Authors: FAN Manhong, Xiao Qian, YU Qinghe, ZHAO Junhao, FAN Manhong

Date: 2025-11-17T00:00:00+00:00

Abstract

Spatiotemporal forecasting of surface soil moisture (SSM) is recognized as a critical scientific issue in precision agricultural irrigation, regional drought monitoring, and early warning systems for extreme precipitation. However, long-term forecasting continues to pose formidable challenges because of the complexity observed across both the spatial and temporal scales. In this study, we used a daily SSM dataset at a $0.05^\circ \times 0.05^\circ$ spatial resolution over the Qilian Mountains, China and proposed a hybrid Convolutional Long Short-Term Memory (ConvLSTM)-Nudging model, which combined deep neural networks with data assimilation to increase the accuracy of long-term SSM forecasting. We trained and evaluated the SSM predictive performance of four models (Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), ConvLSTM, and ConvLSTM with Squeeze-and-Excitation (SE) attention mechanism (ConvLSTM-SE)) in both short-term and long-term scenarios. The results showed that all the models perform well under short-term predictions, but the accuracy decrease substantially in long-term predictions. Therefore, we integrated Nudging technique during the long-term prediction phase to assimilate observational information and rectify model biases. Comprehensive evaluations demonstrate that Nudging significantly improves all the models, with ConvLSTM-Nudging achieving the best performance under the 200-d forecasting scenario. Relative to those of the best-performing ConvLSTM model for long-term forecasts, when observation noise $\delta=0.00$ and observation fraction $\text{obs}=50.0\%$, the coefficient of determination (R²) of ConvLSTM-Nudging increases by approximately 82.1%, while its mean absolute error (MAE) and root mean squared error (RMSE) decrease by approximately 84.8% and 77.3%, respectively; the average Pearson correlation coefficient (r) improves by approximately 23.6%, and Bias is reduced by 98.1%. These results demonstrated that

although pure deep learning models achieve high accuracy in the short-term predictions, they are prone to error accumulation and systematic drift in long-term autoregressive predictions. Integrating data assimilation with deep learning and continuously correcting the state through observation can effectively suppress long-term biases, thereby achieving robust long-term SSM forecasting.

Full Text

Preamble

Journal of Arid Land (2025) 17(11): 1623–1648

doi: 10.1007/s40333-025-0112-9; CSTR: 32276.14.JAL.02501129

Science Press & Springer-Verlag

A Hybrid ConvLSTM-Nudging Model for Predicting Surface Soil Moisture in the Qilian Mountains, China

FAN Manhong*, XIAO Qian, YU Qinghe, ZHAO Junhao

College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China

Abstract

Spatiotemporal forecasting of surface soil moisture (SSM) is recognized as a critical scientific issue in precision agricultural irrigation, regional drought monitoring, and early warning systems for extreme precipitation. However, long-term forecasting continues to pose formidable challenges because of the complexity observed across both spatial and temporal scales. In this study, we used a daily SSM dataset at a $0.05^\circ \times 0.05^\circ$ spatial resolution over the Qilian Mountains, China and proposed a hybrid Convolutional Long Short-Term Memory (ConvLSTM)-Nudging model, which combined deep neural networks with data assimilation to increase the accuracy of long-term SSM forecasting. We trained and evaluated the SSM predictive performance of four models (Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), ConvLSTM, and ConvLSTM with Squeeze-and-Excitation (SE) attention mechanism (ConvLSTM-SE)) in both short-term and long-term scenarios. The results showed that all models perform well under short-term predictions, but accuracy decreases substantially in long-term predictions. Therefore, we integrated the Nudging technique during the long-term prediction phase to assimilate observational information and rectify model biases. Comprehensive evaluations demonstrate that Nudging significantly improves all models, with ConvLSTM-Nudging achieving the best performance under the 200-day forecasting scenario. Relative to the best-performing ConvLSTM model for long-term forecasts, when observation noise $\delta=0.00$ and observation fraction $\text{obs}=50.0\%$, the coefficient of determination (R^2) of ConvLSTM-Nudging increases by approximately 82.1%, while its mean absolute error (MAE) and root mean squared error (RMSE) decrease by approximately 84.8% and 77.3%, respectively; the average Pear-

son correlation coefficient (r) improves by approximately 23.6%, and Bias is reduced by 98.1%. These results demonstrate that although pure deep learning models achieve high accuracy in short-term predictions, they are prone to error accumulation and systematic drift in long-term autoregressive predictions. Integrating data assimilation with deep learning and continuously correcting the state through observation can effectively suppress long-term biases, thereby achieving robust long-term SSM forecasting.

Keywords: data assimilation; surface soil moisture; deep neural networks; Convolutional Long Short-Term Memory (ConvLSTM); Squeeze-and-Excitation (SE) attention mechanism; Nudging; long-term prediction

Citation: FAN Manhong, XIAO Qian, YU Qinghe, ZHAO Junhao. 2025. A hybrid ConvLSTM-Nudging model for predicting surface soil moisture in the Qilian Mountains, China. *Journal of Arid Land*, 17(11): 1623-1648. <https://doi.org/10.1007/s40333-025-0112-9>; <https://cstr.cn/32276.14.JAL.02501129>

1 Introduction

Accurately predicting surface soil moisture (SSM) is critical for hydrological modeling, agricultural management, and climate change research (Vereecken et al., 2008). Nevertheless, long-term SSM prediction remains highly challenging because of the complex interplay among atmospheric forcing, surface heterogeneity, and subsurface hydrological processes (Entin et al., 2000; Koster et al., 2004).

Traditional SSM prediction methods are usually based on physical models. For instance, the Richards (1931) equation, a fundamental physical model for soil moisture prediction, is widely used to simulate soil water movement under both saturated and unsaturated conditions. Soil Vegetation Atmosphere Transfer (SVAT) models, such as the Noah land surface model, are applied at regional scales to simulate soil hydrology, atmosphere-vegetation interactions, and climate change (Ek et al., 2003). The Soil Water Atmosphere Plant (SWAP) model comprehensively accounts for soil moisture dynamics, root water uptake by plants, and the evapotranspiration processes driven by climatic conditions (van Dam et al., 2008). The HYDRUS model operates based on the finite-element method to solve coupled problems involving water flow, heat transfer, and solute transport in unsaturated soils, demonstrating strong applicability and versatility (Šimůnek et al., 2008). Although these physical models offer good interpretability, they often perform poorly in practical applications because of parameter uncertainty and inadequate descriptions of complex physical processes (Li et al., 2022a).

In recent years, deep learning models have been widely applied to soil moisture prediction tasks, demonstrating stronger capabilities for learning nonlinear features as well as superior fitting and generalization performance (Ding et al.,

2024). Convolutional Neural Networks (CNNs), which were originally proposed by LeCun (1989) for image recognition tasks, are well suited for extracting spatial features from structured data. Since soil moisture and meteorological time series data can be structured as image-like grids, CNNs effectively capture their localized spatial features and apply them to soil moisture modeling. Recurrent Neural Networks (RNNs), on the other hand, excel at processing time series data and are capable of capturing sequential dependencies between variables, enabling the modeling of dynamic soil moisture changes (Connor et al., 1994). However, the standard RNN architecture is prone to gradient vanishing or gradient exploding problems when processing long sequences, which limits its effectiveness in long-distance dependency modeling cases (Mikolov et al., 2011). To overcome this limitation, Hochreiter and Schmidhuber (1997) proposed the Long Short-Term Memory (LSTM) neural network, which incorporates a gating mechanism to enhance its ability to model long-term dependencies.

Despite the remarkable success of LSTM in time series modeling tasks, its complex structure still results in high computational costs for both training and prediction (Song et al., 2020). Building upon this foundation, Yu et al. (2021) designed a hybrid model that integrates a CNN and a Gated Recurrent Unit (GRU). The CNN-GRU model consistently outperforms the standalone CNN and GRU models in terms of predicting soil moisture at various depths within the root zone. Nevertheless, the parallel computations of the convolutional and recurrent units in this structure significantly increase the number of model parameters and the computational cost, and the model is prone to overfitting when the sample size is insufficient. In addition, Li et al. (2022b) integrated an attention mechanism into the LSTM model, which further improves its performance across various prediction scenarios and significantly enhances its ability to identify key time steps. Nonetheless, the introduction of the attention mechanism also increases the number of model parameters, making the training process more sensitive to sample size and noise and introducing training instability and overfitting risks. Convolutional LSTM (ConvLSTM) network, which integrates convolutional structures with recurrent gating mechanisms, is capable of simultaneously capturing spatial and temporal dependencies (Gamboa-Villafuella et al., 2021). It has been successfully applied to spatiotemporal dynamic SSM modeling tasks (Shi et al., 2015; Li et al., 2024a; Lü et al., 2024). ConvLSTM with a Squeeze-and-Excitation (SE) attention mechanism (ConvLSTM-SE) model performs quite well in short-term predictions; however, its performance drops sharply as the prediction timescale increases (Li et al., 2024a). Despite the excellent performance of deep learning in soil moisture prediction scenarios, its purely data-driven nature continues to present certain challenges, especially when addressing sudden environmental events (e.g., rainfall) (Wang et al., 2024). The lack of physical constraints leads to the accumulation of errors and reduces the generalizability of the model, thereby limiting its stability and reliability in long-term prediction tasks (Cheng et al., 2023).

Data assimilation was originally designed to provide reliable initial conditions for numerical atmospheric forecasting (Charney et al., 1950) and has been in-

creasingly applied in soil hydrology, where it integrates observational data with physical models to improve spatiotemporal soil moisture estimates and characterize the associated uncertainties. For example, Xu et al. (2020) coupled satellite soil moisture with data assimilation and an integrated drought index at the continental scale, demonstrating that the combined use of these datasets can substantially enhance regional drought detection and monitoring capabilities, thereby underscoring the practical value of assimilation techniques for conducting drought assessments and providing early warnings. Margulis et al. (2002) proposed the integration of the ensemble Kalman filter (EnKF) with a land surface model to effectively balance the computational efficiency and spatiotemporal resolution achieved when capturing fine-scale soil moisture dynamics. Zhu et al. (2017) employed an EnKF-based differential information assimilation method to successfully extract effective information from noisy data. Gruber et al. (2019) designed an adaptive Kalman filter framework that corrected the deviation between the model outputs and the observational data through triple collocation analysis and Monte Carlo simulation processes. Considering the high cost of acquiring SSM data, Wang et al. (2020) proposed a robust data value analysis framework based on a hybrid data assimilation method. Wang et al. (2023) integrated EnKF with HYDRUS-1D model to improve the accuracy of soil moisture simulations and highlighted the critical influences of ensemble size and observation error on the performance of the assimilation system. Although data assimilation substantially improves the consistency between models and observations and enhances short- to mid-term state estimates, its reliance on accurate physical models, the computational burden imposed on high-dimensional nonlinear systems, and the requirement for high-quality observations limit its applicability for use in certain long-term forecasting and sparse-observation scenarios.

Hybrid strategies that integrate data assimilation with deep learning have recently emerged by combining physical constraints with the representational power of data-driven methods. The Nudging technique was first applied in numerical weather forecasting (Anthes, 1974); it dynamically adjusts the state of the utilized model to match observational data through relaxation terms. Pawar et al. (2020) successfully integrated Nudging into the data assimilation process for geophysical flows. Utilizing the Lorenz-96 system as a testbed, they integrated data assimilation with deep learning and compared the performance of Nudging with that of the extended Kalman filter (EKF), EnKF, and deterministic ensemble Kalman filter (DEnKF). The experiments revealed that Nudging has the potential to assimilate very sparse observational data while avoiding matrix operations such as computing the Kalman gain. Menut et al. (2024) conducted a systematic comparison between the configurations “with or without Nudging and with or without online coupling,” finding that adopting Nudging yielded a greater improvement in model skill than online coupling itself did (higher correlations and lower Bias and Root Mean Square Error (RMSE) values). They also reported that Nudging reduced the sensitivity of outputs to different physics configurations by 30.0%–70.0%. Antil et al. (2024) explicitly embedded Nudging into deep network-based data assimilation pipelines, pro-

viding empirical demonstrations of stability and convergence and extending the applicability of Nudging under sparse observation settings. Based on the above theoretical and empirical progress, we propose a ConvLSTM-Nudging framework that combines the spatiotemporal modeling capabilities of ConvLSTM with physics-based state correction, with the aim of improving the accuracy of long-term SSM predictions. The main objectives of this study are to: (1) evaluate the performance of CNN, LSTM, ConvLSTM, and ConvLSTM-SE in SSM modeling tasks and in both short- and long-term SSM prediction scenarios; (2) identify the optimal parameters for the Nudging technique under varying observation fractions and observation errors; and (3) investigate the effectiveness of Nudging for improving long-term prediction accuracy.

2.1 Study Area

The Qilian Mountains (35°49' -39°58' N, 93°33' -103°54' E) are located between western Gansu Province and northeastern Qinghai Province in China, with an altitude range of 1040–5993 m (Fig. 1 [FIGURE:1]). The overall orientation is northwest-southeast, with terrain that is high in the west and low in the east. The area is bordered by the Altun Mountains to the west, the Qinling Mountains and Liupan Mountains to the east, the Qaidam Basin to the south, and the Hexi Corridor to the north. The Qilian Mountains have a semi-arid to arid temperate continental mountain climate, with cool and humid summers and cold and dry winters (Lin et al., 2017; Liu et al., 2024). The average annual sunshine duration is 1744.0 h, the average annual temperature is approximately 5°C, and the average annual precipitation is approximately 250.0 mm, which is concentrated mainly during summer. In recent years, evaporation has been declining annually, with an average evaporation level of 634.73 mm annually, and the rate of decrease is 4.39 mm/a (Zhou and Li, 2022). The spatial distribution of soil moisture content clearly decreases from east to west (Meng et al., 2021).

2.2 Data Sources

The daily SSM dataset with a spatial resolution of $0.05^\circ \times 0.05^\circ$ over the Qilian Mountains from 2017 to 2021 was obtained from [//data.tpdc.ac.cn](http://data.tpdc.ac.cn). The dataset was developed by Qu et al. (2019, 2021), Chai et al. (2020), and Hu et al. (2022) using an optimized downscaling model (RF-OWCM). It combines a multivariate statistical regression framework with the Advanced Very High Resolution Radiometer (AVHRR) brightness temperature (TB) based Soil Moisture Active Passive (SMAP) time-expanded daily $0.25^\circ \times 0.25^\circ$ land SSM dataset in the Qilian Mountains (SMsmapTE, V1; <https://doi.org/10.11888/Geogra.tpdc.270235>).

The 30 m resolution digital elevation model (DEM) data in the Qilian Mountains were provided by the National Cryosphere Desert Data Center (<http://www.ncdc.ac.cn>). It was generated by Zhang et al. (2020) using Shuttle Radar Topography Mission (SRTM) v.4.1 1s (approximately 30 m) segmented

data through format conversion, image stitching, reprojection, and regional clipping.

3.1 ConvLSTM

ConvLSTM extends the inputs and hidden states to three-dimensional tensors (time, height, and width) and replaces the dense connections in the traditional LSTM model with convolutional operations, thereby enabling the capture of local spatial correlations. The ConvLSTM cell processes the input spatiotemporal tensor $X_t \in \mathbb{R}^{C \times H \times W}$ at each time step t while maintaining the hidden state $H_{t-1} \in \mathbb{R}^{K \times H \times W}$ and the cell state $C_{t-1} \in \mathbb{R}^{K \times H \times W}$, where C , K , H , and W denote the number of input channels, number of hidden channels, spatial height, and width of the feature map, respectively. The computational operations are defined as follows (Gamboa-Villafruela et al., 2021; Li et al., 2024a):

Input gate:

$$I_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \quad (1)$$

Forget gate:

$$F_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \quad (2)$$

Candidate cell state:

$$G_t = \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (3)$$

Cell state update:

$$C_t = F_t \circ C_{t-1} + I_t \circ G_t \quad (4)$$

Output gate:

$$O_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \quad (5)$$

Hidden state:

$$H_t = O_t \circ \tanh(C_t) \quad (6)$$

where $*$ represents the convolution operation; \circ represents the Hadamard product; σ denotes the sigmoid function; I_t , F_t , G_t , and O_t denote the input, forget, candidate, and output gates, respectively; C_t is the cell state update; H_t is the hidden state; W_{xi} , W_{xf} , W_{xc} , and W_{xo} (input-to-gate kernels), W_{hi} , W_{hf} , W_{hc} , and W_{ho} (hidden-to-gate kernels), and W_{ci} , W_{cf} , and W_{co} (cell-to-gate weights)

denote the weights of the convolution kernels for the input, forgetting, candidate, and output gates, respectively; and b_i , b_f , and b_o are the corresponding bias terms for the input gate, forget gate, and output gate, respectively.

The ConvLSTM model was built on an encoder-decoder architecture (Fig. 2 [FIGURE:2]). The encoder consisted of two ConvLSTM layers, each with 8 hidden channels and a 3×3 convolution kernel. The input sequence length was set to 14 timesteps, and the same padding was applied to preserve the original spatial resolution. These layers were responsible for extracting the spatiotemporal features of the input sequence. The decoder contained a single ConvLSTM layer with the same number of hidden channels, kernel size, and padding as those of the encoder; this layer was used to generate SSM predictions for the next day. In addition, to prevent model overfitting, we added a dropout layer between the ConvLSTM layers and set the dropout rate to 0.2.

3.2 ConvLSTM-SE

On the basis of the ConvLSTM model, we introduced the SE module (Jin et al., 2022), a channel attention mechanism that enhances the key feature representations of the model by adaptively adjusting the weights of different channels in CNNs. Its core goal is to explicitly model the dependencies between channels and improve the performance of the model by filtering important feature channels and suppressing redundant information. The calculation process of the SE module can be divided into the following three main steps (Hu et al., 2018).

(1) Squeeze (F_{sq}): Global average pooling F_{sq} was performed on the input $X \in \mathbb{R}^{H \times W \times C}$ to compress the spatial features in each channel into a scalar and obtain the channel description vector $z \in \mathbb{R}^{1 \times 1 \times C}$. This step could explicitly model the global channel information and alleviate the dependencies between channels.

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (7)$$

where z_c is the channel description vector in the c th channel; u_c is the feature map in the c th channel; and i and j are the spatial position indices.

(2) Excitation (F_{ex}): The channel description vector z obtained in the previous stage was input into a two-layer fully connected network to generate an attention weight value $s \in \mathbb{R}^{1 \times 1 \times C}$ for each channel. The output dimensionality of the first fully connected layer was $1 \times 1 \times \frac{C}{R}$ (where R is the compression ratio hyperparameter), and the activation function was a Rectified Linear Unit (ReLU). The second layer was restored to $1 \times 1 \times C$, and the activation function was the sigmoid function.

$$s = \sigma(g(z, D)) = \sigma(D_2 \cdot \text{ReLU}(D_1 \cdot z)) \quad (8)$$

where g represents a two-layer fully connected gating function; D represents the weights matrix; θ denotes the ReLU activation function; and D_1 and D_2 denote the weights matrix learned in the two-layer fully connected network.

(3) Scale (F_{scale}): The normalized weight vector s obtained above was applied to the feature X of each channel. That is, the $H \times W$ values of each channel in X were multiplied by the weight of the corresponding channel in s to obtain the desired feature Y , which has the same size as that of feature X . The SE module did not change the size of the feature map.

$$Y_c = s_c \cdot X_c \quad (9)$$

where s_c is the attention weight value in the c th channel; and Y_c is the feature in the c th channel.

The architecture of the ConvLSTM-SE model based on the encoder-decoder framework is shown in Figure 3

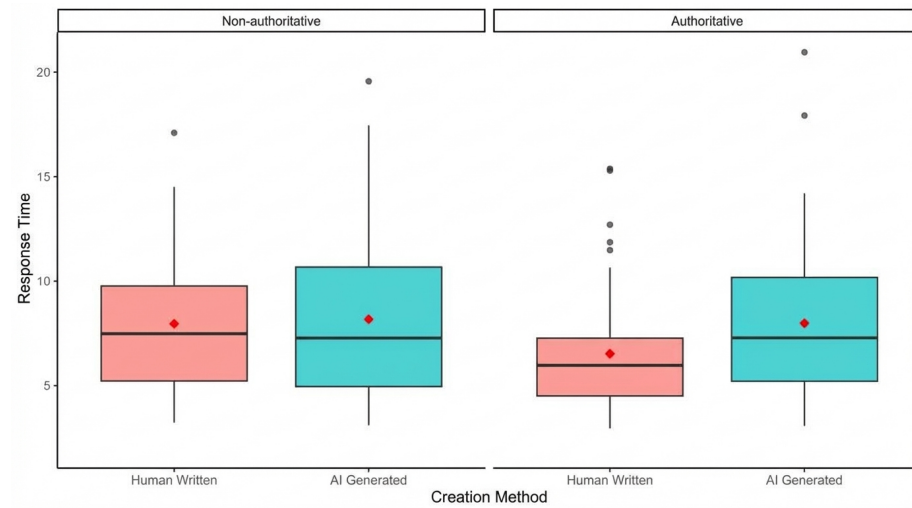


Figure 1: Figure 3

. The model took 14 consecutive days of spatial raster data as its inputs and used a two-layer ConvLSTM encoder to extract spatiotemporal features. In the decoder, the SE attention mechanism was integrated to extract global information through a fully connected layer and global average pooling, and the feature map was adaptively weighted to enhance the representations of salient patterns.

Afterward, the recalibrated features were passed through the decoder ConvLSTM layer to generate a soil moisture forecast for the next day, thus achieving accurate and efficient SSM prediction.

3.3 Nudging

Nudging, also known as Newtonian relaxation, was originally proposed by Anthes (1974) for initializing meteorological models. As an empirical data assimilation method, Nudging has been widely used in meteorological and hydrological models because it is easy to implement and its computational cost is significantly lower than those of variational and ensemble methods (Stauffer and Seaman, 1990). The core idea is to incorporate a feedback term that is proportional to the “observation-forecast” residual into the model equations so that the model state continuously approaches the observations while suppressing large-scale errors without degrading the ability of the model to generate realistic mesoscale structures (Conti et al., 2022).

$$\frac{dX_k}{dt} = M(X_k) + G \cdot (X_{\text{obs}} - X_k) \quad (10)$$

where X_k is the model state (prediction); M is the dynamic model operator; $G \in \mathbb{R}^{n \times m}$ is the Nudging gain matrix; and X_{obs} is the observation.

Based on Equation 10, we coupled Nudging with the model prediction process in discrete time using per-pixel correction. Specifically, the model took the most recent L days of the given SSM sequence as its input and produced a one-step prediction \hat{X}_t . If an observation Y_t is available on day t , with its availability indicated by a binary mask $M_t \in \{0, 1\}^{H \times W}$, we applied a pixel-level update to the model output:

$$X_t^a = \hat{X}_t + \tau \cdot M_t \circ (Y_t - \hat{X}_t) \quad (11)$$

where X_t^a represents the assimilated analysis value; $\tau \in (0, 1]$ is the Nudging coefficient; and no correction is applied when $M_t = 0$. This discretization scheme was equivalent to taking G in Equation 10 as a mask-gated diagonal gain (performing a convex combination only at the observed pixels), thereby balancing numerical stability and computational efficiency. Since the SSM observations in this study directly corresponded to the state variable, the observation operator was always an identity matrix.

We adopted output-layer coupling: the analysis field X_t^a was written back into the input buffer (replacing the newest frame), and together with the preceding $L - 1$ frames, formed the input sequence for the next step; “assimilate whenever observations were available” was the default frequency. To reflect sparse observation scenarios, we downsampled the data to 10.0%, 20.0%, and 50.0%

availability to construct $M_t = 0$, and we injected Gaussian observation noise δ , only at pixels with $M_t = 1$, to evaluate the robustness of the model under noise. Guided by the literature-recommended range of 0.1-1.0 (Pawar et al., 2020), we selected the optimal Nudging coefficient τ via a grid search.

The overall workflow and schematic of the proposed method from data processing and model recursion to Nudging correction are presented in Figure 4 [FIGURE:4]. The research process can be divided into three main steps: (1) **Data processing:** DEM-assisted clipping was used to obtain the SSM raster sequence for the study area, which was then split into training, validation, and test sets; (2) **Model prediction:** This section included the prediction processes of two models: ConvLSTM and ConvLSTM-SE. The 14-day observation data acquired from $t - 14$ to t were used as the initial inputs to generate a prediction for $t + 1$. Afterward, the prediction for the previous time step (or the corrected analysis field) was combined with the observation data derived from the previous 13 days to form a new 14-day sequence, which served as the input for the next step. This process was iterated until $t + 200$ days. A schematic diagram of the ConvLSTM cell and the SE module structure is shown, with the training and validation loss convergence curves displayed on the right; and (3) **Nudging correction:** Observation masks were constructed at fractions of 10.0%, 20.0%, and 50.0%, and Gaussian noise with $\delta = 0.01$ was added to simulate the effect of noise on the observations. The prediction field was corrected online, and the corrected X_t^a was updated based on Equation 11 and written back to the input buffer as the latest frame for the next recursive prediction. The example at the bottom shows the correction results produced at different observation fractions.

3.4 Evaluation Strategy and Metrics

In this study, a multi-metric evaluation framework is employed to systematically assess the performance of the developed model. During the model optimization phase, the training and validation sets are optimized with the mean squared error (MSE) as the objective function, while the mean absolute error (MAE) is used to monitor the convergence of the training process. In the model validation phase, the prediction accuracy is quantitatively evaluated on the test set using three error metrics: MSE, MAE, and RMSE. The coefficient of determination (R^2) is used to assess the ability of the model to explain the variance in the observed data. Pearson correlation coefficient (r) is applied to measure the degree of linear correlation between the predicted and observed values, while prediction bias is identified through a bias analysis. This comprehensive evaluation framework not only achieves a quantitative comparison between the predictions and observations but also provides a multidimensional model performance assessment, thereby offering a robust statistical basis for the reliability of the results (Li et al., 2024b; Lü et al., 2024).

The calculation methods for these statistical indicators are presented as follows:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (SM_i - \hat{SM}_i)^2 \quad (12)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |SM_i - \hat{SM}_i| \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (SM_i - \hat{SM}_i)^2} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (SM_i - \hat{SM}_i)^2}{\sum_{i=1}^N (SM_i - \overline{SM})^2} \quad (15)$$

$$r = \frac{\sum_{i=1}^N (SM_i - \overline{SM})(\hat{SM}_i - \overline{\hat{SM}})}{\sqrt{\sum_{i=1}^N (SM_i - \overline{SM})^2} \sqrt{\sum_{i=1}^N (\hat{SM}_i - \overline{\hat{SM}})^2}} \quad (16)$$

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N (\hat{SM}_i - SM_i) \quad (17)$$

where SM_i is the true SSM value; \overline{SM} is the mean value of SM_i ; \hat{SM}_i is the predicted or assimilated SSM value; $\overline{\hat{SM}}$ is the mean value of \hat{SM}_i ; and N is the number of measured or predicted SSM sample points. Lower values are better for MSE, MAE, RMSE, and the absolute value of Bias, and higher values are better for R^2 and r .

4.1 Model Training

This study utilized a daily SSM dataset with a spatial resolution of $0.05^\circ \times 0.05^\circ$ over the Qilian Mountains from 2017 to 2021 and divided it into strictly chronological and non-overlapping subsets: training (1 January 2017–31 May 2020; 1247 days), validation (1 June 2020–31 May 2021; 365 days), and testing sets (1 June 2021–31 December 2021; 214 days). The subsets were strictly time-ordered with no overlap, and sliding windows were constructed within each subset so that no window crosses a boundary.

We systematically evaluated the training dynamics and generalization of the model. We adopted the Adaptive Moment Estimation (Adam) Optimizer (initial learning rate = 0.017), used a 14-day input sequence and a 1-day prediction horizon, and compared four models: CNN, LSTM, ConvLSTM, and ConvLSTM-SE. The evolution trends exhibited by MSE and MAE over increasing training

epochs on the training set (solid lines) and validation set (dashed lines) are shown in Figure 5 [FIGURE:5]. All model values decrease rapidly during the initial epochs and then stabilize. ConvLSTM and ConvLSTM-SE achieve the lowest overall metrics and the smoothest curves, followed by CNN; LSTM starts with a higher loss and exhibits a brief spike before converging. According to Table 1, the numbers of convergence epochs are 19 for CNN, 25 for LSTM, 6 for ConvLSTM, and 12 for ConvLSTM-SE. The total training times required for 100 epochs are 0.1211 h (CNN), 0.1009 h (LSTM), 5.7206 h (ConvLSTM), and 6.3694 h (ConvLSTM-SE). The peak Graphics Processing Unit (GPU) memory usage levels during training are similar across the models, ranging from 8481.27 to 8820.38 MiB (approximately 7.84–8.61 GiB). However, the final file sizes of the models differ significantly: 0.11 MiB for CNN, 0.17 MiB for ConvLSTM, 0.40 MiB for ConvLSTM-SE, and 254.84 MiB for LSTM. The training and validation curves almost overlap, indicating negligible overfitting and strong training stability and generalizability. Considering both accuracy and stability, ConvLSTM-SE and ConvLSTM perform best overall.

4.2 Short-Term Forecasting

During the short-term prediction process, the sliding window-based prediction method was adopted; that is, a fixed-length input window slides gradually in the temporal dimension, and the observed data are used each time for model training and predicting the next moment. In this study, the observed soil moisture data derived from the first 14 days were used as model inputs each time to predict the soil moisture status on the 15th day.

As shown in Figure 6 [FIGURE:6], all four models capture the seasonal evolution trend of soil moisture, but CNN systematically underestimates the magnitude (persistent negative Bias), and the LSTM model also exhibits underestimation, with significant fluctuations in predictions during certain periods and a marked decrease in correlation. In contrast, ConvLSTM and ConvLSTM-SE track the true values closely, maintaining consistently low daily MSE, MAE, and RMSE and high correlations (R^2 and r). To facilitate a more intuitive comparison, we use a uniform color scale in Figure 7

to display the true values and predicted values (soil moisture color bars) of each model, as well as the prediction errors (error color bars). In the northeastern region and the mountain transition zone with significant humidity gradients, ConvLSTM and ConvLSTM-SE more accurately reproduce the observed spatial distribution and fine details, whereas CNN and LSTM exhibit more significant underestimation effects and striped residuals in high-moisture areas.

Aggregate metrics further confirm the advantage of spatiotemporal modeling: ConvLSTM achieves the best overall performance ($MAE = 0.0047 \text{ m}^3/\text{m}^3$, $RMSE = 0.0068 \text{ m}^3/\text{m}^3$, $R^2 = 0.9728$, and $r = 0.9869$), reducing MAE and RMSE by roughly 47.2% and 52.8%, respectively, relative to those of CNN

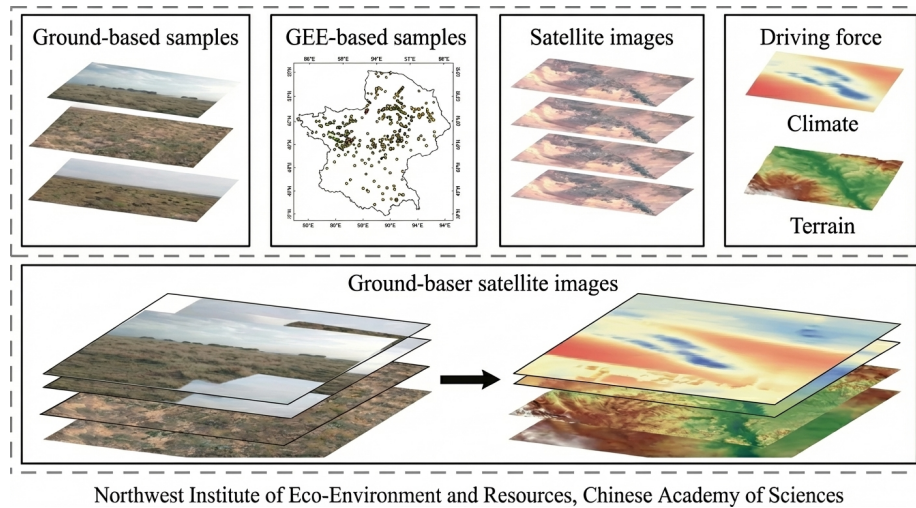


Figure 2: Figure 7

(Table 2). ConvLSTM-SE ranks second in terms of performance, with MAE of $0.0064 \text{ m}^3/\text{m}^3$ and RMSE of $0.0091 \text{ m}^3/\text{m}^3$, representing an approximately 36.8% RMSE reduction relative to that of CNN. In comparison, LSTM performs moderately, with MAE of $0.0072 \text{ m}^3/\text{m}^3$ and RMSE of $0.0097 \text{ m}^3/\text{m}^3$, placing it between ConvLSTM-SE and CNN in terms of accuracy. The Bias results indicate a notable underestimation effect for CNN ($-0.0039 \text{ m}^3/\text{m}^3$), whereas the other models yield Biases that are close to zero. Overall, ConvLSTM and ConvLSTM-SE, which explicitly capture spatiotemporal coupling, outperform the purely temporal (LSTM) and purely spatial (CNN) baselines in terms of accuracy, stability, and spatial fidelity.

4.3 Long-Term Forecasting

The long-term prediction process was experimented with using a recursive prediction method. Specifically, a fixed number of observations is selected for each prediction stage, and the results of the previous prediction are used as new inputs to iteratively predict future time periods. This method continuously feeds previous prediction results into the model, thereby having a lasting effect on subsequent predictions. In the specific experimental design, the observed data acquired from 1 June to 14 June 2021 were used to predict the soil moisture on 15 June 2021. The observed data from 2 June to 14 June 2021, together with the predicted value for 15 June 2021, were subsequently used as inputs to predict the soil moisture on 16 June 2021. This recursive process continued iteratively until the soil moisture prediction for 31 December 2021 was completed. Starting from 28 June 2021, the model input gradually transitioned to rely on

historical predictions, after which the forecasts were generated entirely on the basis of the model's previous outputs. This setup enabled evaluation of error accumulation and drift, and tested model robustness under conditions where no new observations or incomplete observational coverages are available.

Under the recursive setup, the long-range forecasts (Figure 8 [FIGURE:8] for time series and daily metrics, and Figure 9 [FIGURE:9] for spatial fields) exhibit characteristic error accumulation. The detailed performance results are summarized in Table 3. CNN tends to underestimate the results throughout the period, leading to a large negative Bias (Bias = $-0.0281 \text{ m}^3/\text{m}^3$), whereas LSTM shows a pronounced positive drift from September onward with a surge in errors, yielding a strongly degraded fit ($R^2 = -2.4026$ and $r = 0.7804$). In contrast, ConvLSTM and ConvLSTM-SE remain more robust during multistep recursion: their daily MSEs, MAEs, and RMSEs remain low and track the ground truth more closely. The overall performance metrics for ConvLSTM are: MAE = $0.0204 \text{ m}^3/\text{m}^3$, RMSE = $0.0282 \text{ m}^3/\text{m}^3$, $R^2 = 0.5362$, $r = 0.7997$, and Bias = $-0.0105 \text{ m}^3/\text{m}^3$, and for ConvLSTM-SE: MAE = $0.0236 \text{ m}^3/\text{m}^3$, RMSE = $0.0326 \text{ m}^3/\text{m}^3$, $R^2 = 0.3827$, $r = 0.7683$, and Bias = $-0.0139 \text{ m}^3/\text{m}^3$. Relative to CNN (MAE = $0.0375 \text{ m}^3/\text{m}^3$ and RMSE = $0.0492 \text{ m}^3/\text{m}^3$), ConvLSTM reduces MAE and RMSE by approximately 45.6% and 42.7%, respectively; and ConvLSTM-SE achieves an RMSE reduction of approximately 33.7%. In summary, the ConvLSTM model performs best in long-term recursive prediction, but it still has large errors because of the accumulation of multistep recursive errors and thus needs further optimization.

Given that long-term recursive forecasting still results in error accumulation and drift, this study extended the approach from a purely data-driven method to observation-constrained online correction (Nudging). We injected Gaussian random noise with a mean of zero and a variance of $\delta = 0.01$ into the observations to simulate uncertainty, and used noise-free observations ($\delta = 0.00$) as a control to systematically evaluate the feasibility and performance of the Nudging correction method under different observation quality conditions.

Under observation noise $\delta = 0.01$, based on the optimal Nudging coefficient τ for each model ($\tau = 0.7$ for CNN, $\tau = 0.6$ for LSTM, $\tau = 0.5$ for ConvLSTM, and $\tau = 0.6$ for ConvLSTM-SE), we evaluated different observation fractions (obs) (Table 4 ; Figs. 10-12). ConvLSTM-Nudging is optimal in terms of most metrics: when obs = 50.0%, MSE and RMSE are $0.0002 (\text{m}^3/\text{m}^3)^2$ and $0.0156 \text{ m}^3/\text{m}^3$, respectively, decreasing by 66.7% and 34.2% relative to those produced when obs = 10.0%; R^2 increases to 0.8591, and Bias approaches zero ($-0.0004 \text{ m}^3/\text{m}^3$), balancing accuracy and stability. LSTM-Nudging is most sensitive to sparse and noisy observations: at obs = 10.0%, it shows a pronounced positive Bias ($0.0874 \text{ m}^3/\text{m}^3$) and a negative R^2 (-4.4399); when obs = 50.0%, MSE and RMSE decrease to $0.0003 (\text{m}^3/\text{m}^3)^2$ and $0.0181 \text{ m}^3/\text{m}^3$, respectively, and R^2 (0.8091) increases significantly, indicating that sufficient observations significantly suppress error drift and enhance temporal correlations. The radar charts show that as observation fraction increases from 10.0% to 50.0%, all

four models monotonically improve in terms of MSE, MAE, and RMSE metrics, R^2 and r continuously increase, and Bias converges toward zero; among them, ConvLSTM-Nudging demonstrates the best consistency across all six metrics: error convergence, r , and R^2 improve synchronously, with the smoothest curve fluctuations, reflecting overall balance and robustness (Fig. 10 [FIGURE:10]). Notably, although ConvLSTM-SE-Nudging is superior overall to CNN-Nudging and LSTM-Nudging (except at low observation levels), it does not surpass ConvLSTM without SE, which indicates that channel attention may more easily overweight local noise under noisy observations; this effect weakens when observation fraction increases to 50.0%. The time series (Fig. 11 [FIGURE:11]) and spatial distributions (Fig. 12 [FIGURE:12]) further corroborate this finding: CNN-Nudging tends to exhibit amplitude compression during seasonal transitions, and LSTM-Nudging tends to yield platform Biases at low observation fraction values, whereas ConvLSTM-Nudging better tracks the amplitude and phase and produces smaller and more uniform error patches. Overall, Nudging effectively alleviates systematic underestimation or overestimation issues and reduces the areas with high error concentrations. From the parameter perspective, the optimal Nudging coefficient τ falls within 0.5-0.7; convolutional spatiotemporal models tend to adopt a smaller Nudging coefficient τ value (0.5) to avoid forcing instantaneous noise into the dynamics, whereas sequence memory-based LSTM relies more on observation constraints (with the greatest gains derived from increasing observation fraction). In summary, under observation noise $\delta = 0.01$, ConvLSTM-Nudging achieves the most robust overall performance, with the lowest errors, the highest R^2 , and a near-zero Bias, and all four models benefit significantly as observation fraction increases.

Under noise-free observations ($\delta = 0.00$), the four models were evaluated with their optimal Nudging coefficients all set to $\tau = 1.0$ (Table 5 ; Figs. 13-15). Overall, as the observation fraction increases from 10.0% to 50.0%, the error metrics (MSE, MAE, and RMSE) decrease monotonically, the correlation metrics (R^2 and r) increase synchronously, and Bias further approaches zero, showing that high-quality and sufficient observations significantly amplify the gains of Nudging. At obs = 50.0%, ConvLSTM-Nudging attains MSE below $0.0001 \text{ (m}^3/\text{m}^3)^2$, RMSE = $0.0064 \text{ m}^3/\text{m}^3$, $R^2 = 0.9762$, $r = 0.9881$, and Bias = $-0.0002 \text{ m}^3/\text{m}^3$, representing the best overall performance. Relative to CNN-Nudging (RMSE = $0.0120 \text{ m}^3/\text{m}^3$ and $R^2 = 0.9168$), the RMSE is reduced by 46.7%, and R^2 increases by 6.5%. In the time series, ConvLSTM-Nudging almost coincides with the ground truth (Fig. 14 [FIGURE:14]), while in space, it presents the smallest and most uniformly distributed error patches (Fig. 15 [FIGURE:15]). LSTM-Nudging remains sensitive to sparse constraints at low observation fractions (at obs = 10.0%, $R^2 = -2.1510$, and Bias = $0.0556 \text{ m}^3/\text{m}^3$) but improves most significantly as observation fraction increases: at obs = 50.0%, RMSE decreases to $0.0088 \text{ m}^3/\text{m}^3$, and R^2 increases to 0.9554, with $r = 0.9804$ and Bias = $0.0015 \text{ m}^3/\text{m}^3$, which indicates that sufficient observations effectively suppress its long-term recursive drift and enhance its temporal consistency. ConvLSTM-SE-Nudging provides the advantage of “channel attention” under

noise-free conditions: at obs = 50.0%, RMSE = 0.0086 m³/m³, R² = 0.9570, and r = 0.9783, which are comparable to those of LSTM-Nudging and significantly better than those of CNN-Nudging; this differs from the noisy case and indicates that the SE mechanism stably enhances the degree of expressiveness under observations with high signal-to-noise ratios. The radar charts provide consistent evidence that the polygons shrink significantly on error axes and expand on R² and r axes, and the Bias axis converges symmetrically around zero (Fig. 13 [FIGURE:13]). The time series shows that the amplitude compression and phase lag observed during seasonal transitions are essentially eliminated at obs = 50.0% (Fig. 14). The spatial distributions confirm that the areas with high error concentrations decrease significantly, and the residual errors are scattered mainly along transition zones with strong moisture gradients (Fig. 15). Notably, the optimal Nudging coefficient τ uniformly equals 1.0, which indicates that under noise-free conditions, “strong-constraint” correction does not introduce additional noise propagation effects and instead maximizes the model state correction implemented via observations. In summary, when observation noise $\delta = 0.00$, all four models benefit significantly from an increased observation fraction, with ConvLSTM-Nudging achieving the best overall performance, ConvLSTM-SE-Nudging and LSTM-Nudging following closely, and CNN-Nudging relatively behind.

5.1 Model Performance

This study systematically evaluates the predictive performance of four models—CNN, LSTM, ConvLSTM, and ConvLSTM-SE—for SSM at different temporal scales and conducts an in-depth analysis of differences in model structures, the effectiveness of the Nudging correction mechanism, and the effects of observation data quality and fraction on predictive performance.

In the short-term forecasting process, all four models track seasonal fluctuations, but their performances differ markedly (Figs. 6-7; Table 2). ConvLSTM and ConvLSTM-SE lead in both the temporal and spatial dimensions and better preserve gradients and fine textures in regions with strong humidity gradients, such as the northeastern region and mountainous transition zones. CNN exhibits systematic underestimation effects, whereas LSTM also exhibits underestimation, with significant fluctuations in predictions during certain periods and a marked decrease in correlation. Explicit spatiotemporal coupling more effectively extracts spatiotemporal information and, overall, outperforms single-dimensional (purely spatial or purely temporal) modeling (Fu et al., 2022; Ge et al., 2023; Zhang et al., 2023).

Over the 200-day long-term autoregressive forecasting horizon, all models exhibit typical error accumulation effects (Figs. 8-9; Table 3). CNN results in a persistent negative Bias (-0.0281 m³/m³). LSTM results in positive drift starting in September, which leads to a sharp decline in the goodness of fit of this

model ($R^2 = -2.4026$). ConvLSTM and ConvLSTM-SE remain relatively stable; the ConvLSTM model achieves MAE and RMSE of 0.0204 and $0.0282 \text{ m}^3/\text{m}^3$, respectively, reductions of 13.6% and 13.5% relative to those of ConvLSTM-SE, indicating the greater robustness of the ConvLSTM architecture in terms of capturing long-term spatiotemporal dependencies (Habiboullah and Louly, 2023; Huang et al., 2023). Although spatiotemporal coupling mitigates the rapid error divergence process to some extent, iterative error propagation in long-horizon autoregression constrains the predictive performance of ConvLSTM (Ding et al., 2025), leaving substantial room for improvement.

We addressed the error accumulation and drift encountered in long-horizon autoregressive forecasting scenarios by introducing the Nudging correction mechanism and systematically evaluating its effects under different observation accuracies ($\delta = 0.00$ and 0.01) and observation fractions (obs = 10.0%, 20.0%, and 50.0%) (Tables 4-5; Figs. 10-15) and summarizing the average metrics for an intuitive comparison (Fig. 16

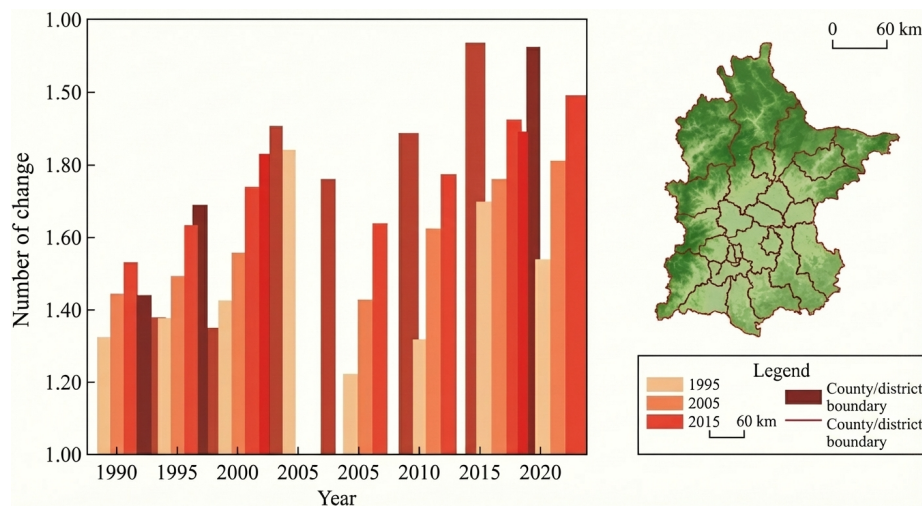


Figure 3: Figure 16

). Overall, CNN-Nudging provides a monotonic improvement as the observation fraction increases under both noise settings, reaching $\text{RMSE} = 0.0120 \text{ m}^3/\text{m}^3$ and $R^2 = 0.9168$ when $\delta = 0.00$ and obs = 50.0%. LSTM-Nudging is most sensitive to sparse and noisy observations: at obs = 10.0% and 20.0%, it yields negative R^2 values, and when $\delta = 0.01$ with obs = 10.0%, the corrected R^2 of -4.4399 is lower than the uncorrected R^2 of -2.4026 , indicating that observation noise may introduce noise or redundancy (Cohen et al., 2013; Goux et al., 2025); the metric improves markedly when observation fraction increases to 50.0%. ConvLSTM-SE-Nudging provides stable gains at high signal-to-noise ratios ($\delta = 0.00$) but tends to overweight local noise under noisy conditions (Liang et al., 2023; Brigato et al., 2025; Feng et al., 2025). ConvLSTM-Nudging

achieves the best overall performance. These results indicate that Nudging significantly suppresses the error accumulation effect in long-horizon autoregression tasks and enhances the stability and spatiotemporal consistency of ConvLSTM (Kozhushko et al., 2022; Antil et al., 2024), but its corrective power remains limited when the initial model errors are large or when the observation fraction is small (Pawar et al., 2020).

The observation fraction has a decisive effect on the effectiveness of Nudging (Fig. 16). As observation fraction increases from 10.0% to 50.0%, the error metrics of all models decrease monotonically, the correlation metrics continue to increase, and Bias approaches zero. Under observation noise $\delta = 0.01$, CNN-Nudging, LSTM-Nudging, ConvLSTM-Nudging, and ConvLSTM-SE-Nudging reduce their RMSEs by approximately 18.9%, 81.2%, 34.2%, and 48.8%, respectively, at $\text{obs} = 50.0\%$ relative to those at $\text{obs} = 10.0\%$; under $\delta = 0.00$, RMSE reductions are 49.4%, 88.0%, 61.2%, and 72.9%, respectively. These results indicate that introducing more observations leads to better correction effects for the models (Celik and Olson, 2023).

5.2 Limitations and Prospects

Although the hybrid ConvLSTM-Nudging model constructed in this study exhibits strong capabilities in the 200-day soil moisture forecasting task, it still has several limitations, which also point out directions for future research. For example, Niu et al. (2025) and Adewole et al. (2024) reported that land surface temperature (LST), root-zone soil moisture (10–40 cm), and the normalized difference vegetation index (NDVI) are key factors that influence soil moisture dynamics, while this study relied only on SSM as the model input. Notably, however, the daily SSM product used here was produced by Qu et al. (2019, 2021), Chai et al. (2020), and Hu et al. (2022) via multisource integration (e.g., leaf area index (LAI), broadband albedo, fractional vegetation cover, gross primary productivity (GPP), and evapotranspiration (ET)), so exogenous meteorological and remote sensing signals are implicitly encoded in the SSM field even though they are not explicitly ingested as drivers during forecasting. Moreover, although Carlson et al. (2024) and Çıbık et al. (2025) proposed strategies for adaptively selecting the Nudging coefficient, this study adopted an empirically fixed value based on previous literature (Pawar et al., 2020), and the adaptability of the model across different regions and environmental settings remained unverified. In the long-term experiment, we intentionally adopted a fully recursive and driver-free design: the model was initialized with the first 14 days of the test period and then rolled forward daily using only antecedent SSM states (and a static DEM), without any exogenous meteorological forcings. This choice (1) isolates the intrinsic predictability of soil moisture arising from state persistence and seasonal regularities; (2) avoids the covariate shifts and compounded forecast errors that would arise from externally predicted precipitation and temperature fields; and (3) enhances the operational robustness and reproducibility

of the model when high-resolution and latency-free drivers are unavailable; the parsimonious setting also reduces the risk of overfitting under recursive use. Although the current driverless design helps focus on the intrinsic dynamics of the system and improve its robustness, as Xu et al. (2021a) noted, uncertainty in Earth system predictions arises from diverse sources, and the accuracy of spatiotemporal predictions is influenced by a combination of the model structure, initial conditions, and exogenous variables. Therefore, the introduction of reliable exogenous drivers (such as precipitation, temperature, and radiation) in the future is expected to further improve the prediction skill of the model, provided that these data have the required spatiotemporal resolutions and coverage levels. In summary, future research could combine multisource data fusion, network structure optimization, and adaptive parameterization, focusing on data quality control (such as field validations and triple configuration evaluations as described by Xu et al., 2021b) and model uncertainty modeling to continuously improve the accuracy, robustness, and cross-scenario transferability of soil moisture predictions.

6 Conclusions

In this paper, a hybrid modeling framework that integrates the ConvLSTM deep learning model and Nudging data assimilation method was proposed to achieve short- and long-term SSM prediction in the Qilian Mountains. The results show that in short-term forecasting scenarios, ConvLSTM and ConvLSTM-SE models, which explicitly model spatiotemporal coupling relationships, significantly outperform the baseline models that rely solely on spatial or temporal modeling, demonstrating higher prediction accuracy and spatial consistency. In long-term dynamic autoregressive forecasting cases, despite the inevitable accumulation of errors, ConvLSTM series models still exhibit stronger robustness. Further analysis reveals that the Nudging correction mechanism effectively suppresses systematic biases and error divergence issues in long-term autoregressive predictions, especially under conditions of no noise or high observation fractions.

The observation fraction is a key factor that influences the Nudging effect, and increasing the observation fraction generally helps improve the performance of the models, with ConvLSTM-Nudging consistently maintaining the best performance across all experimental conditions.

Conflicts of 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

This research was funded by the National Natural Science Foundation of China (42461053), the Department of Education of Gansu Province: Higher Education Innovation Fund Project (2023B-064), the Youth Doctoral Fund Project (2024QB-014), and the Natural Science Foundation of Gansu Province (25JRRA012).

Author Contributions

Conceptualization: FAN Manhong, XIAO Qian; **Data curation:** XIAO Qian, ZHAO Junhao; **Methodology:** FAN Manhong, XIAO Qian; **Investigation:** XIAO Qian, YU Qinghe; **Formal analysis:** FAN Manhong, XIAO Qian; **Writing - original draft preparation:** XIAO Qian; **Writing - review and editing:** FAN Manhong, XIAO Qian; **Funding acquisition:** FAN Manhong; **Resources:** FAN Manhong, ZHAO Junhao; **Supervision:** FAN Manhong; **Project administration:** FAN Manhong; **Software:** XIAO Qian, YU Qinghe; **Validation:** XIAO Qian, YU Qinghe; **Visualization:** XIAO Qian, ZHAO Junhao. All authors approved the manuscript.

References

- Adewole A O, Eludoyin A O, Chirima G J, et al. 2024. Field-scale variability and dynamics of soil moisture in southwestern Nigeria. *Discover Soil*, 25(4): 1827-1847.
- Anthes R A. 1974. Data assimilation and initialization of hurricane prediction models. *Journal of Atmospheric Sciences*, 31(3): 702-719.
- Antil H, Löhner R, Price R. 2024. Data assimilation with deep neural nets informed by Nudging. In: Rozza G, Stabile G, Gunzburger M, et al. *Reduction, Approximation, Machine Learning, Surrogates, Emulators and Simulators: RAMSES*. Cham: Springer Nature, 17-41.
- Brigato L, Morand R, Strømmen K, et al. 2025. Position: There are no champions in long-term time series forecasting. [2025-01-22]. <https://arxiv.org/pdf/2502.14045>.
- Carlson E, Farhat A, Martinez V R, et al. 2024. On the Infinite-Nudging Limit of the Nudging Filter for Continuous Data Assimilation. [2025-01-22]. <https://arxiv.org/pdf/2408.02646>.
- Celik E, Olson E. 2023. Data assimilation using time-delay Nudging in the presence of gaussian noise. *Journal of Nonlinear Science*, 33: 110, doi: 10.1007/s00332-023-09967-1.
- Chai L N, Zhu Z L, Liu S M. 2020. Daily 0.05\$×\$0.05 Land Surface Soil Moisture Dataset of Qilian Mountain Area (2018, SMHiRes, V1). National Tibetan

- Plateau Data Center. [2025-01-20]. <https://doi.org/10.11888/Soil.tpd.270408>.
- Charney J G, Fjörtoft R, Neumann J V. 1950. Numerical integration of the barotropic vorticity equation. *Tellus*, 2(4): 237-254.
- Cheng S H, Engel B A, Liu R, et al. 2023. Impedance factor of hydraulic conductivity for frozen soil based on ice segregation theory and its application. *Water Resources Research*, 59(6): e2022WR033876, doi: 10.1029/2022WR033876.
- Çıbık A, Fang R, Layton W, et al. 2025. Adaptive parameter selection in Nudging based data assimilation. *Computer Methods in Applied Mechanics and Engineering*, 433: 117526, doi: 10.1016/j.cma.2024.117526.
- Cohen A, Davenport M A, Leviatan D. 2013. On the stability and accuracy of least squares approximations. *Foundations of Computational Mathematics*, 13(5): 819-834.
- Connor J T, Martin R D, Atlas L E. 1994. Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5(2): 240-254.
- Conti G, Aydoğdu A, Gualdi S, et al. 2022. On the physical nudging equations. *Climate Dynamics*, 58(5): 1459-1476.
- Ding L, Bai Y, Fan M H, et al. 2024. Using a snow ablation optimizer in an autonomous echo state network for the model-free prediction of chaotic systems. *Nonlinear Dynamics*, 112(13): 11483-11500.
- Ding L, Bai Y L, Zheng D H, et al. 2025. Chaotic climate system forecasting using an improved echo state network with sparse observations. *Science China Earth Sciences*, 68(7): 2346-2360.
- Ek M B, Mitchell K E, Lin Y, et al. 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, 108(D22): 8851, doi: 10.1029/2002JD003296.
- Entin J K, Robock A, Vinnikov K Y, et al. 2000. Temporal and spatial scales of observed soil moisture variations in the extratropics. *Journal of Geophysical Research: Atmospheres*, 105(D9): 11865-11877.
- Feng T, Ni J, Gleichgerrcht E, et al. 2025. SeizureFormer: A Transformer Model for IEA-Based Seizure Risk Forecasting. [2025-01-22]. <https://arxiv.org/pdf/2504.16098v1>.
- Fu E, Zhang Y N, Yang F, et al. 2022. Temporal self-attention-based ConvLSTM network for multivariate time series prediction. *Neurocomputing*, 501: 162-173.
- Gamboa-Villafruela C J, Fernández-Alvarez J C, Márquez-Mijares M, et al. 2021. Convolutional LSTM Architecture for Precipitation Nowcasting Using Satellite

Data. *Environmental Sciences Proceedings*, 8(1): 33, doi: 10.3390/ecas2021-10340.

Ge M Y, Gao W, Zhu M, et al. 2023. Sea ice classification of SAR images based on SE-ConvLSTM spatial-temporal feature fusion. *Remote Sensing Technology and Application*, 38(6): 1306–1316. (in Chinese)

Goux O, Weaver A T, Gürol S, et al. 2025. On the Impact of Observation Error Correlations in Data Assimilation, with Application to along-Track Altimeter Data. *Quarterly Journal of the Royal Meteorological Society*. [2025-01-12]. <https://arxiv.org/pdf/2503.09140>.

Gruber A, Lannoy G D, Crow W. 2019. A Monte Carlo based adaptive Kalman filter framework for soil moisture data assimilation. *Remote Sensing of Environment*, 228(2019): 105–114.

Habiboullah A, Louly M A. 2023. Soil moisture prediction using NDVI and NSMI satellite data: Vit-based models and ConvLSTM-based model. *SN Computer Science*, 4: 140, doi: 10.1007/s42979-022-01554-7.

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

Hu J, Shen L, Sun G. 2018. Squeeze-and-excitation networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society/Computer Vision Foundation. Salt Lake City, USA, 7132–7141.

Hu Z X, Chai L N, Crow W T, et al. 2022. Applying a wavelet transform technique to optimize general fitting models for SM analysis: a case study in downscaling over the Qinghai-Tibet Plateau. *Remote Sensing*, 14(13): 3063, doi: 10.3390/rs14133063.

Huang F N, Zhang Y K, Zhang Y, et al. 2023. Interpreting Conv-LSTM for spatio-temporal soil moisture prediction in China. *Agriculture*, 13(5): 971, doi: 10.3390/agriculture13050971.

Jin X, Xie Y, Wei X S, et al. 2022. Delving deep into spatial pooling for squeeze-and-excitation networks. *Pattern Recognition*, 121(2022): 108159, doi: 10.1016/j.patcog.2021.108159.

Koster R D, Dirmeyer P A, Guo Z C, et al. 2004. Regions of strong coupling between soil moisture and precipitation. *Science*, 305(5687): 1138–1140.

Kozhushko O, Boiko M, Kovbasa M Y, et al. 2022. Field scale computer modeling of soil moisture with dynamic Nudging assimilation algorithm. *Mathematical Modeling and Computing*, 9(2): 203–216.

LeCun Y. 1989. Generalization and Network Design Strategies. [2025-01-18]. <https://masters.donmtu.ru/2012/fknt/umiarov/library/lecun.pdf>.

- Li L, Dai Y J, Wei Z W, et al. 2024a. Enhancing deep learning soil moisture forecasting models by integrating physics-based models. *Advances in Atmospheric Sciences*, 41(7): 1326-1341.
- Li Q L, Li Z Y, Shangguan W, et al. 2022a. Improving soil moisture prediction using a novel encoder-decoder model with residual learning. *Computers and Electronics in Agriculture*, 195: 106816, doi: 10.1016/j.compag.2022.106816.
- Li Q L, Zhu Y H, Shangguan W, et al. 2022b. An attention-aware LSTM model for soil moisture and soil temperature prediction. *Geoderma*, 409: 115651, doi: 10.1016/j.geoderma.2021.115651.
- Li S L, Han Y, Li C X, et al. 2024b. A novel framework for multi-layer soil moisture estimation with high spatio-temporal resolution based on data fusion and automated machine learning. *Agricultural Water Management*, 306: 109173, doi: 10.1016/j.agwat.2024.109173.
- Liang D J, Zhang H X, Yuan D F, et al. 2023. Does Long-Term Series Forecasting Need Complex Attention and Extra Long Inputs? [2025-01-22]. <https://arxiv.org/pdf/2306.05035>.
- Lin P F, He Z B, Du J, et al. 2017. Recent changes in daily climate extremes in an arid mountain region, a case study in northwestern China's Qilian Mountains. *Scientific Reports*, 7(1): 2245, doi: 10.1038/s41598-017-02345-4.
- Liu L Y, Gou X H, Wang X J, et al. 2024. Relationship between extreme climate and vegetation in arid and semi-arid mountains in China: A case study of the Qilian Mountains. *Agricultural and Forest Meteorology*, 348: 109938, doi: 10.1016/j.agrformet.2024.109938.
- Lü X B, Nurmemet I, Xiao S T, et al. 2024. Spatial-temporal simulation and prediction of root zone soil moisture based on Hydrus-1D and CNN-LSTM-attention models in the Yutian Oasis, southern Xinjiang, China. *Pedosphere*, 35(5): 846-857.
- Margulis S A, McLaughlin D, Entekhabi D, et al. 2002. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resources Research*, 38(12): 1299, doi: 10.1029/2001WR001114.
- Meng X J, Mao K B, Meng F, et al. 2021. A fine-resolution soil moisture dataset for China in 2002-2018. *Earth System Science Data*, 13(7): 3239-3261.
- Menut L, Bessagnet B, Cholakian A, et al. 2024. What is the relative impact of nudging and online coupling on meteorological variables, pollutant concentrations and aerosol optical properties? *Geoscientific Model Development*, 17(9): 3645-3665.
- Mikolov T, Kombrink S, Burget L, et al. 2011. Extensions of recurrent neural network language model. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE Signal Processing Society.

Prague, Czech Republic, 5528–5531.

Niu J Q, Liu Z J, Chen F Y, et al. 2025. Variations of soil moisture and its influencing factors in arid and semi-arid areas, China. *Journal of Arid Land*, 17(5): 624–643.

Pawar S, Ahmed S E, San O, et al. 2020. Long short-term memory embedded nudging schemes for nonlinear data assimilation of geophysical flows. *Physics of Fluids*, 32(7): 076606, doi: 10.1063/5.0012853.

Qu Y Q, Zhu Z L, Chai L N, et al. 2019. Rebuilding a microwave soil moisture product using random forest adopting AMSR-E/AMSR2 brightness temperature and SMAP over the Qinghai-Tibet Plateau, China. *Remote Sensing*, 11(6): 683, doi: 10.3390/rs11060683.

Qu Y Q, Zhu Z L, Montzka C, et al. 2021. Inter-comparison of several soil moisture downscaling methods over the Qinghai-Tibet Plateau, China. *Journal of Hydrology*, 592: 125616, doi: 10.1016/j.jhydrol.2020.125616.

Richards L A. 1931. Capillary conduction of liquids through porous mediums. *Physics*, 1(5): 318–333.

Shi X J, Chen Z R, Wang H, et al. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Forecasting. [2025-01-22]. <https://arxiv.org/pdf/1506.04214>.

Šimůnek J, van Genuchten M T, Šejna M. 2008. Development and applications of the HYDRUS and STANMOD software packages and related codes. *Vadose Zone Journal*, 7(2): 587–600.

Song X Y, Liu Y T, Xue L, et al. 2020. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *Journal of Petroleum Science and Engineering*, 186: 106682, doi: 10.1016/j.petrol.2019.106682.

Stauffer D R, Seaman N L. 1990. Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: experiments with synoptic-scale data. *Monthly Weather Review*, 118(6): 1250–1277.

van Dam J C, Groenendijk P, Hendriks R F A, et al. 2008. Advances of modeling water flow in variably saturated soils with SWAP. *Vadose Zone Journal*, 7(2): 640–653.

Vereecken H, Huisman J A, Bogena H, et al. 2008. On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resources Research*, 44(4): W00D06, doi: 10.1029/2008WR006829.

Wang C J, Liu Q M, Yin C S, et al. 2023. Simulation of soil moisture based on ensemble Kalman filter assimilation method and HYDRUS-1D model. *Agricultural Research in the Arid Areas*, 41(2): 141–149. (in Chinese)

- Wang Y K, Shi L S, Lin L, et al. 2020. A robust data-worth analysis framework for soil moisture flow by hybridizing sequential data assimilation and machine learning. *Vadose Zone Journal*, 19(1): e20026, doi: 10.1002/vzj2.20026.
- Wang Y L, Shi L S, Hu Y A, et al. 2024. A comprehensive study of deep learning for soil moisture prediction. *Hydrology and Earth System Sciences*, 28(4): 917–943.
- Xu L, Abbaszadeh P, Moradkhani H, et al. 2020. Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index. *Remote Sensing of Environment*, 250: 112028, doi: 10.1016/j.rse.2020.112028.
- Xu L, Chen N C, Chen Z Q, et al. 2021a. Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions. *Earth-Science Reviews*, 222: 103828, doi: 10.1016/j.earscirev.2021.103828.
- Xu L, Chen N C, Zhang X, et al. 2021b. In-situ and triple-collocation based evaluations of eight global root zone soil moisture products. *Remote Sensing of Environment*, 254: 112248, doi: 10.1016/j.rse.2020.112248.
- Yu J X, Zhang X, Xu L L, et al. 2021. A hybrid CNN-GRU model for predicting soil moisture in maize root zone. *Agricultural Water Management*, 245: 106649, doi: 10.1016/j.agwat.2020.106649.
- Zhang J T, Zhang C Q, Feng Q. 2020. Digital Elevation Model (DEM) Data with the Spatial Resolution of 30 m in the Qilian Mountains. National Cryosphere Desert Data Center. [2025-01-12]. <https://doi.org/10.12072/ncdc.qlsst.db0007.2020>.
- Zhang S Y, Deng Y S, Niu Q R, et al. 2023. Multivariate temporal self-attention network for subsurface thermohaline structure reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 4507116, doi: 10.1109/TGRS.2023.3320350.
- Zhou X R, Li Y. 2022. Response of dry-wet change to millennial and centennial warm periods in the Qilian Mountains. *Acta Geographica Sinica*, 77(5): 1138–1152. (in Chinese)
- Zhu P H, Shi L S, Zhu Y, et al. 2017. Data assimilation of soil water flow via ensemble Kalman filter: Infusing soil moisture data at different scales. *Journal of Hydrology*, 555(2017): 912–925.

Figures

Source: ChinaXiv – Machine translation. Verify with original.

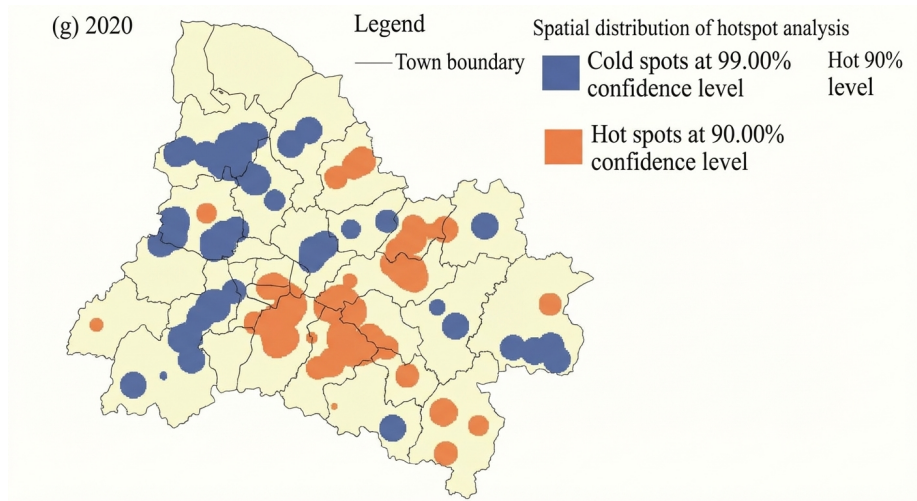


Figure 4: Figure 25

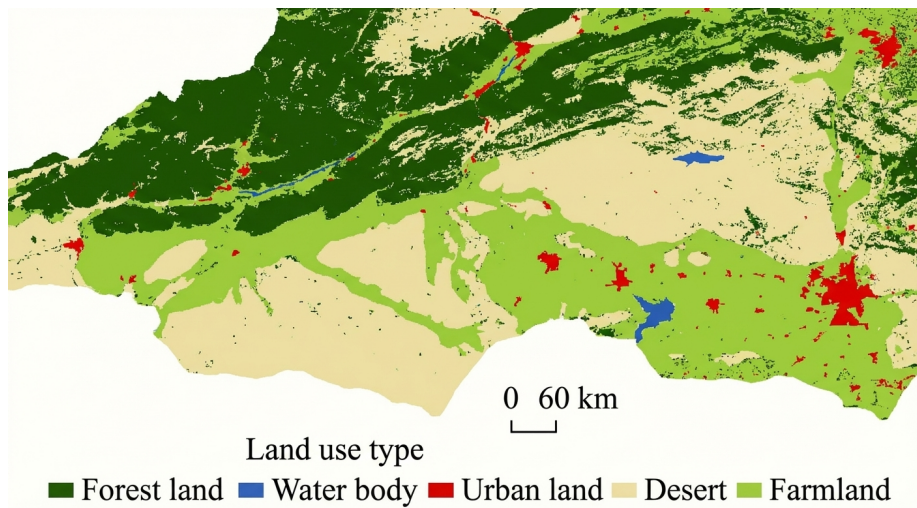


Figure 5: Figure 33

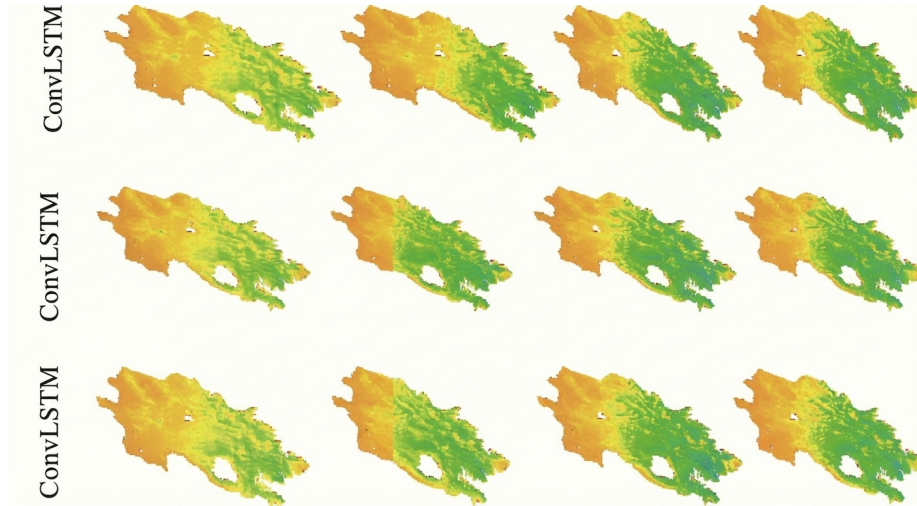


Figure 6: Figure 34

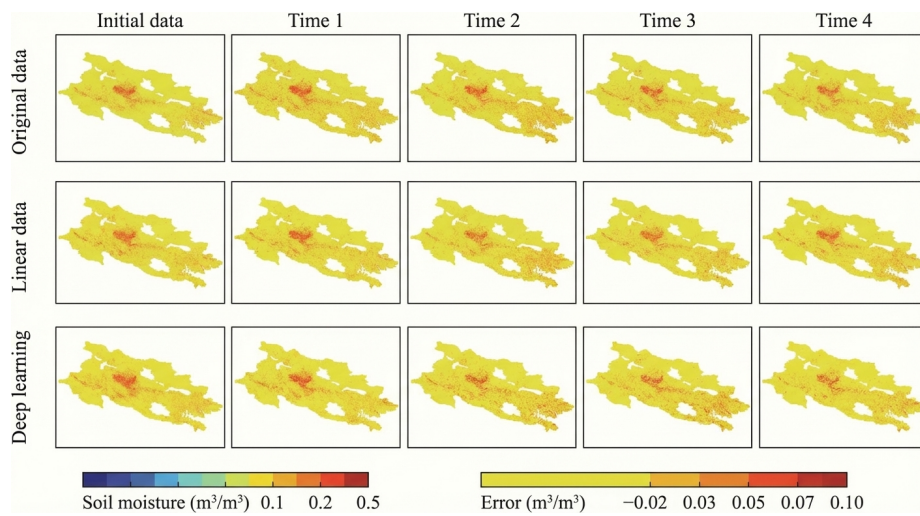


Figure 7: Figure 43

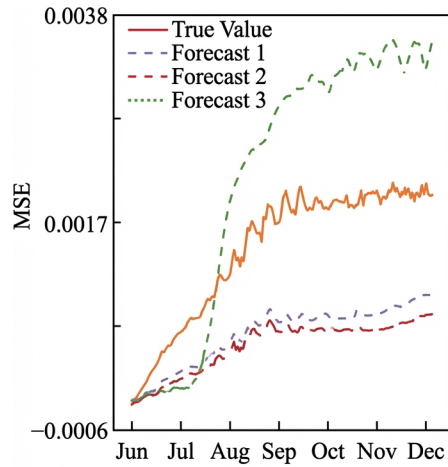


Fig. 6 MSE of Multi-step Forecasting

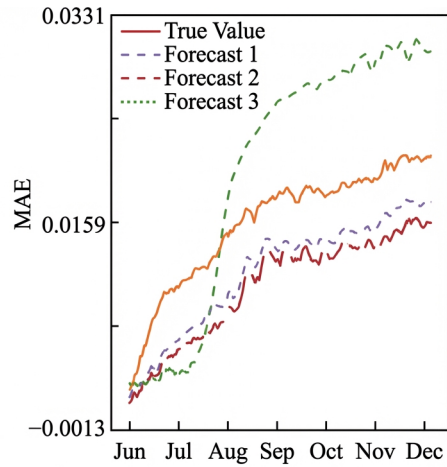


Fig. 7 MAE of Multi-step Forecasting

Figure 8: Figure 48

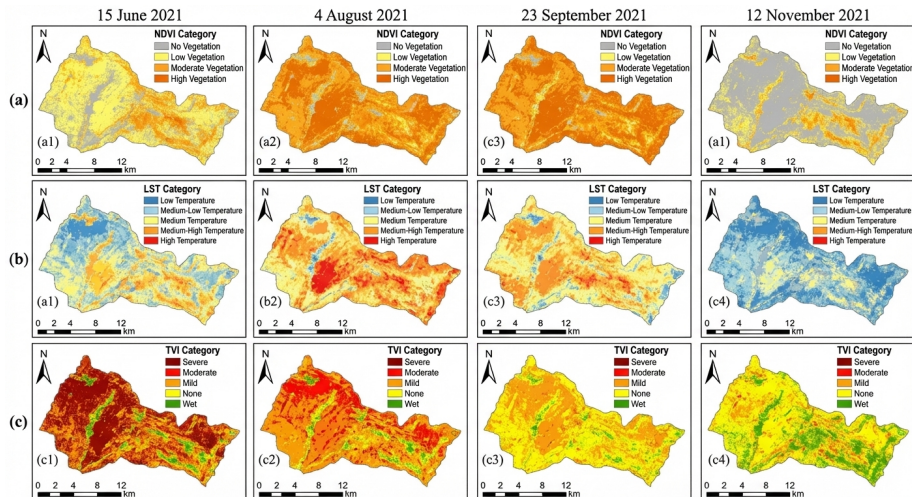


Figure 9: Figure 58

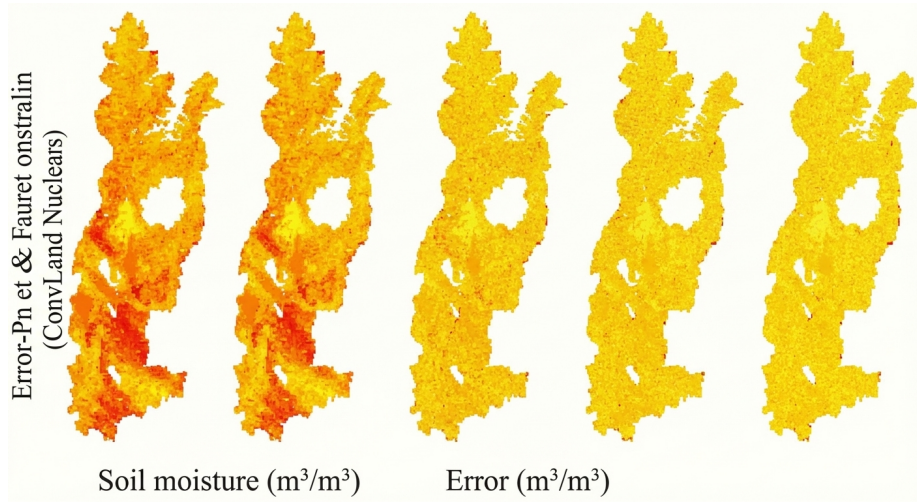


Figure 10: Figure 69

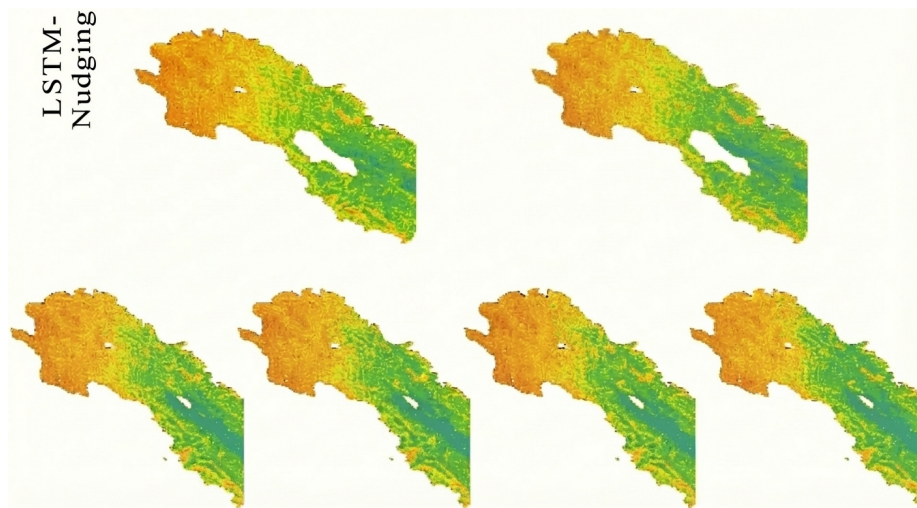


Figure 11: Figure 70

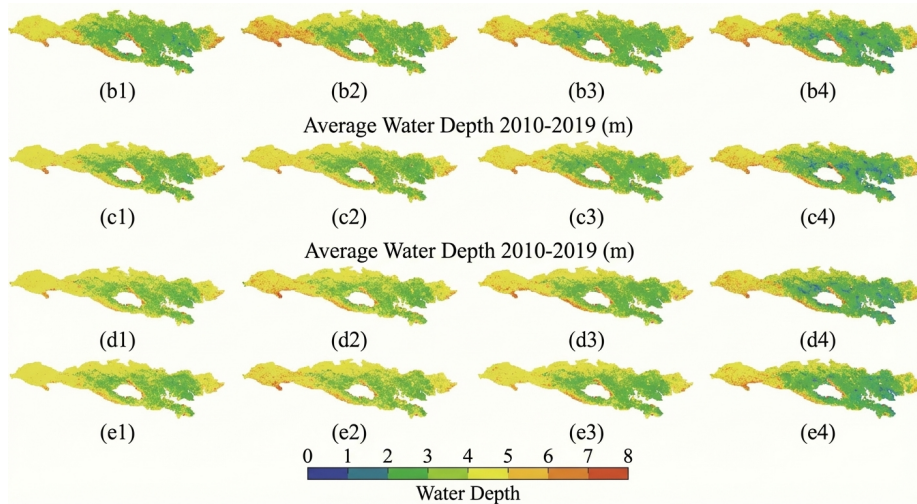


Figure 12: Figure 71

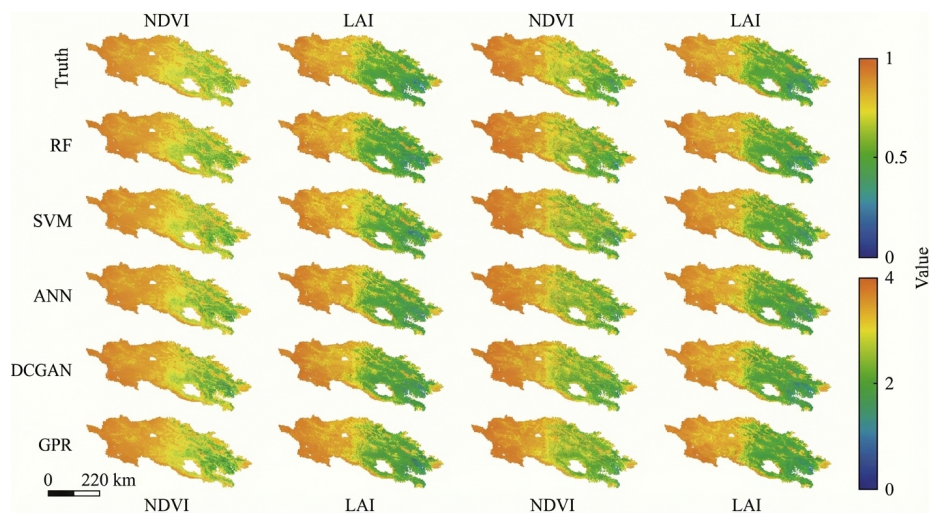


Figure 13: Figure 73

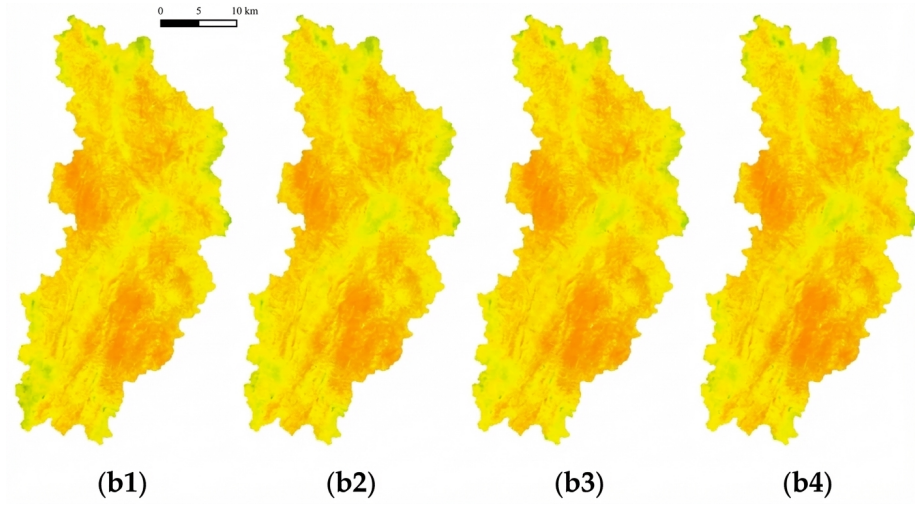


Figure 14: Figure 74

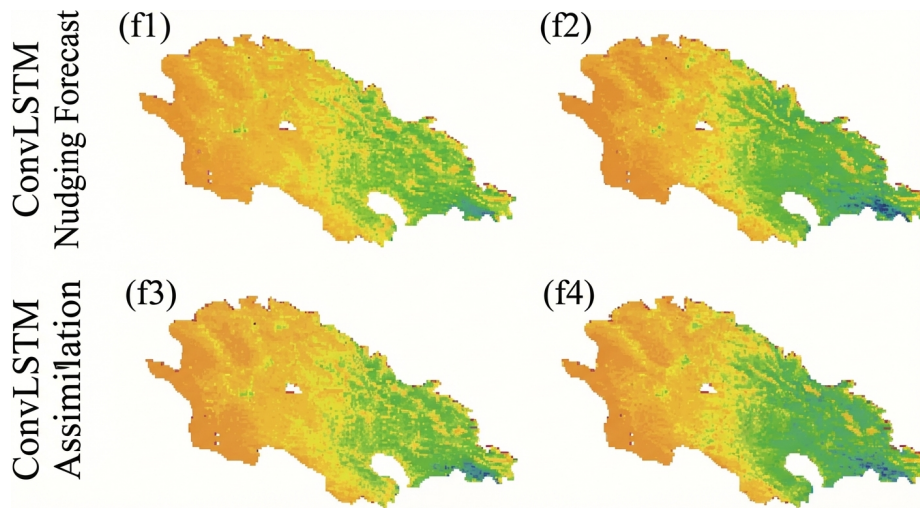


Figure 15: Figure 83