https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Decomposing Non-Stationarity for Soil Moisture Prediction: A Hybrid Statistical-Deep Learning Framework Integrating Periodic, Trend, and Residual Components

Chenlu Yu[1], Dong Wang[*1], Vijay P. Singh[2], Pengcheng Xu[3], Chenjian Yan[1], Xiankui Zeng[1], Jianguo Jiang[1], Mei Li[4], Qun Li[4], Shengtian Zhang[4], Jichun Wu[1]

[1]Key Laboratory of Surficial Geochemistry, Ministry of Education, Department of Hydrosciences, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, P.R. China
[2]6848 Truxton Drive, Dallas, Texas 75231, USA
[3]College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou, P.R. China
[4]Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of the People's Republic of China, Nanjing, P.R. China

Correspondence to: Dong Wang (wangdong@nju.edu.cn)

Abstract. The non-stationarity of hydro-meteorological processes has intensified under climate change, posing substantial challenges for accurate soil moisture (SM) prediction. To address the difficulty in distinguishing different sources of non-stationarity in conventional models, this study proposed a non-stationary SM prediction framework that integrated statistical methods with deep learning (DL) models. The framework focused on separating three types of non-stationary components—periodic fluctuations, long-term trend variations, and residual disturbances—and evaluated their effects on predictive performance. Specifically, independent models were first developed for different months to explicitly characterize periodic features. Statistical methods were then used to construct marginal distributions of hydro-meteorological variables, with physical covariates incorporated into distribution parameters to represent time-varying trends. Furthermore, Natural Gradient Boosting (NGBoost) was employed to perform secondary fitting of prediction residuals, thereby correcting random fluctuations that were not fully captured in the one-stage prediction. A gated recurrent unit with an attention mechanism (GRU-Attention) was selected as the core DL model, and SM prediction experiments were conducted across China to systematically evaluate the two-stage prediction performance under six modeling scenarios. Results showed that: (1) seasonal periodicity was the most robust predictive structure, whereas trend-related non-stationarity provided conditional rather than universal gain and residual learning improved performance only when the first-stage errors retained learnable structure; (2) the prediction skill was strongly organized by aridity, with the most stable performance concentrated in arid and semi-arid regions, while representative-grid and temporal-stability analyses revealed important local departures from regional mean predictability; and (3) the SHAP attribution identified a coherent dry-to-wet transition in dominant controls, from evaporation in hyper-arid regions to antecedent soil-moisture memory in arid regions and precipitation dominance in

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

wetter regions, with tree-based importance rankings supporting the robustness of this driver hierarchy. From the perspective of non-stationary component separation, this study provides a new hybrid modeling approach for SM prediction and offers methodological support for hydrological forecasting, agricultural drought monitoring, and regional water resources management under complex climate change conditions.

1 Introduction

Soil moisture (SM) is a key land-surface state variable that regulates exchanges of water, energy and carbon between the land and atmosphere (Abbes et al., 2024; Shawon et al., 2025). By controlling the partitioning of precipitation into infiltration, evapotranspiration, and runoff, SM shapes drought development, vegetation stress, irrigation demand, flood generation, and land-atmosphere feedbacks (Piao et al., 2010; Seneviratne et al., 2010; Sun et al., 2025). Accurate SM prediction is therefore essential for agricultural drought warning, water-resources management, landslide early warning and climate-risk assessment (Dari et al., 2024; Palazzolo et al., 2025).

Climate change and human disturbance are increasingly weakening the stationarity assumptions that have long supported hydrological prediction. Stationarity assumes that the statistical properties of hydrological variables remain invariant through time, but warming, circulation shifts, land-cover change and water management can alter means, variances, extremes and dependence structures (Apurv & Cai, 2019; Blöschl et al., 2019; Luke et al., 2017; Milly et al., 2008). These changes propagate into SM, because it is a land-surface response to precipitation, radiation, evaporative demand, vegetation dynamics, and antecedent water storage. Climate-model and observational studies show that soil water availability can change unevenly across depths, regions and seasons, which complicates the use of stationary prediction rules (Berg et al., 2017; Priya et al., 2024; You et al., 2025). In SM modelling, Kathuria et al. (2019) showed that spatial variance and correlation structures can become non-stationary under surface heterogeneity, while Zhu et al. (2025) demonstrated that explicitly extracting periodic SM signals can improve global forecasting. Baumberger et al. (2025) further showed that recurrent neural networks could model soil temperature and moisture time series, but their process reflectivity still required explicit evaluation. These studies indicate that non-stationary SM prediction is no longer a peripheral extension of stationary modelling. It is becoming a necessary response to changing hydroclimatic conditions.

The main unresolved issue is that SM non-stationarity is multidimensional rather than a single trend-like disturbance. At least three forms are relevant. Periodic non-stationarity reflects the recurrent seasonal organization of SM, which arises from annual cycles in precipitation, radiation, evapotranspiration and vegetation activity. Fourier-based, wavelet-based, and seasonal-decomposition approaches show that isolating periodic or multi-scale structure can reduce the burden placed on sequence models and improve SM, drought or hydrological forecasting (Rezaiy & Shabri, 2025; Verbesselt et al., 2010; J. Zhu et al., 2025). Trend-related non-stationarity reflects gradual or covariate-linked changes in marginal distributions, including shifts in mean wetness, variability or extremes. This form is commonly represented through time-varying statistical models, non-stationary frequency analysis or distributional regression (Apurv & Cai, 2019; W. Li et al., 2026;

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Rigby & Stasinopoulos, 2005). Residual non-stationarity reflects the remaining structured errors after a primary model has captured dominant seasonal and trend components. Residual learning can improve nonlinear online prediction when past errors contain stable information, but it can amplify noise when residuals are weakly structured (Ilhan et al., 2024). In hydrology, decomposition and hybrid modelling have been applied to streamflow, drought and flood-frequency prediction, yet comparable work in SM prediction remains limited and usually focuses on one component at a time (W. Li et al., 2026; Liu et al., 2024; Z. Wang et al., 2024). A key gap is therefore not merely whether SM is non-stationary, but how non-stationary components differ in their contribution to predictive skill.

The second core problem is that no single prediction method can simultaneously represent distributional shifts, nonlinear temporal memory, and residual error structure, and addressing this gap requires combining statistical structure with deep learning rather than treating them as competing alternatives. Statistical distributional models are well suited to trend-related non-stationarity, because their parameters can be linked to covariates or time. They provide interpretable descriptions of changes in marginal behavior, including location and scale (Rigby & Stasinopoulos, 2005). However, these models are less able to learn high-dimensional nonlinear memory from hydro-meteorological predictors. Deep learning (DL) models, including recurrent, gated, and attention-based architectures, can learn such memory from large data sets and have substantially advanced hydrological prediction (Feng et al., 2020; Liu et al., 2025; Nearing et al., 2021). Their weakness is that high predictive skill does not automatically identify the hydrological source of that skill, especially under non-stationary conditions. Recent studies therefore increasingly favor hybrid frameworks that combine process knowledge, statistical constraints or multi-source information with DL flexibility (Liu et al., 2024; Shen et al., 2023; Z. Wang et al., 2024). Residual learning can also be viewed as a post-processing step that calibrates the remaining relationship between first-stage predictions and observations. Learners, such as NGBoost, are useful for this purpose, because they estimate conditional distributions, but they add value only when the first-stage residuals are systematic rather than random (Duan et al., 2020; Ilhan et al., 2024). For non-stationary SM prediction, a defensible hybrid framework should therefore separate distributional shifts, learn nonlinear temporal dependence, and explicitly test whether residual correction adds information.

A further challenge is to distinguish physically meaningful predictors from variables that are merely useful to a fitted model. Correlation screening metrics can select predictors that improve apparent accuracy, but they may not identify variables that participate in the underlying hydrological dynamics. Convergent cross mapping (CCM), based on nonlinear state-space reconstruction, offers a stricter test, because it asks whether one variable contains recoverable information about another within a shared dynamical attractor (Sugihara et al., 2012). Recent hydrological applications have used CCM and related causal methods to diagnose groundwater-streamflow interactions, SM-precipitation coupling, drought propagation, evapotranspiration controls and seasonal shifts in catchment causality (Ban et al., 2025; Bonotto et al., 2022; Pang & Yang, 2025; J. Yu et al., 2026; Zhao et al., 2024). In parallel, explainable artificial intelligence has become important for interpreting hydrological prediction models. SHapley Additive exPlanations (SHAP) quantify how a trained model assigns predictive sensitivity to individual inputs (Lundberg et al., 2020; Lundberg & Lee, 2017), and recent hydrological and SM studies have used SHAP to reveal regional differences in dominant controls and site-specific predictor behavior (Hosseini et

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

al., 2025; Nikraftar et al., 2025). These two perspectives are complementary rather than interchangeable: CCM targets causal participation in system dynamics, whereas SHAP targets contributions within a fitted prediction model. Combining them is especially useful for SM prediction, because antecedent SM, precipitation, evaporation, and radiation can differ in causal relevance, predictive contribution, and hydroclimatic dependence.

Here, we developed a statistical-DL hybrid framework to evaluate the predictive roles of periodic, trend-related, and residual components in non-stationary SM prediction. We applied it to monthly SM across mainland China from 1950 to 2020, compared six model configurations and evaluated performance across aridity zones and representative grid cells. The study addressed three questions: (i) Which non-stationary structures provide transferable predictive information in SM series; (ii) where the proposed framework remains robust across aridity zones and representative local grids; and (iii) how dominant hydro-meteorological controls shift across the dry-to-wet gradient. The remainder of the paper is organized as follows: Sect. 2 describes the framework and evaluation design, Sect. 3 introduces the study area and data, Sect. 4 presents the prediction and attribution results, Sect. 5 discusses mechanisms and limitations, and Sect. 6 summarizes the conclusions.

2 Methodology

The methodological framework consisted of six linked modules, considering three non-stationary components (Fig. 1). First, candidate predictors were screened using convergent cross mapping (CCM) so that the forecasting models relied on variables containing nonlinear dynamical information, rather than on variables that were only statistically correlated with SM. Second, generalized additive models for location, scale and shape (GAMLSS) were used to represent monthly, distributional and trend-related non-stationarity through marginal cumulative distribution transformations. Third, gated recurrent unit (GRU) networks with an attention layer were used for one-stage prediction, followed by a second-stage residual learning module (natural-gradient boosting, NGBoost) to assess whether the remaining prediction errors retained learnable structure. Fourth, prediction skill was evaluated across five aridity zones and representative grids to quantify spatial robustness and local predictability. Finally, SHapley Additive exPlanations (SHAP), together with alternative feature-selection diagnostics, were used to identify dominant controls and evaluate the robustness of model attribution. The methodological details of each module are described below.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

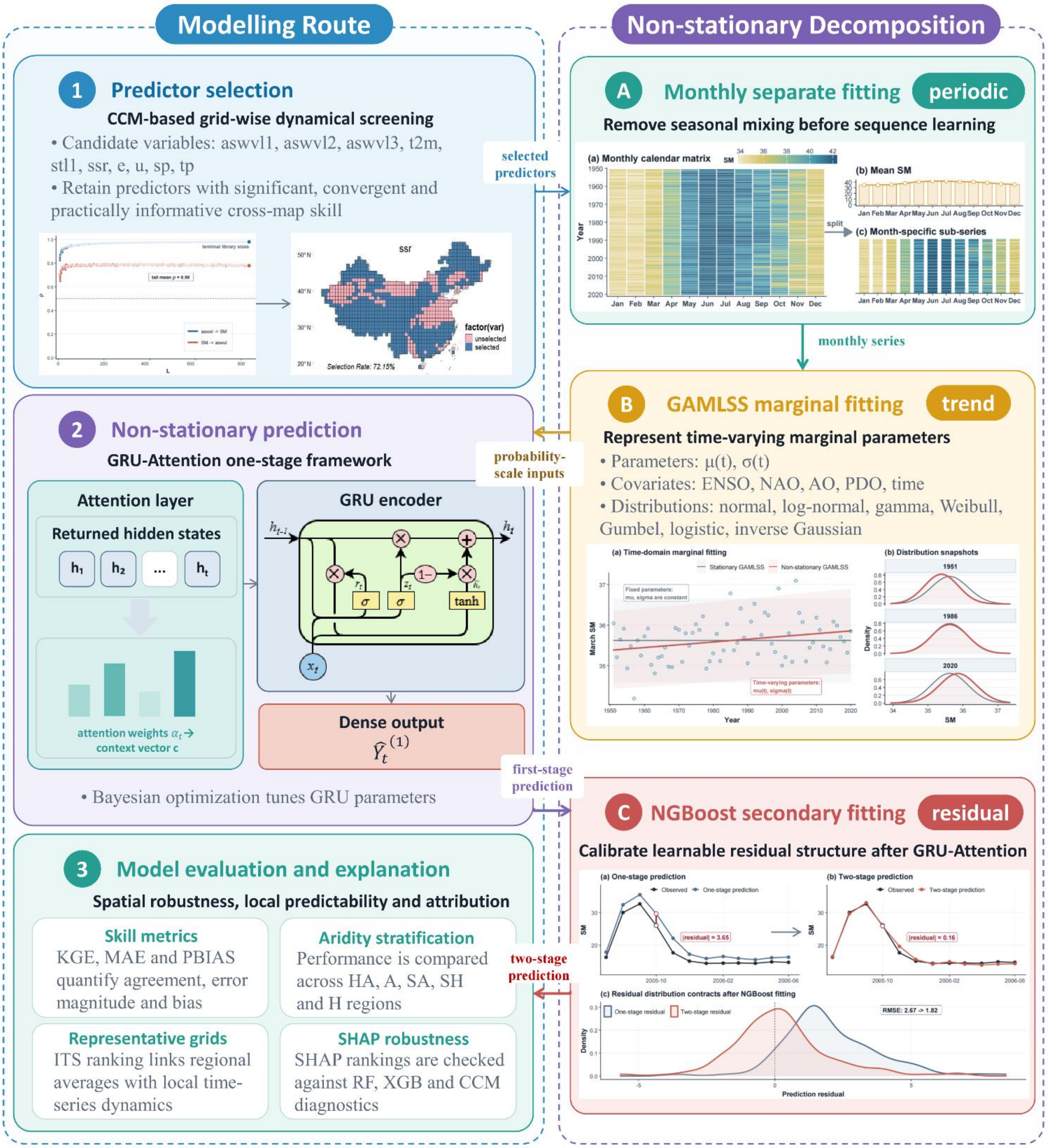

Figure 1: Framework of this study.

2.1 Causality-Guided Predictor Screening

Predictor screening was performed using convergent cross mapping (CCM), an empirical dynamic modelling method based on Takens' state-space reconstruction theorem (Sugihara et al., 2012; Takens, 1981). CCM tests whether the reconstructed manifold of one variable contains information about the state of another variable in a coupled nonlinear system. This property makes CCM more restrictive than conventional correlation screening, because cross mapping can, in principle, be examined in both directions and can therefore diagnose asymmetric information transfer and feedback. In this study, CCM was used as a prediction-oriented screening step before model training: a predictor was retained only when its reconstructed manifold provided significant and convergent information about the SM series at the same grid cell. Detailed theoretical treatments and hydrological applications are available in Sugihara et al. (2012), Bonotto et al. (2022) and Zhao et al. (2024).

For each grid cell, the target series was monthly SM, denoted as $Y(t)$. Candidate predictors included antecedent SM at one-, two- and three-month lags (aswvl1, aswvl2 and aswvl3), 2 m air temperature (t2m), soil temperature level 1 (stl1), surface net solar radiation (ssr), total evaporation (e), 10 m wind speed (u), surface pressure (sp), and total precipitation (tp). For each candidate predictor $X(t)$, the delay-coordinate shadow manifold was reconstructed as

$$M_X(t) = [X(t), X(t-\tau), \ldots, X(t-(E-1)\tau)\,] \tag{1}$$

where $M_X(t)$ is the reconstructed state vector, E is the embedding dimension, and τ is the time delay. The embedding dimension was selected objectively rather than fixed a priori. Candidate dimensions from E=2 to E=5 were evaluated using simplex projection, and the E that maximized prediction skill was adopted for the corresponding predictor manifold. This choice followed standard empirical dynamic modelling practice: the embedding should be large enough to unfold the system trajectory, but not so large that the attractor becomes sparse and noisy.

For a target time t, the E+1 nearest neighbors of $M_X(t)$ were identified in the reconstructed manifold. Their distances were transformed into normalized exponential weights:

$$w_i(t) = \frac{u_i}{\sum_{j=1}^{E+1} u_j} \tag{2}$$

$$u_i = exp\left(-\frac{d\left(M_X(t), M_X(t_i)\right)}{d\left(M_X(t), M_X(t_1)\right)}\right) \tag{3}$$

where t_i is the time index of the i-th nearest neighbor, $d\big(M_X(t), M_X(t_i)\big)$ is the distance between the target state and its i-th neighbor in the reconstructed manifold, and $d\big(M_X(t), M_X(t_1)\big)$ is the distance to the closest neighbor. The cross-mapped estimate of SM was then calculated as:

$$\widehat{Y(t)}|M_X = \sum w_i(t) Y(t_i) \tag{4}$$

Here, $Y(t_i)$ is the observed SM value at neighbor time index t_i. Cross-map skill for the predictor-to-SM direction was measured as the Pearson correlation between observed and reconstructed SM:

$$\rho_{X\to Y}(L) = \; cor\left(Y(t), \widehat{Y(t)}\;|M_X\right) \tag{5}$$

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

where L is the library size. The reverse quantity, $\rho_{Y\rightarrow X}(L)$, can be obtained analogously by reconstructing the SM manifold and estimating $X(t)$. In a dynamically coupled system, ρ should increase and then stabilize as L increases, because a larger library samples the attractor more completely. In the implementation, each candidate predictor was evaluated separately for each grid cell using the R package multispatialCCM. A predictor was retained, when the predictor-to-SM cross mapping satisfied three criteria. The CCM significance test indicated $p<0.05$; the terminal segment of the ρ-L curve was convergent, with $|\text{slope}| < 0.01$ and a standard deviation of the last three ρ values smaller than 0.05; and the mean ρ of the last three library sizes exceeded 0.5. These thresholds were used to retain variables that were statistically significant, dynamically convergent and practically informative. When CCM could not be estimated because of numerical instability, Pearson correlation was used only as a fallback to avoid dropping the grid from subsequent modelling. In this fallback pathway, predictors with $|r| > 0.6$ were retained, but this was treated as a pragmatic data-processing safeguard rather than evidence of nonlinear dynamical coupling.

2.2 Statistical Representation of Non-stationary Soil Moisture Marginals

Distributional non-stationarity in SM was represented before sequence modelling using generalized additive models for location, scale and shape (GAMLSS). Unlike conventional marginal fitting, which assumes fixed distributional parameters, GAMLSS allowed the parameters of a response distribution to vary with explanatory covariates through link functions (Rigby & Stasinopoulos, 2005). This framework was appropriate for SM, because changes in large-scale climate variability, evaporative demand and long-term background conditions may alter not only the central tendency of SM but also its dispersion. In this study, GAMLSS was used as a probability-scale transformation rather than as the final prediction model, with the aim of separating distributional shifts from the temporal dependence subsequently learned by the DL model.

For each grid cell and calendar month *m*, the monthly SM observation in year *t*, denoted by $Y_{m,t}$, was assumed to follow a parametric marginal distribution:

$$Y_{m,t} \sim D(\theta_{m,t}) \tag{6}$$

where $D(\cdot)$ is a candidate distribution family, and $\theta_{m,t}$ is its parameter vector. In the general GAMLSS framework, it may include location, scale, and shape parameters, commonly written as:

$$\theta_{m,t} = (\mu_{m,t}, \sigma_{m,t}, \upsilon_{m,t}, \tau_{m,t}) \tag{7}$$

where μ represents the location parameter, σ represents the scale parameter, and ν and τ, when present, describe higher-order distributional properties, such as skewness and tail behavior (Rigby & Stasinopoulos, 2005; Stasinopoulos et al., 2017). In this study, seven candidate marginal families were evaluated: normal, log-normal, gamma, Weibull, Gumbel, logistic, and inverse Gaussian distributions. These families provide flexible support for symmetric, skewed and positive-valued marginal behaviors. For the candidate families used here, the fitted parameter vector was represented by the location and scale parameters, μ and σ, which were either held constant or allowed to vary with external covariates.

The distinction between stationary and non-stationary marginal models was defined by whether the distributional parameters were time-invariant. In the stationary case, the parameter vector was fixed, so that all observations from the same calendar month were assumed to arise from the same marginal distribution. In contrast, the non-stationary formulation allowed one or more elements of $\theta_{m,t}$ to change with climate covariates or a temporal trend. GAMLSS extended conventional distribution fitting by linking each distributional parameter to covariates through a monotonic link function, thereby ensuring that the fitted parameter remained within its admissible range. For a generic parameter θ_k, the GAMLSS predictor can be written as:

$$g_k(\theta_{k,m,t}) = \beta_{0,k} + \beta_{1,k} \cdot cov_t \quad (8)$$

where $g_k(\cdot)$ is the link function for parameter k, and $\beta_{0,k}$ and $\beta_{1,k}$ are the intercept and coefficient of the corresponding linear predictor, respectively. cov_t is one non-stationary candidate driver. The candidate covariates were the four monthly large-scale climate indices, ENSO, NAO, AO and PDO, together with a linear time index. In this study, only linear covariate effects were considered. Each covariate was tested separately rather than in a multivariable marginal model, which reduced overparameterization given the limited sample size within each calendar-month series.

For each candidate distribution and covariate, three non-stationary parameter structures as well as the stationary one were fitted. For example, the first allowed only the location parameter to vary:

$$g_\mu(\mu_{m,t}) = \beta_{0,\mu} + \beta_{1,\mu} \cdot cov_t, \;\; g_\sigma(\sigma_{m,t}) = \beta_{0,\sigma} \quad (9)$$

The second allowed only the scale parameter to vary, and the third allowed both location and scale to vary with the same covariate. All candidate marginal models were estimated by maximum likelihood using the GAMLSS framework, and model selection was based on both goodness of fit and parsimony. First, the fitted cumulative distribution function was compared with the empirical distribution using the Kolmogorov-Smirnov test. Candidate models with $p<0.05$ were discarded, because they failed to provide an acceptable marginal fit at the 5% significance level. Among the retained models, the final marginal distribution was selected by prioritizing a larger *p*-value and then a lower corrected Akaike information criterion (Akaike, 1974; Hurvich & Tsai, 1989).

After the optimal marginal model had been selected, each monthly SM observation was transformed to the probability scale using the fitted cumulative distribution function (CDF):

$$U_{m,t} = F_D(Y_{m,t}; \theta_{m,t}), \quad 0 < U_{m,t} < 1 \quad (10)$$

where $F_D(\cdot)$ is the CDF of the selected distribution D. After prediction, the estimated probability-scale values were transformed back to SM units through the inverse CDF.

2.3 Deep Learning-Based Two-Stage Prediction

The predictive component was implemented as a two-stage learning framework. The first stage used a gated recurrent unit with an attention layer (GRU-Attention) to learn nonlinear relationships between the selected predictors and SM. The second stage used natural-gradient boosting (NGBoost) to calibrate the mapping between the one-stage prediction and the

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

observed SM. For calendar-month-specific configurations, each calendar month was modelled independently, with the first 50 years used for model fitting and the remaining 20 years used for evaluation. For full-series configurations, the first 600 monthly observations were used for training and the remaining 240 observations were used for testing. All predictors and targets were scaled using min-max normalization before GRU training, and predictions were transformed back to the original SM scale before second-stage modelling.

The GRU is a gated recurrent architecture developed as a simplified variant of long short-term memory (LSTM). It uses fewer gates than LSTM by combining the memory cell and hidden state, which reduces the number of trainable parameters, while retaining the ability to regulate information flow through time (Baumberger et al., 2025; Chung et al., 2014). This property is useful for monthly SM modelling, because each calendar-month model has a limited number of annual samples. Given an input vector x_t and previous hidden state h_{t-1}, the GRU transition was expressed as (Ahmed et al., 2021; J. Yu et al., 2021; Zheng et al., 2024):

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{11}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{12}$$

$$\widetilde{h_t} = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \tag{13}$$

$$h_t = (1 - z_t) \odot \widetilde{h_t} + z_t \odot h_{t-1} \tag{14}$$

Here, z_t and r_t are the update and reset gates, respectively; $\widetilde{h_t}$ is the candidate hidden state; $\tanh(\cdot)$ is the hyperbolic tangent activation; $\sigma(\cdot)$ is the logistic sigmoid function; $\odot$ denotes element-wise multiplication; and W, U, and b are trainable weight matrices and bias vectors. The reset gate controls the contribution of previous hidden information when constructing the candidate state, whereas the update gate controls the degree to which the previous state is retained.

An attention operation was then applied to the returned GRU representation before the final dense layer. Attention mechanisms were originally introduced to allow sequence models to assign unequal weights to different hidden representations, thereby reducing reliance on a single compressed state vector (Bahdanau et al., 2016; Q. Li et al., 2022). In a standard additive attention formulation, the score and normalized weight for the hidden state hidden state h_t can be written as:

$$e_t = \mathrm{v}_a{}^T \tanh(W_a h_t + b_a) \tag{15}$$

$$\alpha_t = \frac{exp(e_t)}{\sum_k exp(e_k)}, \quad \sum_t \alpha_t = 1 \tag{16}$$

and the context vector is:

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{17}$$

In these equations, e_t is the attention score; α_t is the normalized attention weight; T is the sequence length; v_a, W_a and b_a are trainable attention parameters; and c is the attention-weighted context vector. The one-stage prediction is obtained from a dense output layer:

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

$$\widehat{Y}_t^{(1)} = W_o \mathrm{c} + b_o \tag{18}$$

This formulation allows the network to emphasize the most informative hidden representation for SM prediction. In the implementation, the selected predictor vector was reshaped into the three-dimensional input format required by the Keras GRU layer, and the GRU output was passed through a softmax-based attention layer before the final dense layer. Hyperparameters were selected using Bayesian optimization, which is commonly used. when model evaluation is computationally expensive because it learns a surrogate function of the objective and chooses new candidates by balancing exploration and exploitation (Snoek et al., 2012). In the paper, the search ranges were 10~80 GRU units, 10^{-4}~10^{-2} for the learning rate, 16~64 for the batch size, and 0~0.3 for dropout rate. Candidate networks were trained for a fixed maximum of 50 epochs with early stopping based on the validation loss. The final GRU-Attention model was then refitted using the best hyperparameter combination that minimized the mean squared error.

The second stage used NGBoost to calibrate the relationship between one-stage predictions and observed SM. NGBoost is a probabilistic boosting method that estimates a full conditional distribution rather than only a point estimate. It optimizes distributional parameters by minimizing a proper scoring rule and uses natural-gradient updates to account for the geometry of the parameter (Duan et al., 2020). In this study, the one-stage training predictions were first transformed back to the original SM scale and then used as the input to the NGBoost model. Let $P_\theta\left(Y_t \middle| \widehat{Y}_t^{(1)}\right)$ denote the conditional distribution of SM given the one-stage prediction. NGBoost estimates the distributional parameters by solving the following equation:

$$\theta^* = \underset{\theta}{argmin} \sum_{t=1}^{n} S\left(P_\theta\left(\cdot \middle| \widehat{Y}_t^{(1)}\right), Y_t\right) \tag{19}$$

where $S(\cdot)$ is a proper scoring rule and n is the number of training samples. In natural-gradient boosting, the parameter update can be written conceptually as:

$$\theta^{(m+1)} = \theta^{(m)} - \rho_m I^{-1}\left(\theta^{(m)}\right) \nabla_\theta S\left[P_{\theta^{(m)}}\left(\cdot \middle| \widehat{Y}^{(1)}, Y\right)\right] \tag{20}$$

where ρ_m is the learning rate at boosting iteration m, $\nabla_\theta S$ is the score gradient and $I(\theta)$ is the Fisher information matrix. The final two-stage point prediction was taken as the expectation of the learned conditional distribution:

$$\widehat{Y}_t^{(2)} = E\left[Y_t \middle| P_{\theta^*}\left(\cdot \middle| \widehat{Y}_t^{(1)}\right)\right] \tag{21}$$

In this framework, the GRU-Attention model provided the primary nonlinear prediction from selected covariates, whereas NGBoost tested whether the remaining relationship between the first-stage prediction and the observed SM contained calibratable distributional structure.

2.4 Representative Grid Selection for Regional Evaluation

Representative grids were selected to link the aridity-zone behavior with local time-series dynamics. Following the temporal-stability framework commonly used in SM monitoring, the departure of grid i from the regional mean at time t was quantified using the relative difference (Brocca et al., 2009; Jacobs, 2004; Vachaud et al., 1985):

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

$$RD_{i,t} = \frac{SM_{i,t} - \overline{SM_t}}{\overline{SM_t}} \tag{22}$$

where $SM_{i,t}$ is the SM value at grid i, and $\overline{SM_t}$ is the mean SM over all grids within the same aridity region at time t. The mean relative difference (MRD) and the standard deviation of relative difference (SDRD) were then computed over the total data length T:

$$MRD_i = \frac{1}{T}\sum_{t=1}^{T} RD_{i,t} \tag{23}$$

$$SDRD_i = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(RD_{i,t} - MRD_i\right)} \tag{24}$$

Here, MRD_i measures whether a grid is persistently wetter or drier than the regional mean, whereas $SDRD_i$ measures the temporal variability of that departure. A grid with both low absolute MRD and low SDRD is therefore more representative of the regional SM dynamics. However, these two criteria can diverge. For example, a grid may have MRD_i close to zero but still exhibit large temporal fluctuations around the regional mean. To avoid selecting grids that are unbiased on average but temporally unstable, the two quantities were combined into an index of temporal stability (ITS):

$$ITS_i = \sqrt{(MRD_i\,)^2 + (SDRD_i)^2} \tag{25}$$

Lower ITS values indicate stronger temporal stability and greater representativeness of the regional SM regime. Within each aridity region, all grids were ranked by ITS and visualized together with MRD and SDRD. The representative grid was selected from the lowest-ITS candidates after confirming that its time series retained physically plausible local variability. These representative grids were then used to evaluate local prediction dynamics and to compare regional-average skill with grid-level predictability.

2.5 SHapley Additive exPlanations (SHAP)-Based Model Interpretation

Model interpretation was based on SHapley Additive exPlanations (SHAP), a game-theoretic attribution method that assigns each predictor a local contribution to an individual model prediction (Lundberg & Lee, 2017; Shapley, 1953). SHAP is well suited to tree-based hydrological prediction models because it provides additive, observation-level explanations while preserving consistency in feature attribution (Lundberg et al., 2020). In this study, SHAP was applied to XGBoost models fitted for the full study domain and separately for each aridity region, using the same candidate predictors.

For a fitted model f and predictor vector x, the SHAP representation decomposes the model output into a baseline prediction and the sum of feature-specific contributions:

$$f(x) = \phi_0 + \sum_{j=1}^{p} \phi_j \tag{26}$$

© Author(s) 2026. CC BY 4.0 License.

where ϕ_0 is the expected model output, ϕ_j is the SHAP contribution of predictor j, and p is the number of predictors. A positive ϕ_j indicates that predictor j increases the predicted SM relative to the baseline, whereas a negative ϕ_j indicates a decreasing effect. The Shapley value for predictor j is defined as:

$$\phi_j = \sum\nolimits_{S\subseteq F\setminus\{j\}} \frac{|S|!\,(p-|S|-1)!}{p!}\left[f_{S\cup\{j\}}\left(x_{S\cup\{j\}}\right) - f_S(x_S)\right] \tag{27}$$

where F is the full predictor set, S is a subset of predictors excluding j, and the bracketed term represents the marginal contribution of adding predictor j to subset S. This formulation evaluates a predictor by averaging its contribution over all possible predictor coalitions, reducing the dependence of attribution on an arbitrary variable ordering.

Global driver importance was quantified using the mean absolute SHAP value:

$$I_j = \frac{1}{n}\sum_{i=1}^{n}\left|\phi_{i,j}\right| \tag{28}$$

where n is the number of samples and $\phi_{i,j}$ is the SHAP value of predictor j for sample i. Predictors with larger I_j were interpreted as having stronger overall influence on the modelled SM variability.

Because SHAP explains the behavior of a fitted model rather than proving physical causality, the SHAP-derived rankings were interpreted as model-based attribution diagnostics. To evaluate whether the identified dominant drivers were robust to the attribution method, SHAP rankings were compared with random forest (RF) feature importance, XGBoost gain-based importance, and CCM selection frequency. These comparisons were used as robustness diagnostics rather than as independent causal proof.

2.6 Model Configurations and Evaluation Methods

Six experimental configurations were designed to isolate the effects of seasonal separation, large-scale climate indices and statistical marginal representation on SM prediction. M1 and M2 were fitted using the full monthly time series, whereas M3-M6 used calendar-month-specific modelling. M3 used non-stationary GAMLSS marginal transformation, M4 used the corresponding stationary marginal transformation, and M5-M6 removed the marginal transformation to test the contribution of monthly separation alone. The configurations are summarized in Table 1.

Table 2: Six experimental model configurations used in this study.

Model	Periodic separation	Climate covariates	Statistical marginal representation	Purpose
M1	No	No	No	Baseline using selected predictors
M2	No	Yes	No	Tests added value of climate covariates without monthly separation
M3	Yes	Yes	Non-stationary	Tests non-stationary marginal representation under monthly modelling

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

M4	Yes	No	Stationary	Tests stationary marginal representation under monthly modelling
M5	Yes	Yes	No	Tests monthly separation without marginal transformation
M6	Yes	No	No	Tests monthly separation plus climate covariates without marginal transformation

Prediction performance was evaluated using three complementary deterministic metrics. Kling-Gupta efficiency (KGE) was used as the primary metric. because it jointly evaluates linear association, variability and bias, which makes it more informative than correlation or error magnitude alone for hydrological model assessment (Gupta et al., 2009):

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{29}$$

$$\alpha = \frac{\sigma(\widehat{SM})}{\sigma(SM)}, \beta = \frac{mean(\widehat{SM})}{mean(SM)} \tag{30}$$

where SM and $\widehat{SM}$ denote the observed and predicted soil moisture series, respectively; r is the Pearson correlation coefficient, α is the variability ratio, and β is the bias ratio. KGE has an optimal value of 1, and higher values indicate better agreement between prediction and observation.

Mean absolute error (MAE) was used to quantify the average magnitude of prediction errors in the original SM units. Unlike squared-error metrics, MAE is less dominated by a small number of large errors and is therefore useful for evaluating typical prediction deviations (Willmott & Matsuura, 2005):

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|SM_t - \widehat{SM}_t\right| \tag{31}$$

where SM_t and $\widehat{SM}_t$ are the observed and predicted SM values at time t. Lower MAE indicates smaller average prediction error.

Percent bias (PBIAS) was used to assess systematic overestimation or underestimation, a diagnostic commonly used in hydrological model evaluation:

$$PBIAS = \frac{100 \times \sum_{t=1}^{T}\left(\widehat{SM}_t - SM_t\right)}{\sum_{t=1}^{T}(SM_t)} \tag{32}$$

Under this sign convention, positive PBIAS indicates overestimation, negative PBIAS indicates underestimation and values closer to zero indicate smaller systematic bias.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

3 Study Area and Data

3.1 Study Area

The study area covers mainland China (10-54°N, 72-136°E), encompassing strong gradients in elevation, climate and water-energy conditions. Topography descends broadly from the Tibetan Plateau and western mountain systems to the eastern lowlands and coastal plains (Fig. 2(a)), creating spatial contrasts in atmospheric forcing, surface energy balance and soil thermal regimes. Summer monsoon rainfall supplies much of the annual water input in eastern and southern China, whereas northwestern China is dominated by limited precipitation, high evaporative demand and persistent water limitation (Piao et al., 2010). Northern and northeastern China experience pronounced seasonal temperature variability and soil thermal contrasts, while southwestern and plateau regions are shaped by elevation-dependent radiation, temperature and precipitation gradients. These interacting controls generate strong spatial heterogeneity and temporal non-stationarity in SM dynamics.

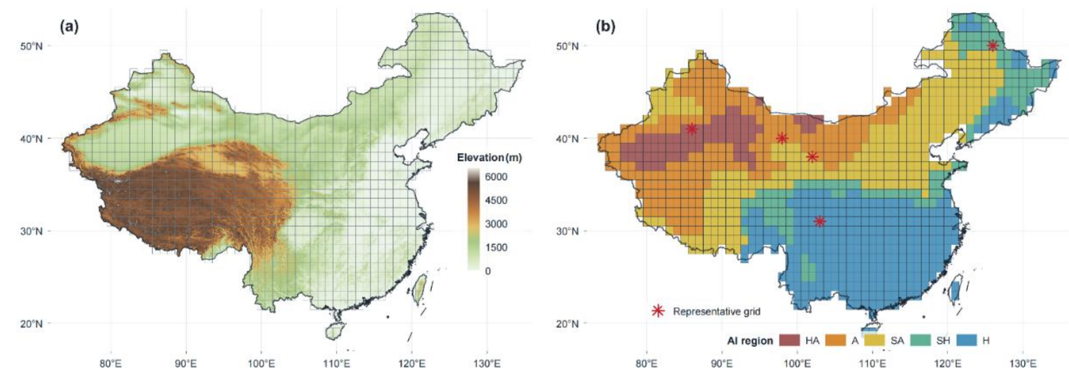

Figure 2: Hydroclimatic and topographic stratification of the study area. (a) Digital elevation model of China. (b) Aridity-index-based hydroclimatic regions. Red stars indicate the representative grid cells selected for region-specific evaluation.

The domain was further stratified by the aridity index (AI), defined as the ratio of long-term precipitation to potential evapotranspiration. Following the UNEP aridity classification (UNEP, 1992), five hydroclimatic zones were considered: Hyper-Arid (HA, AI < 0.03), Arid (A, 0.03 ≤ AI < 0.20), Semi-Arid (SA, 0.20 ≤ AI < 0.50), Dry Sub-Humid (SH, 0.50 ≤ AI < 0.65) and Humid (H, AI ≥ 0.65). These zones represent a continuous dry-to-wet transition from strongly water-limited landscapes to humid monsoon regions (Fig. 2(b)). This stratification provides the spatial basis for evaluating whether prediction skill, representative-grid behavior, and dominant SM controls vary systematically across hydroclimatic regimes (Seneviratne et al., 2010).

3.2 Data

The primary data source was ERA5-Land data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5-Land is produced by replaying the land component of ERA5 and provides a physically consistent representation of terrestrial water and energy states over multiple decades (Muñoz-Sabater et al., 2021). This study used monthly data from January 1950 to December 2020 on the native 0.1° × 0.1° grid. Surface SM was represented by volumetric soil water layer 1 (swvl1), corresponding to the 0-7 cm surface soil layer. Candidate hydro-meteorological and land-surface predictors included 2 m temperature (t2m), soil temperature level 1 (stl1), surface net solar radiation (ssr), total evaporation (e), 10 m u-component of wind (u_10), 10 m v-component of wind (v_10), surface pressure (sp), total precipitation (tp), soil type (slt) and type of high vegetation (tvh). The two wind components were combined into scalar 10m wind speed (u). All variables were resampled to a 1° modelling grid. Missing cells and non-land cells outside the China mask were excluded, yielding 966 valid grid cells.

Large-scale climate variability was represented by four monthly climate indices: Niño 3.4 for the El Niño-Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), the Arctic Oscillation (AO), and the Pacific Decadal Oscillation (PDO). These indices were included as covariates because remote ocean-atmosphere modes can modulate regional temperature, precipitation and circulation anomalies, thereby influencing SM through both water-supply and energy-demand pathways (Seneviratne et al., 2010). All four monthly indices were obtained from the NOAA Physical Sciences Laboratory and aligned with the ERA5-Land monthly time axis.

4 Results

4.1 Effects of Periodic Separation, Trend Modeling, and Residual Correction on Soil Moisture Prediction

4.1.1 Separating Seasonal Periodicity Provides the Most Robust Skill Gain

The ablation comparison first confirmed that separating the seasonal periodic component was the dominant source of improvement in one-stage SM prediction (Fig. 3). Compared with the full-series configurations (M1 and M2), the calendar-month-specific configurations (M3~M6) generally shifted the KGE distributions toward higher values, indicating that removing seasonal mixing before sequence learning improved the representation of recurrent SM dynamics. Differences among the four periodic-separated models further clarified the advantage of integrating statistical marginal modeling with DL. Although M5 achieved the highest average monthly KGE among the four models, M3 and M4 showed more balanced error behavior across months. In particular, M4 achieved the lowest monthly mean MAE among all models (1.250), followed by M3 (1.318), whereas the purely deep-learning models M5 and M6 showed larger mean MAE values of 1.362 and 1.426, respectively. The contrast was more evident for systematic bias. M3 and M4 remained closer to zero in most months, with monthly mean PBIAS values of -0.228 and -0.033, respectively, while M5 and M6 showed stronger positive bias, both exceeding 0.80 on average.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

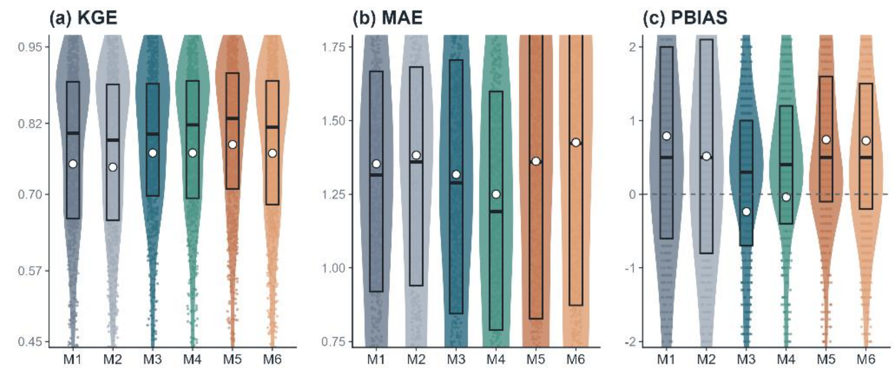

Figure 3: Ablation performance of periodic-component separation in one-stage SM prediction. Points represent grid-scale results, boxes denote the interquartile range and median, and white circles indicate model means.

Fig. 4 further evaluated the monthly behavior of the four periodic-separated models (M3~M6). KGE was the highest in winter, with the cross-model monthly mean reaching 0.651 in January and remaining relatively high in December to February, whereas the lowest skill occurred in September (0.527). When grouped by season, mean KGE was the highest in winter (0.634) and the lowest in autumn (0.538), indicating that SM predictability after periodic separation still varied strongly within the annual cycle. This seasonal contrast was accompanied by a systematic change in absolute error: MAE was the lowest in February (1.083), increased through summer, and reached its maximum in October (1.618), with the seasonal mean rising from approximately 1.21 in winter-spring to 1.58 in autumn. The monthly PBIAS pattern provided additional information on the type of remaining prediction error. Positive bias was most evident in March-April and November-December, whereas negative bias was concentrated in May-June, with the cross-model mean PBIAS reaching 1.944 in April and -1.985 in June. Across months, M5 achieved the highest average KGE (0.594), but M3 and M4 generally maintained lower MAE and smaller absolute bias, especially for the transition months when the sign and magnitude of PBIAS changed rapidly. Thus, monthly results showed that periodic separation improved the representation of recurrent seasonal behavior, but the residual prediction difficulty was concentrated in transitional and late-growing-season months, when SM dynamics were more strongly affected by changing water-energy conditions.

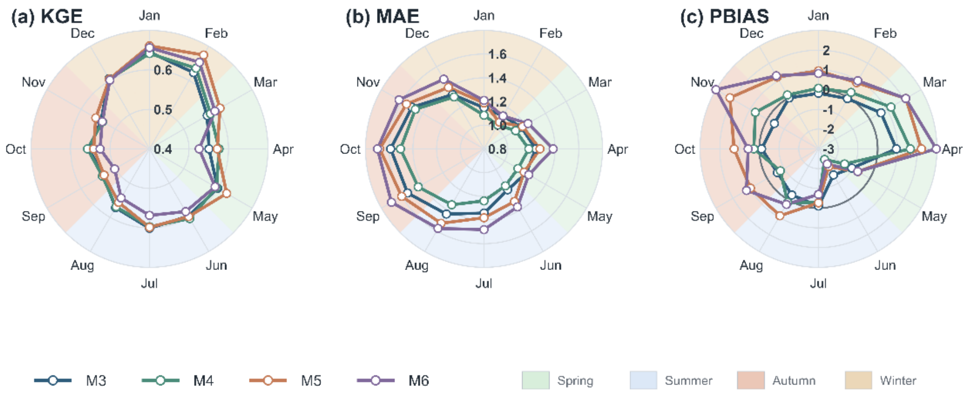

Figure 4: Monthly prediction skill after periodic-component separation.

4.1.2 Trend Non-Stationarity Provides Conditional Rather Than Universal Predictive Gain

We first used the Mann-Kendall (M-K) test to diagnose trend-related non-stationarity in the SM series across months and grid cells (Fig. 5). Significant trends were widespread, but they did not dominate the domain in any month (Fig. 5(a)): the combined proportion of significant increasing and decreasing trends ranged from 30.1% to 44.2%. Significant decreases were generally more frequent than significant increases, whereas non-significant increasing and decreasing tendencies together occupied a large fraction of the grids. The annual map showed that these trend classes were also spatially organized (Fig. 5(b)). Significant decreases were concentrated across large parts of northern, northeastern and eastern China, while significant increases appeared more locally in western and central regions. Monthly M-K maps in the Supplements showed that this broad spatial contrast persisted through the annual cycle, although the areal extent of each trend class varied by month. Thus, trend non-stationarity was a recurrent feature of the SM record, but it was expressed unevenly across both space and seasons.

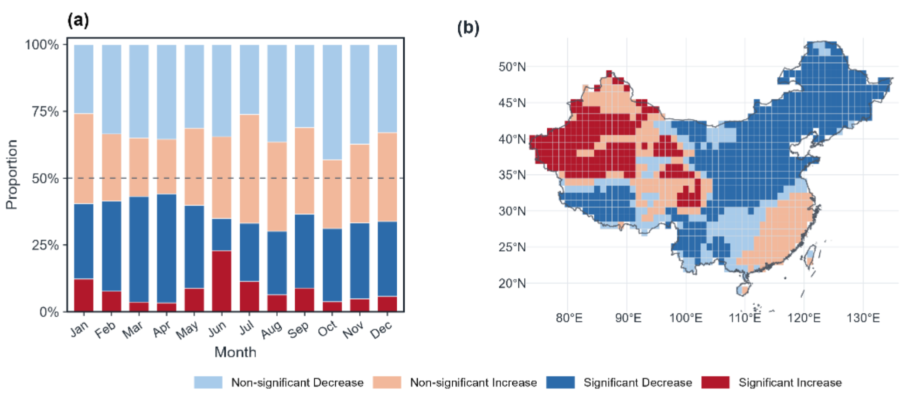

Figure 5: Spatiotemporal diagnostics of trend-related soil moisture non-stationarity. (a) Monthly proportions of four M-K trend classes. (b) Annual spatial distribution of the same trend classes.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

We next asked whether the diagnosed trend non-stationarity corresponded to a consistent gain from non-stationary marginal fitting (Fig. 6). Here, ΔKGE was defined as KGE(M3)-KGE(M4), so positive values denoted higher skill for the non-stationary-margin model and negative values denoted higher skill for the stationary-margin counterpart. The annual ΔKGE map showed a fine-grained mixture of positive and negative values rather than a spatial pattern that followed a single trend class (Fig. 6(a)). Consistent with this spatial mosaic, the grid-wise comparison of M3 and M4 showed strongly aligned annual KGE values (R^2 = 0.89), with most points close to the 1:1 line (Fig. 6(b)). This indicated that the two marginal representations produced broadly comparable annual prediction skill, rather than a domain-wide improvement from the non-stationary formulation.

The trend-class distributions further constrained this result (Fig. 6(c)). M3 and M4 had strongly overlapping KGE distributions in all four M-K categories, including both significant decreasing and significant increasing grids. The median ΔKGE values were similar in the two non-significant trend classes and significant increasing class, but slightly negative in the significant decreasing class. The monthly three-panel comparisons in the Supplements were consistent with the annual analysis: monthly R^2 between M3 and M4 ranged from 0.38 to 0.66, monthly median ΔKGE remained close to zero (-0.0106 to 0.0095). Together, the annual and monthly evidence showed that trend-related non-stationarity was a useful diagnostic property of SM dynamics, but that its detection alone did not imply additional predictive information from non-stationary marginal fitting.

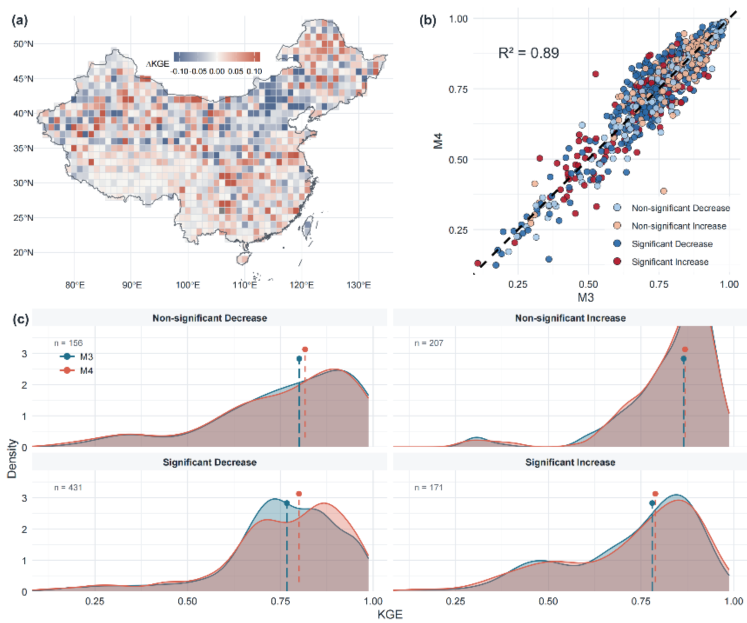

Figure 6: Trend-related non-stationarity and predictive gain from non-stationary marginal fitting. (a) Annual spatial distribution of ΔKGE. (b) Annual grid-wise comparison of KGE between M3 and M4. (c) KGE distributions of M3 and M4 within four M-K trend classes. ΔKGE is defined as KGE(M3)-KGE(M4), with positive values indicating higher skill for the non-stationary-margin model.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

4.1.3 Residual Learning Helps Only When Residuals Are Learnable

For M1~M6, residuals were modelled with NGBoost to construct two-stage predictions, and XGBoost and Random Forest (RF) were included as additional residual-learning baselines (Fig. 7(a)). Across the three residual learners, XGBoost achieved the highest skill for M1, M2, M5 and M6, with mean KGE values of 0.799, 0.787, 0.815, and 0.808, respectively. By contrast, the GRU-based learner achieved the highest performance for the statistically informed configurations M3 and M4 (mean KGE = 0.794 and 0.805). RF consistently produced lower two-stage KGE across all six configurations (0.712~0.755). The paired one-stage and two-stage GRU results further showed that the gain from residual learning was strongly model dependent (Fig. 7(b)). M1~M4 all had positive median gains (ΔKGE = +0.022, +0.025, +0.012 and +0.020, respectively), and these gains were spatially widespread, improving 77%, 76%, 63%, and 71% of grid cells. In contrast, the purely DL-driven baselines, M5 and M6, had slightly negative median gains (-0.006 and -0.005) and improved in fewer than half of the grid cells (42% and 45%). Thus, the second stage improved performance mainly when the first-stage model left residual errors that remained learnable.

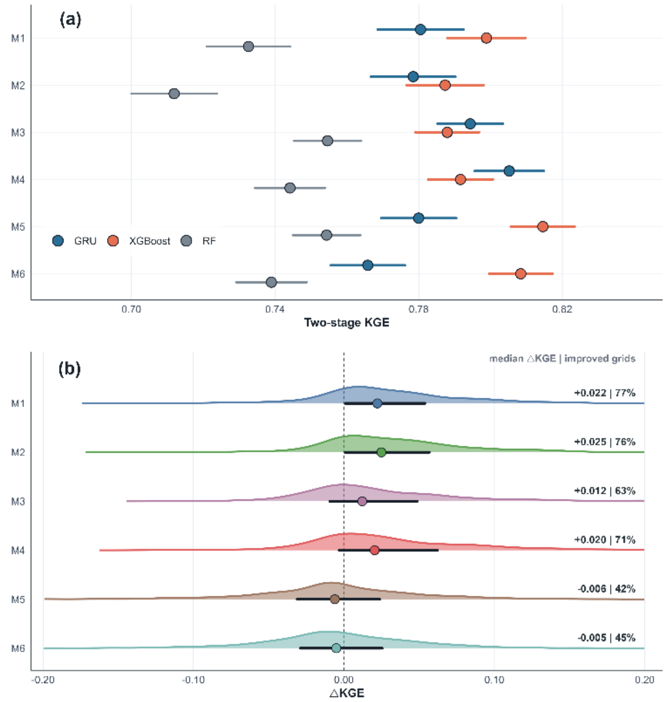

Figure 7: Residual-learner performance and GRU-based two-stage gain. (a) Mean two-stage KGE and 95% confidence intervals for GRU, XGBoost and Random Forest (RF) across the six model configurations. (b) Grid-cell ΔKGE distributions for GRU.

The gain-structure analysis showed that the magnitude of two-stage improvement depended on one-stage model skill (Fig. 8). For M3 and M4, the relationship between one-stage KGE and percentage KGE improvement was strongly negative and statistically significant, indicating that the largest relative improvements occurred in grid cells where the initial

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

statistical-DL prediction was weaker. The 20% KGE-bin summaries confirmed this pattern: mean improvement was high in low-skill bins and declined towards zero as one-stage KGE approached 0.8~1.0. M1 and M2 showed weaker but still significant negative relationships, with positive average gains across much of the middle KGE range. By contrast, M5 and M6 showed no meaningful monotonic association between one-stage KGE and improvement, and their binned mean gains fluctuated around zero. These results indicated that the largest and most spatially consistent improvements occurred when the first-stage model retained correctable residual structure, whereas the purely DL-driven baselines gained little from the additional residual-learning step. Residual learning was therefore beneficial only when the remaining errors were structured enough to be learned, rather than serving as an automatically effective second-stage correction.

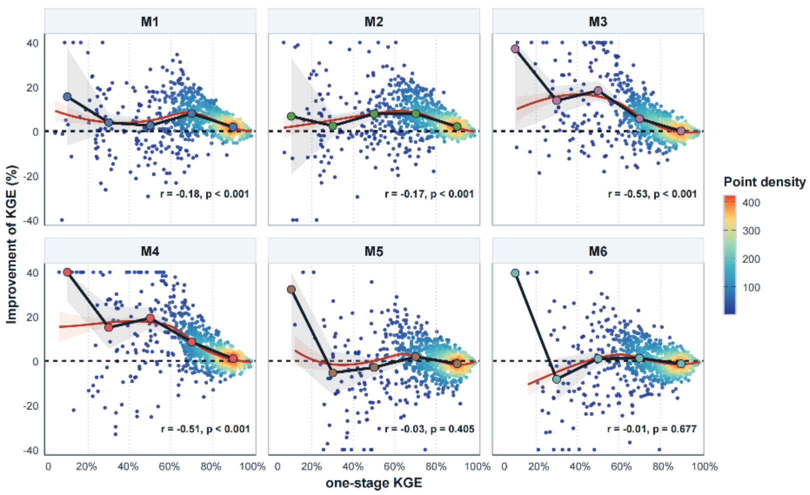

Figure 8: Dependence of percentage KGE improvement on one-stage KGE. Each panel shows grid-cell point density, a LOESS trend and mean improvement within 20% one-stage KGE bins.

4.2 Spatial Robustness across Different Aridity Zones and Representative Grids

4.2.1 Aridity-zone Dependence of Prediction Skill

To evaluate whether two-stage prediction skill was spatially transferable across contrasting hydroclimatic settings, the study domain was stratified into five aridity-index (AI) regions (UNEP, 1992): Hyper-Arid (HA), Arid (A), Semi-Arid (SA), Dry Sub-Humid (SH), and Humid (H). The resulting performance distributions showed a clear aridity-zone dependence rather than a spatially uniform skill pattern (Fig. 9). Both M3_two and M4_two achieved their highest median KGE in the A and SA regions, with median values of 0.86 and 0.89 in A and 0.82 and 0.85 in SA, respectively (Fig. 9(a)). The HA region had lower median skill (0.73 for both models) and a broader lower tail, indicating less stable prediction under extremely dry conditions. The SH region also showed moderate skill (0.77 and 0.78), whereas the H region retained high median performance (0.83 and 0.84) but contained a visible group of low-skill grids. This non-monotonic response indicated that aridity did not act as a simple linear control on prediction skill (Fig. 9(b)-(c)): for both M3_two and M4_two, KGE increased from the hyper-arid boundary toward the arid to semi-arid range, decreased across part of the sub-humid to humid transition,

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

and then rose again at the wettest end of the sampled AI range. Overall, M4_two was generally slightly higher than M3_two, particularly in the A and SA regions, but the two models shared the same spatial organization of prediction skill.

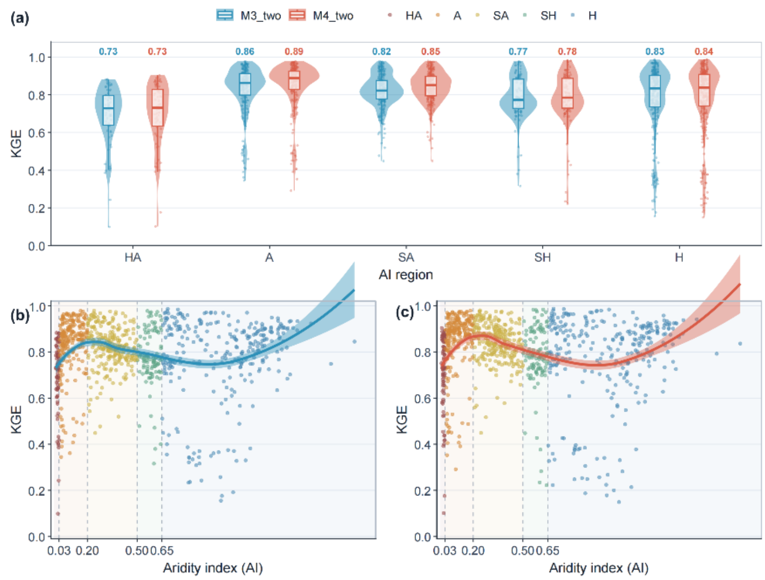

Figure 9: Aridity-zone dependence of two-stage prediction skill. (a) KGE distributions of M3_two and M4_two across five AI regions. Numbers above the violins denote regional median KGE. (b, c) Continuous relationships between aridity index and KGE for M3_two and M4_two, respectively. Dashed vertical lines indicate the AI thresholds used to define the five regions.

Seasonal aggregation further showed that the AI dependence of skill was not confined to the annual summary. Across both models, winter (DJF) consistently produced the highest mean KGE in every AI region, whereas autumn (SON) generally had the lowest skill (Fig. 10). For M3_two, winter mean KGE reached 0.74 in A and 0.76 in SA, compared with 0.61 in HA and 0.59 in H. For M4_two, the corresponding winter values increased to 0.78 in A and 0.79 in SA, while remaining lower in HA (0.63) and H (0.59). The lowest seasonal values occurred in HA in autumn and summer, with mean KGE values of 0.48 and 0.50 for M3_two and 0.49 and 0.52 for M4_two. The monthly heat map in Fig. S14 provided a higher-resolution robustness check for this seasonal pattern. It showed that the winter advantage was primarily associated with high skill in December to February, especially in the A and SA regions, whereas weaker skill in HA and parts of the humid region persisted across multiple months rather than being caused by a single anomalous month. Overall, these seasonal and monthly patterns confirmed that the spatial robustness of the two-stage models was structured by a persistent aridity-dependent predictability gradient, with the most stable skill concentrated in the arid and semi-arid zones rather than at either hydroclimatic extreme.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

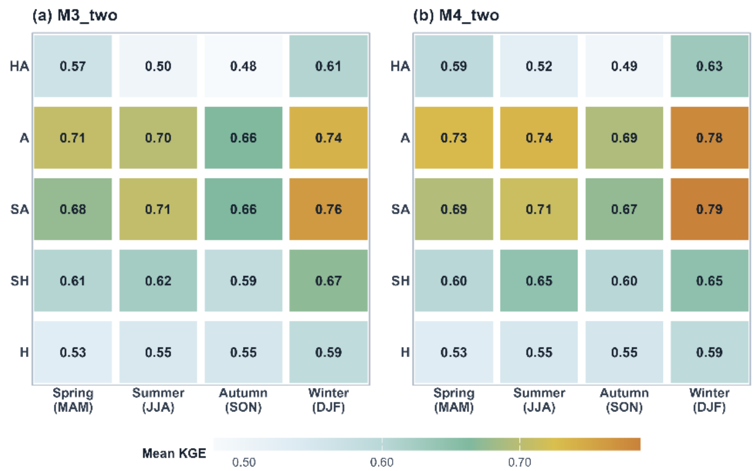

Figure 10: Seasonal mean KGE across AI regions for (a) M3_two and (b) M4_two.

4.2.2 Representative-grid Evaluation and Local Predictability

A representative-grid analysis was used to test whether a grid that best represented the regional SM state also exhibited representative prediction skill. Within each AI region, grids were ranked using the index of temporal stability (ITS), together with the mean relative difference (MRD) and the standard deviation of relative difference (SDRD), and the top-ranked candidates were displayed for comparison (Fig. 11(a)). The ranked ITS profiles differed substantially among the five AI regions, confirming that temporal stability itself was region dependent. For the representative grids, M3_two and M4_two reproduced the observed SM fluctuations with strongly region-dependent accuracy (Fig. 11(b)-(f)). The highest local skill occurred in the humid representative grid, where KGE reached 0.917 for M3_two and 0.914 for M4_two, and both simulations closely tracked the pronounced seasonal cycle. The SA representative grid also showed high skill, with KGE values of 0.804 and 0.857, indicating that both models captured the dominant temporal variability in this intermediate dryland setting. By contrast, the HA grid showed only moderate skill (0.534 and 0.525), and the A representative grid had the weakest local performance, with KGE equal to 0.407 for both models. This was notable because the A region had the highest regional KGE distribution in Fig. 9. The mismatch demonstrates that a grid selected for regional temporal stability may not represent the average predictability of that region, especially where local SM dynamics are heterogeneous.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

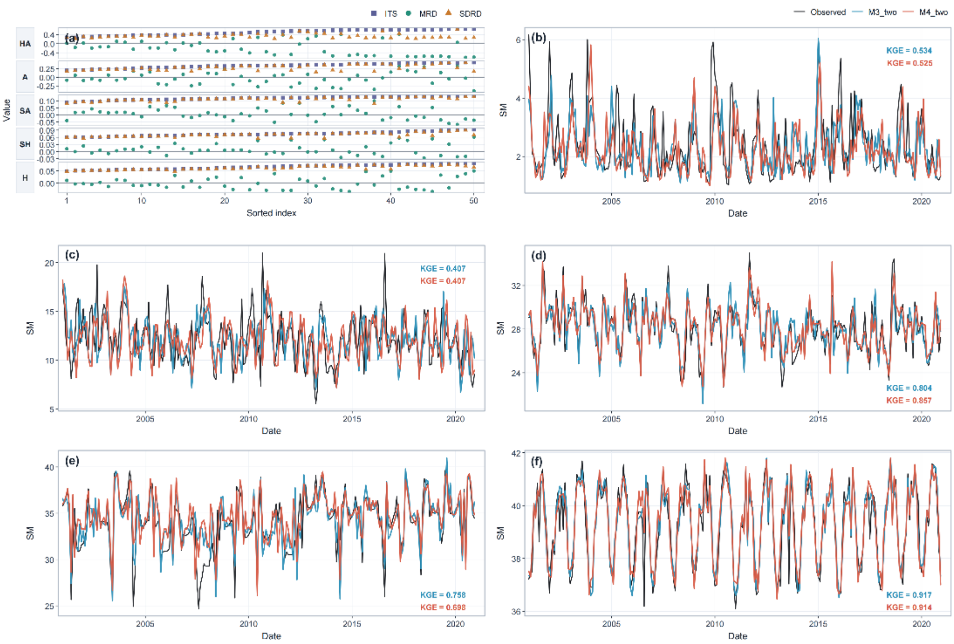

Figure 11: Linear relationships between normalized ITS and KGE across AI regions and model configurations.

The relationship between local temporal stability and prediction skill was further evaluated across all two-stage model configurations using the normalized ITS within each AI region (Fig. 12). The linear fits revealed broadly consistent regional patterns across the six models, but the direction and significance of the ITS-KGE relationship differed among AI zones. In most non-humid regions, KGE increased with normalized ITS, indicating that grids with weaker temporal stability tended to show equal or higher prediction skill. This increasing tendency was most consistent in the SA region, where all six models showed significant positive slopes, and was also evident in SH for M3_two-M6_two. The A region showed the same positive direction across models, although the relationships were generally weaker, whereas HA was more variable and M1_two was the main exception to the otherwise increasing pattern. By contrast, the H region showed the opposite behavior: KGE generally decreased as normalized ITS increased, meaning that more temporally stable humid grids tended to be more predictable. These results indicated that ITS did not provide a universal monotonic indicator of grid-level prediction skill; instead, its relationship with KGE depended strongly on hydroclimatic context. Across the six models, the LOESS curves showed broadly similar regional shapes, confirming that the ITS-KGE relationship was not specific to a single model configuration (Fig. S15~S17). Thus, temporal stability alone was insufficient to explain local predictability, showing that regional mean performance captured the dominant aridity-dependent structure of model skill, whereas ITS-based grid-level gradients exposed important local departures from that regional pattern.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

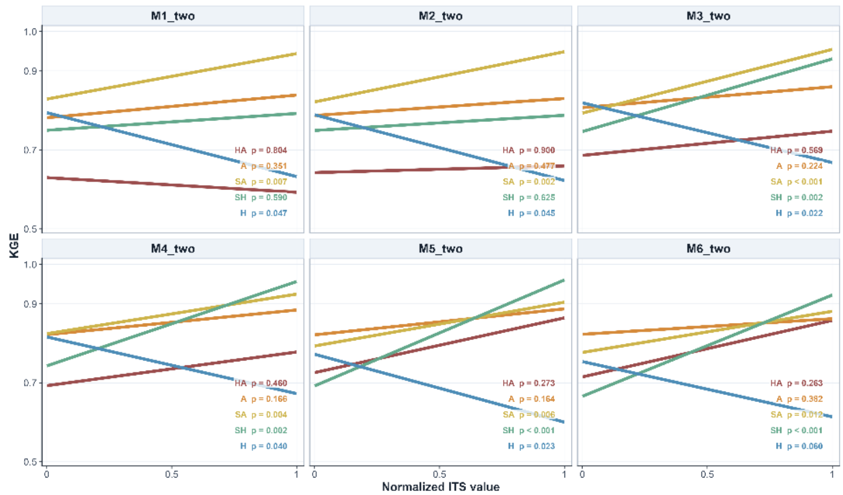

Figure 12: Linear relationships between normalized ITS and KGE across AI regions and model configurations.

4.3 SHAP-based Explainable Machine Learning for Driver Attribution and Mechanism Interpretation

4.3.1 SHAP-derived Dominant Controls and Aridity-dependent Driver Transitions

To resolve how the prediction framework encoded hydro-meteorological controls on SM, we interpreted the trained XGBoost component using SHapley Additive exPlanations (SHAP; Fig. 13). The SHAP summaries indicated that the model relied on a small group of physically interpretable controls, but the leading driver changed along the aridity gradient. In the hyper-arid region, evaporation (e) contributed most strongly, accounting for 21% of the relative mean absolute SHAP contribution, followed by antecedent SM (aswvl1) and precipitation (tp). The leading driver then shifted to aswvl1 in the arid region (28%) and to tp in the semi-arid region (29%). In dry sub-humid and humid regions, tp remained dominant, with its contribution increasing from 38% in SH to 43% in H. Together, these patterns defined a dry-to-wet transition from evaporative limitation in the most water-stressed grids, through an antecedent-memory regime, to precipitation-input dominance in wetter environments. Across all zones, ssr remained a secondary energy-related constraint, whereas aswvl2, aswvl3, sp and u clustered near zero and added limited information.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

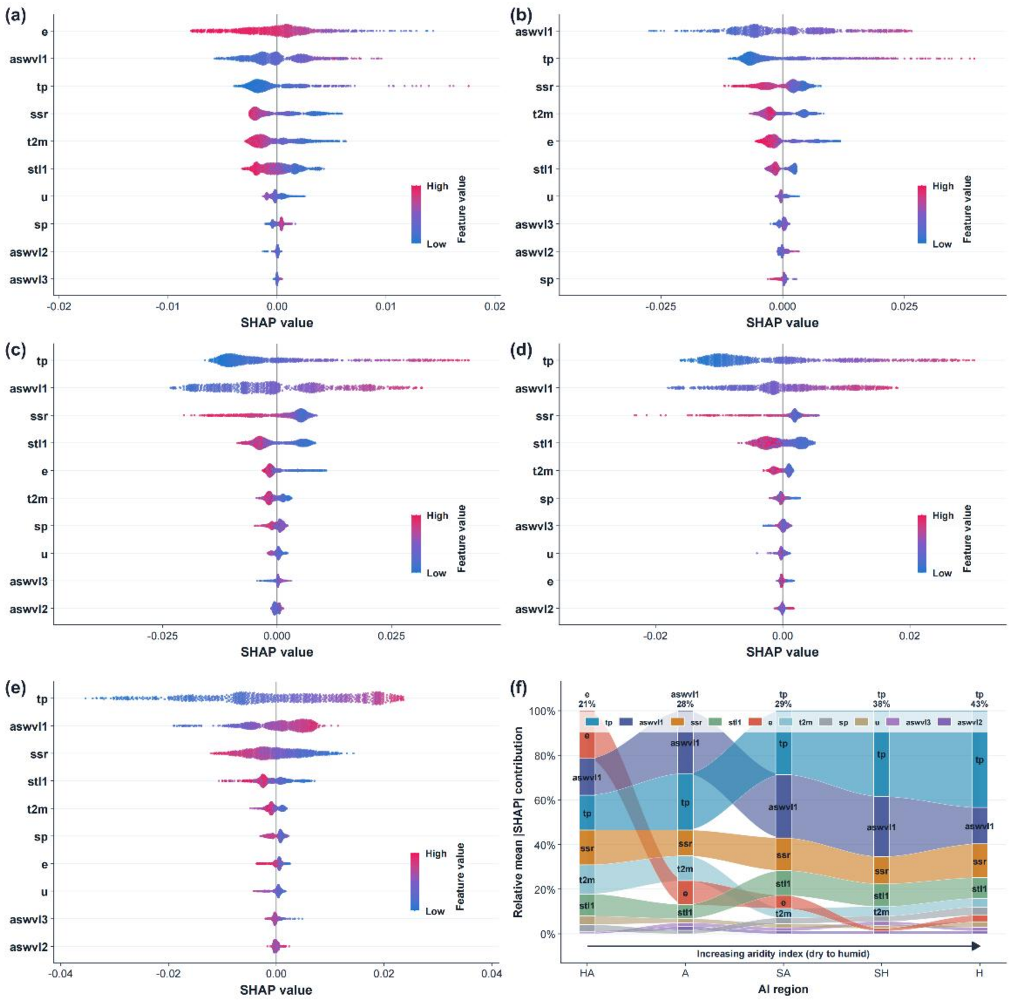

Figure 13: SHAP-derived dominant controls and aridity-dependent driver transitions. (a-e) SHAP summary distributions for HA, A, SA, SH and H, respectively. Points denote grid-level attribution values, and colors indicate relative feature values. (f) Relative mean absolute SHAP contribution along the aridity gradient.

The SHAP dependence plots further indicated that the dominant controls acted through distinct marginal-response forms across zones (Fig. 14). In HA, evaporation had a monotonic negative effect, with SHAP values declining as e increased. This pattern identified evaporative loss as the principal limiting process in the driest grids. Across all regions, aswvl1 and tp showed positive nonlinear marginal responses, indicating that antecedent wetness and precipitation increased the modelled SM contribution. In A, SA, SH and H, ssr ranked among the top three controls and displayed a nonlinear negative response. SHAP values were close to zero or weakly positive at lower radiation levels, but declined rapidly beyond a region-specific radiation range. This response was strongest in the semi-arid and sub-humid panels, where high ssr values were associated with sharply negative SHAP contributions.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

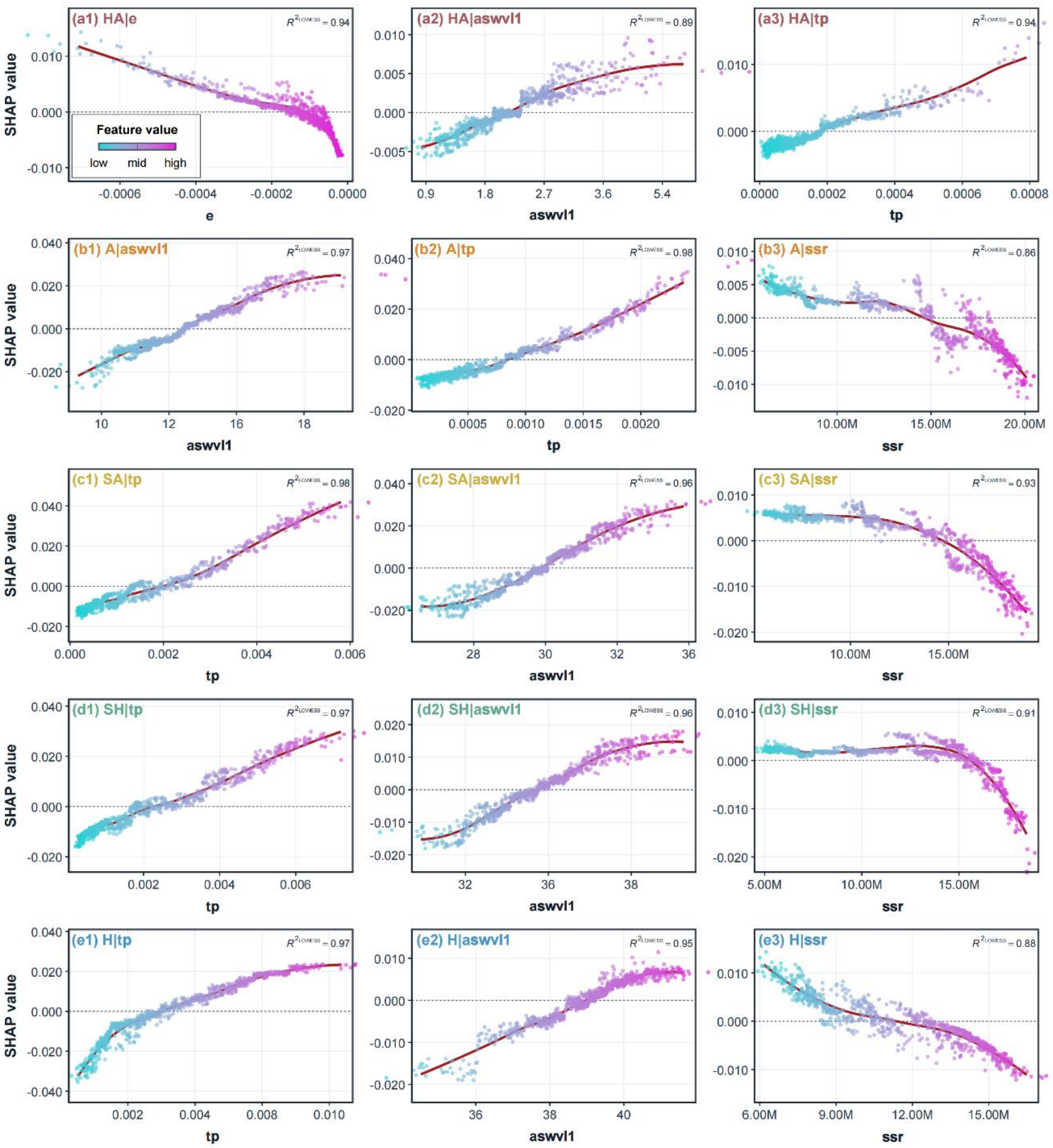

Figure 14: SHAP dependence of the top three drivers within each aridity zone.

4.3.2 Rank Robustness of SHAP Attribution Against Alternative Selection Methods

The SHAP ranking was evaluated against three alternative selection or importance diagnostics to determine whether the inferred driver hierarchy was method-dependent (Fig. 15). SHAP showed consistently high rank agreement with Random Forest (RF) and XGBoost (XGB) feature importance across the five aridity zones and the overall domain (Fig. 15(a)). Rank

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

correlations ranged from 0.81 to 0.92 for SHAP versus RF and from 0.70 to 0.92 for SHAP versus XGB. The strongest agreement occurred in HA and the overall domain for the SHAP-XGB comparison, and in SA for the SHAP-RF comparison. By contrast, SHAP-CCM rank correlations ranged from 0.10 in HA to 0.53 for the overall domain, indicating that causal-selection frequency and predictive attribution magnitude captured related but non-equivalent information. This contrast was also evident in the feature-level relationship between normalized mean |SHAP| importance and normalized CCM selection frequency (Fig. 15(b)). The dominant variables tp and aswvl1 occupied the high-importance, high-support part of the feature space and also received strong RF/XGB support. However, aswvl2 and aswvl3 had high CCM selection frequencies but low SHAP importance, indicating that deeper antecedent soil-water layers were frequently selected by CCM but added little marginal predictive information once aswvl1 and tp were included. Conversely, e and ssr retained measurable SHAP importance despite more moderate CCM support. The supplementary full-ranking comparison in Fig. S18 gave the same overall contrast: SHAP, RF and XGB prioritized the main water-supply and memory variables, whereas CCM ranked antecedent soil-water layers more uniformly across zones. Thus, SHAP attribution was robust relative to tree-based importance rankings, while its divergence from CCM showed that attribution magnitude, split-based importance and nonlinear causal selection diagnosed different aspects of the predictor system.

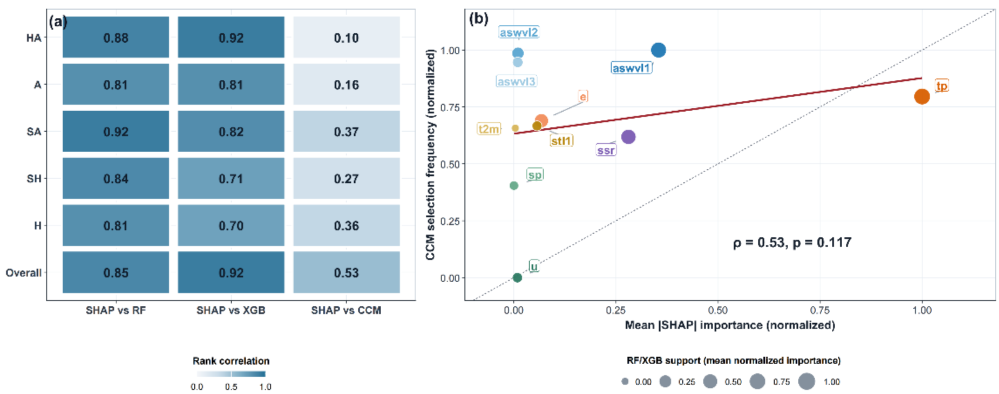

Figure 15: Cross-method robustness of SHAP attribution. (a) Rank correlations between SHAP and Random Forest (RF), XGBoost (XGB) and CCM feature rankings for each aridity zone and for the overall domain. (b) Relationship between normalized mean |SHAP| importance and normalized CCM selection frequency for the ten predictors. Circle size indicates RF/XGB support, calculated as the mean normalized importance across the two tree-based methods.

5 Discussion

5.1 Causality-Guided Predictor Selection and Explainable Driver Attribution

In non-stationary SM prediction, the rational selection of predictors is essential for improving model stability, enhancing physical interpretability, and ensuring the reliability of prediction results. In this study, Convergent Cross Mapping (CCM) was employed to identify causal drivers among ten candidate variables, and the top 4 selected variables are shown in Fig. 16. Results indicated that antecedent SM and precipitation exhibited consistently high selection frequencies

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

across most grids, with aswvl1~3 ranking among the variables with the highest regional mean selection rates. This finding suggested that the SM system possessed pronounced memory effects and state persistence, meaning that its evolution was governed not only by instantaneous meteorological forcing but also by long-term constraints imposed by antecedent wetness conditions.

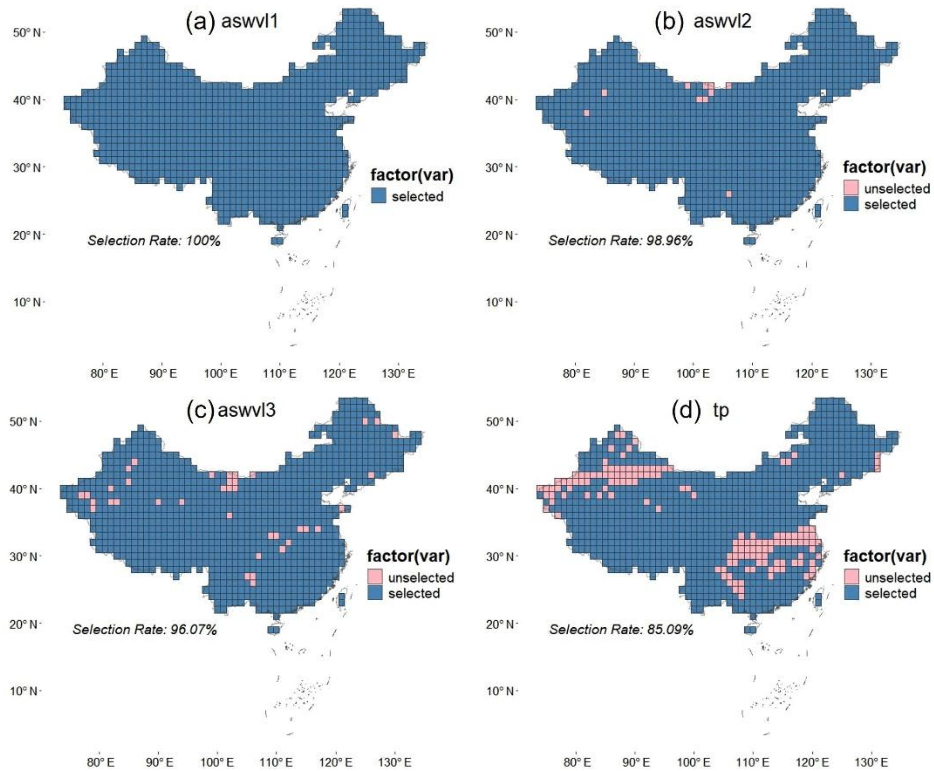

Figure 16: Spatial frequency results of the top 4 driving variables selected by CCM.

Compared with traditional statistical-based correlation approaches, such as Pearson/Spearman correlation analysis and Mutual Information (MI), the major advantage of CCM lies in its capability to characterize nonlinear dynamical relationships under non-stationary conditions. The Pearson/Spearman methods primarily quantify linear or monotonic statistical dependencies and are prone to generating spurious correlations in the presence of common trends or low-frequency variability. Although MI can capture certain nonlinear relationships, it remains fundamentally a statistical dependency measure and is often unable to distinguish true causal drivers from pseudo-associations induced by common external forcing (Laarne et al., 2022; Liang et al., 2025; Runge, 2018). In recent years, Granger causality analysis and other predictive causality frameworks have been increasingly applied in hydro-meteorological and drought-propagation studies, because they infer causal directions based on predictive improvements in time series analysis (Nowack et al., 2020; Pang & Yang, 2025; Zhou et al., 2025). However, the classical Granger framework is inherently built upon linear autoregressive assumptions and generally requires strong stationarity conditions. Consequently, its applicability may be limited for SM systems characterized

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

by pronounced nonlinear feedbacks, time-varying relationships, and multiscale coupling processes (Liang et al., 2025). By contrast, the CCM method adopted in this study is grounded in Takens' state-space reconstruction theory (Takens, 1981). Rather than relying on predefined functional forms, CCM identifies causal relationships by examining whether variables share information from the same system attractor manifold (Sugihara et al., 2012). Recent studies have applied CCM or its extended variants to groundwater-streamflow interactions, wetland hydrological connectivity, soil moisture-precipitation coupling, drought-type propagation, and catchment-scale hydrological causality, demonstrating its potential for revealing nonlinear and non-separable causal structures in complex water-related systems (Ban et al., 2025; Bonotto et al., 2022; S. Lee et al., 2023; J. Yu et al., 2026; Zhao et al., 2024). This framework enables CCM to extract intrinsic dynamical signals from noisy and non-stationary time series, making it particularly suitable for causal inference in complex hydrological systems. For strongly nonlinear SM systems with feedback mechanisms, recent soil moisture and drought studies further suggest that CCM-based approaches can help characterize causal pathways and propagation chains that are difficult to resolve using conventional correlation or purely predictive methods (Pang & Yang, 2025; Sun et al., 2025; You et al., 2025; Zhao et al., 2024). Therefore, CCM provides a more physically meaningful basis for predictor selection under non-stationary conditions.

An important finding of this study is the apparent discrepancy between CCM-based feature selection and SHAP-based feature contribution analysis. Although aswvl2 and aswvl3 were frequently identified as important variables by CCM, their SHAP contribution values were substantially lower than those of precipitation (tp) and the most recent SM state (aswvl1). While this phenomenon appeared contradictory at the first glance, it actually reflected the fundamental distinction between causal relevance and predictive contribution. From the perspective of dynamical systems theory, SM evolution is a continuous process characterized by strong memory effects, temporal autocorrelation, and lagged responses to meteorological forcing (Basir et al., 2024; R. Li et al., 2024; Palazzolo et al., 2025). To adequately reconstruct the underlying state-space manifold, sufficient historical state information must be incorporated to reduce the "projection overlap" problem associated with low-dimensional embeddings (Sugihara et al., 2012). Therefore, although longer-lag SM variables (aswvl2 and aswvl3) are important for representing the long-term memory structure of the system and contribute substantially to dynamical reconstruction, much of their information content can already be effectively represented by the most recent SM condition (aswvl1) due to strong temporal autocorrelation among lagged states (Basir et al., 2024; Ley et al., 2024; R. Li et al., 2024). As a result, the additional predictive information provided by aswvl2 and aswvl3 is relatively limited, leading to lower marginal contributions in the SHAP analysis. This interpretation is consistent with recent explainable hydrological modeling studies showing that SHAP quantifies how strongly model predictions depend on a variable under the fitted model structure, rather than whether that variable is dynamically causal in the underlying physical system (Ley et al., 2024; Nikraftar et al., 2025). These findings indicate that CCM and SHAP characterize variable importance from two fundamentally different perspectives: CCM emphasizes whether a variable genuinely participates in the dynamical evolution of the system, namely its dynamical causality, whereas SHAP focuses on the extent to which model

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

predictions depend on a variable, namely its predictive sensitivity (L. Slater et al., 2025). Therefore, the two approaches are not inherently contradictory but rather complementary in revealing the roles of driving factors from different dimensions.

Overall, results suggested that predictor selection in non-stationary SM prediction should not rely solely on statistical correlations or machine-learning-based contribution metrics, but should additionally consider the causal participation of variables within the system dynamics. Accordingly, this study proposes that CCM should be employed as a “physical filter” to ensure the dynamical completeness of the predictor set, whereas SHAP can serve as a “sensitivity evaluator” to identify dominant predictive variables and optimize computational efficiency. Such a framework is consistent with recent efforts to combine physical mechanisms, causal inference, and explainable machine learning in hydrological prediction, where interpretability is improved by distinguishing physically meaningful drivers from variables that merely improve local predictive accuracy (Hosseini et al., 2025; Ley et al., 2024; Liang et al., 2025; S. Wang et al., 2024). This facilitates the incorporation of physical consistency into data-driven models, thereby improving the reliability and interpretability of predictions in complex hydrological systems.

5.2 Mechanistic Contributions of Periodic, Trend, and Residual Components to Prediction Performance

In this study, a systematic comparison of different model structures revealed that SM predictive performance under non-stationary conditions is not determined by a single non-stationary characteristic, but rather by the combined influence of periodic, trend, and residual components. These three components correspond to seasonal regularity, long-term state drift, and unexplained residual variability, respectively, each exerting a distinct mechanistic impact on predictive performance.

First, explicitly separating the periodic component significantly improved model performance. As shown in Fig. 3, models M3~M6 with period separation consistently outperformed the unprocessed models M1 and M2 across KGE, MAE, and PBIAS metrics, indicating that the periodic component represents a critical structural feature that stabilizes SM predictions. This enhancement was primarily attributable to the pronounced seasonal cyclicity inherent in SM (J. Zhu et al., 2025). Without period separation, models must simultaneously fit intra-annual cyclic fluctuations, long-term trends, and short-term stochastic disturbances, resulting in overlapping signals across different time scales and increased model complexity (Verbesselt et al., 2010). In non-stationary contexts, the unseparated seasonal component may be incorrectly attributed to trends or residuals, biasing trend estimates, complicating residual structures, and ultimately reducing model generalization (Wu et al., 2007). By isolating the periodic component, the model can first remove the most stable and repeatable seasonal structure, thereby focusing on capturing more complex dynamics, such as nonlinear responses, state memory, and anomalous perturbations (Ahmed et al., 2021; Liao et al., 2023; J. Wang et al., 2023; Z. Wang et al., 2024). Moreover, the observed variability in monthly predictive performance further highlighted the seasonally dependent nature of SM processes (Fig. 4). Different months exhibited distinct water supply, loss, and lagged response mechanisms; therefore, modeling separately by month or seasonal phase allowed the model to more accurately reflect dominant hydrological processes throughout the year.

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Second, although the non-stationary statistical models (M4) effectively captured trend-related changes in SM sequences and achieved overall good predictive performance (KGE = 0.77), it did not consistently outperform the stationary ones (M3) in practice. Spatial distributions of KGE differences across grids revealed that even in grids identified as non-stationary, the stationary model may achieve comparable or better predictive performance (Fig. 6). This indicated that process non-stationarity does not automatically confer predictive superiority (Luke et al., 2017). The fundamental reason lies in the difference between process identification and predictive modeling objectives. Non-stationarity tests typically rely on statistical features, such as changes in mean, variance, or trends to determine whether a time series exhibits long-term drift (Apurv & Cai, 2019; Kathuria et al., 2019; W. Li et al., 2026). In contrast, predictive performance depends on whether the model can extract sufficiently stable, transferable, and informative features from the inputs (Liu et al., 2025; Nearing et al., 2021; Z. Wang et al., 2024). In SM systems, recent SM states, precipitation, radiation, and evapotranspiration already encapsulated much of the information about the current state and potential future dynamics. In regions with strong memory effects, antecedent SM effectively encoded cumulative precipitation, evaporative deficits, and soil water storage, indirectly capturing trend-related information (Feng et al., 2020; Kratzert et al., 2019; Lees et al., 2021; C. Yu et al., 2025). Consequently, explicitly adding a trend component may provide little additional predictive information in some non-stationary grids. Furthermore, non-stationary models generally involve higher parameter complexity or increased structural flexibility; if the trend signal is weak, sample size is limited, or the relationship between trend and prediction target is unstable, fitting the trend may introduce additional uncertainty or even misinterpret noise as signal, increasing the risk of overfitting (Guan et al., 2015; Hastie et al., 2009). For DL models, which optimize primarily for fit rather than physical consistency, internal feature representations do not necessarily correspond to hydrological mechanisms, and "effectiveness" is judged by predictive error minimization rather than adherence to physical laws (Hosseini et al., 2025; Ley et al., 2024; Shen et al., 2023; Willard et al., 2023). ITS-based representative grid analysis further showed that grids representative of regional mean SM did not necessarily reflect representative predictive performance (Fig. 11), highlighting the decoupling between physical representativeness of SM sequences and their learnability by predictive models. These findings underscore the importance of integrating physical constraints, state information, and data-driven modeling, rather than focusing solely on predictive metrics.

Finally, residual analysis demonstrated that NGBoost residual learning substantially improved the performance of statistical-DL hybrid models (M3 and M4) but had negligible effect—or may even degrade performance—when applied to purely DL models (M5 and M6) (Fig. 8). The effectiveness of residual learning depended on the presence of systematic, learnable structure in the initial model residuals (Ilhan et al., 2024; Liu et al., 2024). For M3 and M4, the residuals from the statistical models were relatively structured, containing identifiable biases that NGBoost could learn and correct. In contrast, the residuals of M5 and M6 after one-stage prediction were largely random or non-Gaussian, with weak conditional dependencies, making further modeling less effective. When residuals lacked stable signals, applying a second-stage model risks overfitting noise, reducing generalization (De Oliveira et al., 2022; Hamoudia et al., 2023; Livieris et al., 2020; Theodorakos et al., 2022). LOESS analysis further supported this explanation: grids with lower initial KGE in statistical

© Author(s) 2026. CC BY 4.0 License.

models showed larger two-stage improvement, indicating that systematic error remained to be captured, whereas the correlation between initial performance and secondary gain was weaker for DL models, reflecting residual randomness and limited learnability (Fig. 9). Residual learning should therefore be treated as a targeted correction for structured first-stage errors, not as an automatically beneficial add-on.

In summary, under non-stationary conditions, SM predictive performance was jointly influenced by periodic, trend, and residual components, but each operates via distinct mechanisms. As a result, future non-stationary SM prediction can be improved in three ways: (i) prioritize the separation of seasonal periodic components to reduce sequence non-stationarity and increase stability; (ii) selectively incorporate non-stationary trends, avoiding unnecessary complexity and integrating physical constraints into DL frameworks; and (iii) design residual learning modules specifically for statistical or structured models to fully exploit systematic error, rather than applying second-stage residual correction indiscriminately to all models.

5.3 Limitations and Future Work

Although the non-stationary SM prediction framework developed in this study achieved satisfactory performance in terms of periodic component separation, non-stationary marginal distribution modeling, and residual-based two-stage prediction, several limitations remain and should be further addressed in future research.

First, the framework decomposed SM non-stationarity into periodic, trend and residual components, but SM dynamics are also controlled by land-surface heterogeneity and human disturbance. Variables such as soil texture, soil hydraulic properties, vegetation condition, land use, irrigation and groundwater interactions were not fully represented. These factors can alter infiltration, retention, evapotranspiration and vertical water movement, particularly at local scales (Dari et al., 2024; Pan et al., 2026; Qiu et al., 2024; Rani et al., 2024; Tian et al., 2023). Although several hydro-meteorological variables were incorporated as predictors in this study, some influencing factors were still not fully considered. For example, Pan et al. (2026) have shown that soil texture and soil hydraulic properties substantially affect soil water retention, infiltration, and vertical water movement. Qiu et al. (2024) proved that vegetation cover and land-use change may regulate SM dynamics by modifying canopy interception, root water uptake, and evapotranspiration. Meanwhile, irrigation and human water use may also alter the natural evolution of regional SM (Dari et al., 2024; Rani et al., 2024; Tian et al., 2023). Therefore, at the local scale, our proposed model may still have limited ability to fully explain anomalous SM fluctuations caused by land-surface heterogeneity and anthropogenic disturbances. Future work should integrate multi-source remote sensing products, vegetation indices, soil property datasets, land-use information and irrigation records to better constrain local SM variability.

Second, the current trend modelling strategy captured gradual covariate-linked changes in marginal distributions, but it did not fully represent abrupt shifts, regime changes, episodic extremes or stage-dependent responses. The weak relationship between M-K trend detection and ΔKGE indicates that trend non-stationarity should not be incorporated solely because it is statistically detectable. Future studies could combine trend diagnostics with change-point detection, dynamic parameter models, time-varying distributions or state-space formulations to distinguish predictive trends from non-informative drift (T. Lee & Ouarda, 2023; Luke et al., 2017; L. J. Slater et al., 2021). Such approaches may be especially important under climate

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

change, where SM responses can shift nonlinearly after drought, heatwave or land-use disturbance events (Kathuria et al., 2019; C. Zhu et al., 2023).

Third, although this study evaluated accuracy, spatial robustness and interpretability, it did not fully optimize computational efficiency or operational scalability. On the one hand, monthly modelling, residual correction and explainability analyses increase the number of models, diagnostics and post-processing steps. On the other hand, although the Attention mechanism enhances models' ability to capture key temporal information, it also further increases model complexity (Vaswani et al., 2017). This is acceptable for a mechanistic evaluation, but it may constrain high-resolution or near-real-time SM prediction. Future work should examine lightweight neural architectures, transfer learning, multi-task learning, pruning and knowledge distillation to reduce computational cost while retaining physical interpretability (Cheng et al., 2024).

Overall, future research on non-stationary SM prediction should, while continuing to improve predictive accuracy, place greater emphasis on the coordinated optimization of physical process constraints, non-stationary structure identification, computational efficiency, and practical application requirements. By integrating multi-source data, causal inference, physical constraints, and explainable artificial intelligence, SM prediction models are expected to achieve greater reliability, generalizability, and application value under complex climate change conditions.

6 Conclusions

This study developed a statistical-DL hybrid prediction framework that integrates causal predictor selection, statistical marginal distribution modeling, deep learning-based one-stage prediction, and residual-based two-stage prediction. The framework was used to systematically compare the effects of three types of non-stationary components—periodic, trend, and residual components—on SM prediction performance. The model applicability and predictor contribution mechanisms under different aridity conditions were further analyzed. The main conclusions are as follows:

(1) Seasonal periodicity is the dominant source of transferable predictive structure in SM series. Trend-related non-stationarity is a useful diagnostic feature but does not consistently improve prediction when incorporated into marginal distributions, and residual learning should be applied selectively because it improves performance only when the remaining first-stage errors are structured and learnable.

(2) The two-stage framework exhibits clear hydroclimatic dependence, with the most stable skill concentrated in arid and semi-arid regions rather than at the driest or wettest extremes. Representative-grid and ITS-based analyses further show that regional mean performance and local grid-level predictability are not interchangeable, highlighting the need to assess model robustness across both aridity zones and local temporal-stability gradients.

(3) SHAP analysis identified a coherent dry-to-wet transition in dominant controls, from evaporation in hyper-arid grids to antecedent SM memory in arid grids and precipitation dominance in wetter regions. The agreement between SHAP and

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

tree-based importance rankings, together with its divergence from CCM, indicates that predictive attribution and dynamical causal participation are complementary for explaining non-stationary SM prediction.

Overall, the proposed non-stationary SM prediction framework can effectively identify and utilize periodic, trend-related, and residual structural information in SM series, thereby improving prediction accuracy, while enhancing the physical interpretability of the model. This study provides methodological guidance and theoretical support for SM prediction, agricultural drought monitoring, and regional water resources management under complex climate change conditions.

Code and data availability

The ERA5-Land monthly averaged data used in this study are publicly available from the Copernicus Climate Data Store (CDS), European Centre for Medium-Range Weather Forecasts (ECMWF): https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land-monthly-means. The four monthly large-scale climate indices used as covariates, including Niño 3.4 (ENSO index), the North Atlantic Oscillation (NAO), the Arctic Oscillation (AO), and the Pacific Decadal Oscillation (PDO), are available from the NOAA Physical Sciences Laboratory monthly climate time-series archive: https://psl.noaa.gov/data/timeseries/month/. The Global Aridity Index and Potential Evapotranspiration Climate Database version 3 (Global-AI_PET_v3), used to derive the aridity index (AI) and hydroclimatic zones, is available through https://cgiarcsi.community/2019/01/24/global-aridity-index-and-potential-evapotranspiration-climate-database-v3/. All Codes are available on request.

Author contributions

CLY: Conceptualization, Formal analysis, Methodology, Data curation, Writing – original draft. **DW:** Conceptualization, Methodology, Funding acquisition, Supervision, Writing – review & editing. **VS:** Writing – review & editing**. PCX:** Data curation. **CJY:** Data curation. **XKZ:** Data curation. **JGJ:** Data curation. **ML:** Data curation. **QL:** Data curation. **STZ:** Data curation. **JCW:** Writing – review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

This study was supported by the National Science and Technology Major Project on Comprehensive Environmental Governance in the Beijing-Tianjin-Hebei Region (2025ZD1205901) and the National Natural Science Foundation of China (No. 42571040).

References

Abbes, A. B., Jarray, N., & Farah, I. R. (2024). Advances in remote sensing based soil moisture retrieval: Applications, techniques, scales and challenges for combining machine learning and physical models. Artificial Intelligence Review, 57(9), 224. https://doi.org/10.1007/s10462-024-10734-1

Ahmed, A. A. M., Deo, R. C., Raj, N., Ghahramani, A., Feng, Q., Yin, Z., & Yang, L. (2021). Deep Learning Forecasts of Soil Moisture: Convolutional Neural Network and Gated Recurrent Unit Models Coupled with Satellite-Derived MODIS, Observations and Synoptic-Scale Climate Index Data. Remote Sensing, 13(4), 554. https://doi.org/10.3390/rs13040554

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Apurv, T., & Cai, X. (2019). Evaluation of the Stationarity Assumption for Meteorological Drought Risk Estimation at the Multidecadal Scale in Contiguous United States. Water Resources Research, 55(6), 5074–5101. https://doi.org/10.1029/2018WR024047

Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural Machine Translation by Jointly Learning to Align and Translate (arXiv:1409.0473). arXiv. https://doi.org/10.48550/arXiv.1409.0473

Ban, J., Ni, S., Han, F., He, B., & Wang, W. (2025). Uncovering nonlinear causal relationships and propagation dynamics of drought types in Xinjiang using convergent cross mapping. Journal of Environmental Management, 393, 127165. https://doi.org/10.1016/j.jenvman.2025.127165

Basir, Md. S., Noel, S., Buckmaster, D., & Ashik-E-Rabbani, M. (2024). Enhancing Subsurface Soil Moisture Forecasting: A Long Short-Term Memory Network Model Using Weather Data. Agriculture, 14(3), 333. https://doi.org/10.3390/agriculture14030333

Baumberger, M., Haas, B., Tewes, W., Risse, B., Meyer, N., & Meyer, H. (2025). Gated recurrent units for modelling time series of soil temperature and moisture: An assessment of performance and process reflectivity. Environmental Modelling & Software, 183, 106245. https://doi.org/10.1016/j.envsoft.2024.106245

Berg, A., Sheffield, J., & Milly, P. C. D. (2017). Divergent surface and total soil moisture projections under global warming. Geophysical Research Letters, 44(1), 236–244. https://doi.org/10.1002/2016GL071921

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., … Živković, N. (2019). Changing climate both increases and decreases European river floods. Nature, 573(7772), 108–111. https://doi.org/10.1038/s41586-019-1495-6

Bonotto, G., Peterson, T. J., Fowler, K., & Western, A. W. (2022). Identifying Causal Interactions Between Groundwater and Streamflow Using Convergent Cross-Mapping. Water Resources Research, 58(8), e2021WR030231. https://doi.org/10.1029/2021WR030231

Brocca, L., Melone, F., Moramarco, T., & Morbidelli, R. (2009). Soil moisture temporal stability over experimental areas in Central Italy. Geoderma, 148(3–4), 364–374. https://doi.org/10.1016/j.geoderma.2008.11.004

Cheng, H., Zhang, M., & Shi, J. Q. (2024). A Survey on Deep Neural Network Pruning: Taxonomy, Comparison, Analysis, and Recommendations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 10558–10578. https://doi.org/10.1109/TPAMI.2024.3447085

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling (arXiv:1412.3555). arXiv. https://doi.org/10.48550/arXiv.1412.3555

Dari, J., Filippucci, P., & Brocca, L. (2024). The development of an operational system for estimating irrigation water use reveals socio-political dynamics in Ukraine. Hydrology and Earth System Sciences, 28(12), 2651–2659. https://doi.org/10.5194/hess-28-2651-2024

De Oliveira, J. F. L., Silva, E. G., & De Mattos Neto, P. S. G. (2022). A Hybrid System Based on Dynamic Selection for Time Series Forecasting. IEEE Transactions on Neural Networks and Learning Systems, 33(8), 3251–3263. https://doi.org/10.1109/TNNLS.2021.3051384

Duan, T., Avati, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A. Y., & Schuler, A. (2020). NGBoost: Natural Gradient Boosting for Probabilistic Prediction (arXiv:1910.03225). arXiv. https://doi.org/10.48550/arXiv.1910.03225

Feng, D., Fang, K., & Shen, C. (2020). Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales. Water Resources Research, 56(9), e2019WR026793. https://doi.org/10.1029/2019WR026793

Guan, M., Sillanpää, N., & Koivusalo, H. (2015). Modelling and assessment of hydrological changes in a developing urban catchment. Hydrological Processes, 29(13), 2880–2894. https://doi.org/10.1002/hyp.10410

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hamoudia, M., Makridakis, S., & Spiliotis, E. (Eds.). (2023). Forecasting with Artificial Intelligence: Theory and Applications. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35879-1

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hosseini, F., Prieto, C., & Álvarez, C. (2025). An explainable AI approach for interpreting regionally optimized deep neural networks in hydrological prediction. Journal of Hydrology, 661, 133689. https://doi.org/10.1016/j.jhydrol.2025.133689

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297–307.

Ilhan, E., Koc, A. B., & Kozat, S. S. (2024). Exploiting residual errors in nonlinear online prediction. Machine Learning, 113(9), 6065–6091. https://doi.org/10.1007/s10994-024-06554-7

Jacobs, J. (2004). SMEX02: Field scale variability, time stability and similarity of soil moisture. Remote Sensing of Environment, 92(4), 436–446. https://doi.org/10.1016/j.rse.2004.02.017

Kathuria, D., Mohanty, B. P., & Katzfuss, M. (2019). A Nonstationary Geostatistical Framework for Soil Moisture Prediction in the Presence of Surface Heterogeneity. Water Resources Research, 55(1), 729–753. https://doi.org/10.1029/2018WR023505

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences, 23(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Laarne, P., Amnell, E., Zaidan, M. A., Mikkonen, S., & Nieminen, T. (2022). Exploring Non-Linear Dependencies in Atmospheric Data with Mutual Information. Atmosphere, 13(7), 1046. https://doi.org/10.3390/atmos13071046

Lee, S., Lee, B., Lee, J., Song, J., & McCarty, G. W. (2023). Detecting causal relationship of non-floodplain wetland hydrologic connectivity using convergent cross mapping. Scientific Reports, 13(1), 17220. https://doi.org/10.1038/s41598-023-44071-0

Lee, T., & Ouarda, T. B. M. J. (2023). Trends, Shifting, or Oscillations? Stochastic Modeling of Nonstationary Time Series for Future Water-Related Risk Management. Earth's Future, 11(7), e2022EF003049. https://doi.org/10.1029/2022EF003049

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., & Dadson, S. J. (2021). Benchmarking data-driven rainfall–runoff models in Great Britain: A comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. Hydrology and Earth System Sciences, 25(10), 5517–5534. https://doi.org/10.5194/hess-25-5517-2021

Ley, A., Bormann, H., & Casper, M. (2024). Linking explainable artificial intelligence and soil moisture dynamics in a machine learning streamflow model. Hydrology Research, 55(6), 613–627. https://doi.org/10.2166/nh.2024.003

Li, Q., Zhu, Y., Shangguan, W., Wang, X., Li, L., & Yu, F. (2022). An attention-aware LSTM model for soil moisture and soil temperature prediction. Geoderma, 409, 115651. https://doi.org/10.1016/j.geoderma.2021.115651

Li, R., Zhang, S., Li, F., Lin, X., Luo, M., Wang, S., Yang, L., & Zhao, X. (2024). Impact of time-lagging and time-preceding environmental variables on top layer soil moisture in semiarid grasslands. Science of The Total Environment, 912, 169406. https://doi.org/10.1016/j.scitotenv.2023.169406

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Li, W., Xiong, L., Zhou, Y., Li, M., Li, R., & Xu, C.-Y. (2026). Introducing an explainable neural network framework for nonstationary flood frequency analysis. Journal of Hydrology, 665, 134729. https://doi.org/10.1016/j.jhydrol.2025.134729

Liang, H., Wang, W., Liu, D., Chen, B., Guo, L., Liu, H., Yu, S., & Zhang, D. (2025). Inferring causal associations in hydrological systems: A comparison of methods. Stochastic Environmental Research and Risk Assessment, 39(6), 2427–2448. https://doi.org/10.1007/s00477-025-02977-3

Liao, S., Wang, H., Liu, B., Ma, X., Zhou, B., & Su, H. (2023). Runoff Forecast Model Based on an EEMD-ANN and Meteorological Factors Using a Multicore Parallel Algorithm. Water Resources Management, 37(4), 1539–1555. https://doi.org/10.1007/s11269-023-03442-y

Liu, J., Koch, J., Stisen, S., Troldborg, L., & Schneider, R. J. M. (2024). A national-scale hybrid model for enhanced streamflow estimation – consolidating a physically based hydrological model with long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 28(13), 2871–2893. https://doi.org/10.5194/hess-28-2871-2024

Liu, J., Shen, C., O'Donncha, F., Song, Y., Zhi, W., Beck, H. E., Bindas, T., Kraabel, N., & Lawson, K. (2025). From RNNs to Transformers: Benchmarking deep learning architectures for hydrologic prediction. Hydrology and Earth System Sciences, 29(23), 6811–6828. https://doi.org/10.5194/hess-29-6811-2025

Livieris, I. E., Stavroyiannis, S., Pintelas, E., & Pintelas, P. (2020). A novel validation framework to enhance deep learning models in time-series forecasting. Neural Computing and Applications, 32(23), 17149–17167. https://doi.org/10.1007/s00521-020-05169-y

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the U nited S tates. Water Resources Research, 53(7), 5469–5494. https://doi.org/10.1002/2016WR019676

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, 4768–4777.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity Is Dead: Whither Water Management? Science, 319(5863), 573–574. https://doi.org/10.1126/science.1151915

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J.-N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data, 13(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What Role Does Hydrological Science Play in the Age of Machine Learning? Water Resources Research, 57(3), e2020WR028091. https://doi.org/10.1029/2020WR028091

Nikraftar, Z., Parizi, E., Saber, M., Boueshagh, M., Tavakoli, M., Esmaeili Mahmoudabadi, A., Ekradi, M. H., Mbuvha, R., & Hosseini, S. M. (2025). An Interpretable Machine Learning Framework for Unraveling the Dynamics of Surface Soil Moisture Drivers. Remote Sensing, 17(14), 2505. https://doi.org/10.3390/rs17142505

Nowack, P., Runge, J., Eyring, V., & Haigh, J. D. (2020). Causal networks for climate model evaluation and constrained projections. Nature Communications, 11(1), 1415. https://doi.org/10.1038/s41467-020-15195-y

Palazzolo, N., Cancelliere, A., Zofei, R. D., & Peres, D. J. (2025). Use of delayed ERA5-Land soil moisture products for improving landslide early warning. Natural Hazards and Earth System Sciences, 25(12), 4907–4919. https://doi.org/10.5194/nhess-25-4907-2025

Pan, Y., Wang, D., Lyu, S., Li, X., Li, S., & Su, Y. (2026). Evaluating saturated hydraulic conductivity schemes: Impacts on soil moisture simulations and soil texture-dependent applicability. Journal of Hydrology, 671, 135129. https://doi.org/10.1016/j.jhydrol.2026.135129

Pang, M., & Yang, S. (2025). Unraveling drought propagation dynamics using causal inference from time series. Environmental Research Letters, 20(12), 124030. https://doi.org/10.1088/1748-9326/ae151e

Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., Liu, H., Ma, Y., Ding, Y., Friedlingstein, P., Liu, C., Tan, K., Yu, Y., Zhang, T., & Fang, J. (2010). The impacts of climate change on water resources and agriculture in China. Nature, 467(7311), 43–51. https://doi.org/10.1038/nature09364

Priya, K., Reddy, V. M., Ray, L. K., & Das, J. (2024). Non-Stationary Flash Drought Analysis and Its Teleconnection With Low-Frequency Climatic Oscillations. Hydrological Processes, 38(10), e15315. https://doi.org/10.1002/hyp.15315

Qiu, D., Xu, R., Gao, P., & Mu, X. (2024). Effect of vegetation restoration type and topography on soil water storage and infiltration capacity in the Loess Plateau, China. CATENA, 241, 108079. https://doi.org/10.1016/j.catena.2024.108079

Rani, A., Sinha, N. K., Jyoti, B., Kumar, J., Kumar, D., Mishra, R., Singh, P., Mohanty, M., Jayaraman, S., Chaudhary, R. S., Lenka, N. K., Kumari, N., & Srivastava, A. (2024). Spatiotemporal Variations in Near-Surface Soil Water Content across Agroecological Regions of Mainland India: 1979–2022 (44 Years). Remote Sensing, 16(16), 3108. https://doi.org/10.3390/rs16163108

Rezaiy, R., & Shabri, A. (2025). Integrating wavelet transform and support vector machine for improved drought forecasting based on standardized precipitation index. Journal of Hydroinformatics, 27(2), 320–337. https://doi.org/10.2166/hydro.2025.292

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized Additive Models for Location, Scale and Shape. Journal of the Royal Statistical Society Series C: Applied Statistics, 54(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Runge, J. (2018). Causal network reconstruction from time series: From theoretical assumptions to practical estimation. Chaos: An Interdisciplinary Journal of Nonlinear Science, 28(7), 075310. https://doi.org/10.1063/1.5025050

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. Earth-Science Reviews, 99(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Shapley, L. (1953). A Value for n-Person Games. Contributions to the Theory of Games II. In H. W. Kuhn (Ed.), Classics in Game Theory (pp. 307–317). Princeton University Press. https://doi.org/doi:10.1515/9781400829156-012

Shawon, S. M., Neha, N. I., Jui, A. N., Dey, N., & Zubair, H. T. (2025). Advances in soil moisture measurement techniques and prediction using artificial intelligence: An extensive and systematic review. Smart Agricultural Technology, 12, 101613. https://doi.org/10.1016/j.atech.2025.101613

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., … Lawson, K. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. Nature Reviews Earth & Environment, 4(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

Slater, L., Blougouras, G., Deng, L., Deng, Q., Ford, E., Hoek van Dijke, A., Huang, F., Jiang, S., Yinxue, L., Moulds, S., Schepen, A., Yin, J., & Zhang, B. (2025). Challenges and opportunities of ML and explainable AI in large-sample hydrology. Philosophical Transactions A, 383. https://doi.org/10.1098/rsta.2024.0287

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T., Murphy, C., & Wilby, R. L. (2021). Nonstationary weather and water extremes: A review of methods for their detection, attribution, and management. Hydrology and Earth System Sciences, 25(7), 3897–3935. https://doi.org/10.5194/hess-25-3897-2021

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms (arXiv:1206.2944). arXiv. https://doi.org/10.48550/arXiv.1206.2944

Stasinopoulos, {D. Mikis}, Rigby, R., Heller, G., Voudouris, V., & {De Bastiani}, F. (2017). Flexible regression and smoothing: Using GAMLSS in R. Chapman & Hall/CRC. https://doi.org/10.1201/b21973

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting Causality in Complex Ecosystems. Science, 338(6106), 496–500. https://doi.org/10.1126/science.1227079

Sun, J., Yang, K., He, X., Wang, G., Wang, Y., Yu, Y., & Lu, H. (2025). Causal pathways underlying global soil moisture–precipitation coupling. Nature Communications, 16(1), 8935. https://doi.org/10.1038/s41467-025-63999-7

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), Dynamical Systems and Turbulence, Warwick 1980 (pp. 366–381). Springer Berlin Heidelberg.

Theodorakos, K., Agudelo, O. M., Espinoza, M., & De Moor, B. (2022). Decomposition-Residuals Neural Networks: Hybrid System Identification Applied to Electricity Demand Forecasting. IEEE Open Access Journal of Power and Energy, 9, 241–253. https://doi.org/10.1109/OAJPE.2022.3145520

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Tian, X., Dong, J., Jin, S., He, H., Yin, H., & Chen, X. (2023). Climate change impacts on regional agricultural irrigation water use in semi-arid environments. Agricultural Water Management, 281, 108239. https://doi.org/10.1016/j.agwat.2023.108239

UNEP. (1992). World Atlas of Desertification. Edward Arnold. https://books.google.com/books?id=5bGMQgAACAAJ

Vachaud, G., Passerat De Silans, A., Balabanis, P., & Vauclin, M. (1985). Temporal Stability of Spatially Measured Soil Water Probability Density Function. Soil Science Society of America Journal, 49(4), 822–828. https://doi.org/10.2136/sssaj1985.03615995004900040006x

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in Neural Information Processing Systems (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. Remote Sensing of Environment, 114(1), 106–115. https://doi.org/10.1016/j.rse.2009.08.014

Wang, J., Wang, X., & Khu, S. T. (2023). A Decomposition-based Multi-model and Multi-parameter ensemble forecast framework for monthly streamflow forecasting. Journal of Hydrology, 618, 129083. https://doi.org/10.1016/j.jhydrol.2023.129083

Wang, S., Liu, Y., Wang, W., Zhao, G., & Liang, H. (2024). Interpretable machine learning guided by physical mechanisms reveals drivers of runoff under dynamic land use changes. Journal of Environmental Management, 367, 121978. https://doi.org/10.1016/j.jenvman.2024.121978

Wang, Z., Xu, N., Bao, X., Wu, J., & Cui, X. (2024). Spatio-temporal deep learning model for accurate streamflow prediction with multi-source data fusion. Environmental Modelling & Software, 178, 106091. https://doi.org/10.1016/j.envsoft.2024.106091

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2023). Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. ACM Computing Surveys, 55(4), 1–37. https://doi.org/10.1145/3514228

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30, 79–82. https://doi.org/10.3354/cr030079

Wu, Z., Huang, N. E., Long, S. R., & Peng, C.-K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. Proceedings of the National Academy of Sciences, 104(38), 14889–14894. https://doi.org/10.1073/pnas.0701020104

You, Z., Sun, X., Sun, H., Chen, L., Lu, M., Xue, J., Ban, X., Yan, B., Tuo, Y., Qin, H., Zhang, L., & Zhang, W. (2025). Mechanisms of meteorological drought propagation to agricultural drought in China: Insights from causality chain. Npj Natural Hazards, 2(1), 24. https://doi.org/10.1038/s44304-025-00073-8

https://doi.org/10.5194/egusphere-2026-4803
Preprint. Discussion started: 21 September 2026
© Author(s) 2026. CC BY 4.0 License.

Yu, C., Wang, D., Singh, V. P., Xu, P., Zhang, A., Ye, X., Zeng, X., Jiang, J., & Wu, J. (2025). Rolling forecast of soil moisture under non-stationary conditions: A robust framework incorporating time-varying dynamics within and between variables. Journal of Hydrology, 662, 133832. https://doi.org/10.1016/j.jhydrol.2025.133832

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. https://doi.org/10.1016/j.agwat.2020.106649

Yu, J., Zhang, Y., Liang, H., Liu, H., Liu, J., Mello, C. R., Di, C., & Guo, L. (2026). A dual-method, multi-scale causal framework reveals seasonal shifts in hydrological causality of a headwater catchment. Journal of Hydrology, 667, 134891. https://doi.org/10.1016/j.jhydrol.2025.134891

Zhao, Y., Ma, E., Zhou, Z., Zou, Y., Cao, Z., Cai, H., Li, C., Yan, Y., & Chen, Y. (2024). Detecting the Non-Separable Causality in Soil Moisture-Precipitation Coupling With Convergent Cross-Mapping. Water Resources Research, 60(3), e2023WR034616. https://doi.org/10.1029/2023WR034616

Zheng, W., Zheng, K., Gao, L., Zhangzhong, L., Lan, R., Xu, L., & Yu, J. (2024). GRU–Transformer: A Novel Hybrid Model for Predicting Soil Moisture Content in Root Zones. Agronomy, 14(3), 432. https://doi.org/10.3390/agronomy14030432

Zhou, X., Sheng, Z., Manevski, K., Zhao, R., Zhang, Q., Yang, Y., Han, S., Liu, J., & Yang, Y. (2025). Can system dynamics explain long-term hydrological behaviors? The role of endogenous linking structure. Hydrology and Earth System Sciences, 29(1), 159–177. https://doi.org/10.5194/hess-29-159-2025

Zhu, C., Li, S., Hagan, D. F. T., Wei, X., Feng, D., Lu, J., Ullah, W., & Wang, G. (2023). Long-Term Characteristics of Surface Soil Moisture over the Tibetan Plateau and Its Response to Climate Change. Remote Sensing, 15(18), 4414. https://doi.org/10.3390/rs15184414

Zhu, J., Wang, S., & Li, Q. (2025). Establishing a periodic SM model with Fourier analysis for enhancing global soil moisture forecasting. Scientific Reports, 15(1), 15962. https://doi.org/10.1038/s41598-025-97347-y