



Local hydrological characteristics shape 1–14-day observation-driven soil moisture predictability

Feini Huang^{1,2} †, Ziqian Li^{2,3} †, Lijun Wang^{4,5,6}, Gustau Camps-Valls², Wei Su³,
Maria Piles², Ruqing Zhang⁷, Wei Shangguan^{1*}, Yongjiu Dai¹

¹Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Province
Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences,
Sun Yat-Sen University, Zhuhai 519082, China.

²Image Processing Laboratory (IPL), University of València, Valencia, Spain

³College of Land Science and Technology, China Agricultural University, Beijing 100083, China

⁴Guangdong Research Institute of Water Resources and Hydropower, Guangzhou, 510635, China

⁵Technology Innovation Center for Basin-wide Perception and Intelligent Regulation of Water
Network in the Guangdong-Hong Kong-Macao Greater Bay Area, Ministry of Water Resources,
510635, China

⁶Center of Water Resources and Environment, School of Civil Engineering, Sun Yat-sen University,
Guangzhou, China

⁷Shanwei Meteorological Service, Shanwei 516600, China

† These authors contributed equally to this work and share first authorship

*Correspondence: shgwei@mail.sysu.edu.cn; Tel.: +86-13466654629

Abstract

Data-driven machine learning approaches have advanced Soil Moisture (SM) forecasting at 1–14-day lead times using observational time series, but the hydrological controls on spatially heterogeneous forecast skill remain poorly understood, hindering targeted model improvement. Here, we show that spatial heterogeneity in data-driven SM predictability is associated with site-specific hydrological characteristics, particularly soil texture and precipitation extremes. Using a stacking ensemble, we established an empirical benchmark of SM predictability at 1–14-day lead times across three soil depths at 73 monitoring sites in Guangdong Province, China. Forecast skill separated into two regimes (slow-decay v.s. fast-decay), characterized by fast or slow decay as lead time increased, respectively, reflecting contrasting soil water dynamics. Event-scale diagnostics showed that fast-decay sites underwent stronger vertical redistribution and deeper SM responses during extreme precipitation, eroding the SM memory that supports longer-lead forecasts. A joint threshold of profile-wide silt content (>70%) and extreme daily precipitation (>28 mm/d) delineated that fast-decay regime. We synthesized these characteristics into Fast-Decay Index (FDI), which summarizes the local hydrological conditions associated with fast loss of predictability. Because its inputs are available prior to model training, the FDI has the potential to support pre-modeling assessment of SM predictability at new sites, pending independent-site validation. The FDI provides a physically interpretable diagnostic for comparing site-specific predictability and may inform targeted model development and observation design.

Keywords: Soil moisture, Predictability, Hydrological control, Machine learning, Stacking ensemble

1. Introduction

Soil Moisture (SM) is a key variable in the terrestrial water cycle, regulating land-atmosphere exchange through its control on the partitioning of energy and water fluxes (Seneviratne et al., 2010; Vereecken et al., 2022). Reliable SM forecasts at daily lead



times from 1 to 14 days can support early warning of agricultural droughts (Zhang et al., 2021), floods (Tramblay et al., 2021) and water scarcity (Krishnamurthy et al., 2022), as well as irrigation (Roy et al., 2021) and disaster-preparedness decisions (Krishnamurthy et al., 2022; Du et al., 2024). Operational SM forecasts commonly rely on land-surface or hydrological models driven by meteorological forecasts. Their skill can be substantially constrained by errors in forecast precipitation and other atmospheric forcings, particularly in the monsoon climates. In parallel, data-driven machine learning (ML) approaches have improved SM forecasting by extracting nonlinear dependencies from observational time series (Fang et al., 2019; Li et al., 2022a; Wang et al., 2024; Xiong et al., 2025; Katipoğlu et al., 2026). In autoregressive settings, these models can exploit information embedded in historical in-situ SM observations without requiring explicit forecasts of future meteorological inputs. In the present study, data-driven SM forecasting refers to a supervised machine-learning framework that uses historical SM observations and hydro-meteorological covariates available at the forecast initialization time, without requiring forecasts of future meteorological forcing. Here, we define data-driven SM predictability operationally as the out-of-sample forecast skill attainable using a standardized multi-model framework. Despite these advances, SM forecast skill is often strongly spatially heterogeneous, which nearby sites can exhibit markedly different lead-time-dependent skill even under broadly similar climatic conditions (Xu et al., 2025). Reported skill also varies with the model class, data product, and predictor set used (Draper et al., 2013; Kim et al., 2021; Kivi et al., 2023; Zha et al., 2025; Hendrickx et al., 2026). For example, in certain catchments, predictive skill remains resilient across extended lead times, while under similar climate conditions, performance degrades rapidly (Singla et al., 2012). This pronounced spatial heterogeneity can be observed both regionally (Song et al., 2021; Wang et al., 2025) and globally (Kim et al., 2021). This heterogeneity limits both hydrological interpretation and practical model development. Without understanding why predictability declines rapidly at some sites but persists at others, it remains difficult to determine where additional model complexity, alternative predictors, or process-aware strategies are most needed.

Current explanations of SM predictability have largely emphasized either external atmospheric forcing or SM memory. The former assumes that predictable precipitation and temperature anomalies can translate into predictable SM states, whereas the latter attributes predictability to the persistence of antecedent SM in the soil profile (Nicolai-Shaw et al., 2016; Kim et al., 2021). Soil texture and hydraulic properties influence this persistence by regulating infiltration, storage, drainage, and evapotranspiration (Saboori et al., 2025; Rahmati et al., 2024). Extreme precipitation can, however, rapidly alter these processes by inducing strong vertical redistribution of water through the profile, potentially weakening the temporal continuity on which autoregressive forecasts depend (Chen et al., 2025). Data quality and sensor uncertainty can further obscure these relationships at point scales (Kivi et al., 2023; Hendrickx et al., 2026). Information-theoretic analyses (Bennett et al., 2019), memory-based metrics (Rahmati et al., 2024), and error-decomposition approaches have each clarified part of the problem (Fang et al., 2018; Martínez-de la Torre and Miguez-Macho, 2019; Li et al., 2025; An et al., 2025), but they seldom link dynamic atmospheric disturbances, profile-scale soil properties, and lead-time-dependent forecast skill within one physically interpretable framework.

The unresolved question is therefore how local hydrological processes filter precipitation disturbances into either persistent or rapidly changing SM trajectories, and how this filtering shapes data-driven predictability. A quantitative link between readily



available soil and precipitation characteristics and the expected decay of forecast skill would help move beyond trial-and-error model development. Such a link would not replace site-specific forecast evaluation, but it could provide a candidate basis for pre-modeling assessment of forecastability if it proves transferable to independent locations. We hypothesize that soil physical properties and precipitation states interact nonlinearly to generate distinct hydrological regimes of SM predictability. Where the soil profile buffers incoming precipitation through matrix storage, subsequent moisture changes should remain relatively gradual and temporally persistent, allowing historical SM observations to retain predictive information over longer lead times. Conversely, where weak retention coincides with intense precipitation, rapid vertical redistribution and deeper SM responses may reduce SM memory and accelerate the loss of forecast skill. We therefore expect predictability to organize into regimes distinguished by the rate at which skill decays with lead time and by threshold combinations of soil-profile properties and precipitation disturbances.

To test this hypothesis, we analyze observations from 73 monitoring sites in Guangdong Province, China, at 10, 30, and 60 cm depths. We first use a stacking ensemble comprising 18 structurally diverse base learners to establish a consistent empirical benchmark of 1–14-day SM predictability and to characterize site-level error-growth trajectories. We then identify predictability regimes from similarity of these trajectories and test their physical consistency using model-agnostic information-entropy metrics and event-scale SM-response diagnostics. Next, we examine how the ensemble’s predictive behavior and the identified regimes relate to climatic, soil, and geographic characteristics, with particular attention to the joint effects of soil texture and extreme precipitation. Finally, we synthesize the resulting controls into a Fast-Decay Index (FDI) that summarizes hydrological conditions associated with rapid predictability loss. The FDI is presented as a physically interpretable candidate diagnostic for future pre-modeling assessment; its ability to screen forecastability at new sites requires independent-site validation. By linking profile-scale hydrological processes to spatial differences in data-driven SM predictability, this study provides a process-oriented basis for more targeted forecasting-model development and observation design.

2. Data and methods

2.1 Study area and data processing

2.1.1 Study area

Guangdong province (20°09'–25°31'N, 109°45'–117°20'E), located in southern China, spans a pronounced physiographic gradient from mountainous terrain in the north, through the Pearl River Delta (PRD), to low-lying coastal plains in the south (Fig. 1; Li et al., 2001). The region has a humid subtropical monsoon climate and experiences marked seasonal and frequent typhoon-related heavy precipitation, and recurrent seasonal droughts (Liu et al., 2001; Wang et al., 2020). These contrasts in topography, hydrological connectivity, precipitation variability, and soil conditions provide a suitable natural setting for examining how local hydrological characteristics regulate lead-time-dependent SM predictability.

The locations of the 73 SM monitoring sites are shown in Fig. 1d. To characterize the regional hydroclimatic setting, we used ERA5-Land reanalysis data to calculate the long-term standard deviations of air temperature, precipitation, and surface SM over 1950–2024 (Fig. 1a, b, and d, respectively). We additionally calculated the runoff coefficient for 2010–2023 as the ratio of accumulated total runoff to accumulated precipitation (Fig. 1c). The resulting maps illustrate substantial spatial heterogeneity in climatic variability, runoff-generation tendency, and SM variability across the province.



These gridded diagnostics are used to provide regional context for the station-scale analysis; the subsequent analyses test the proposed hydrological controls using in situ observations and site-specific soil and precipitation characteristics.

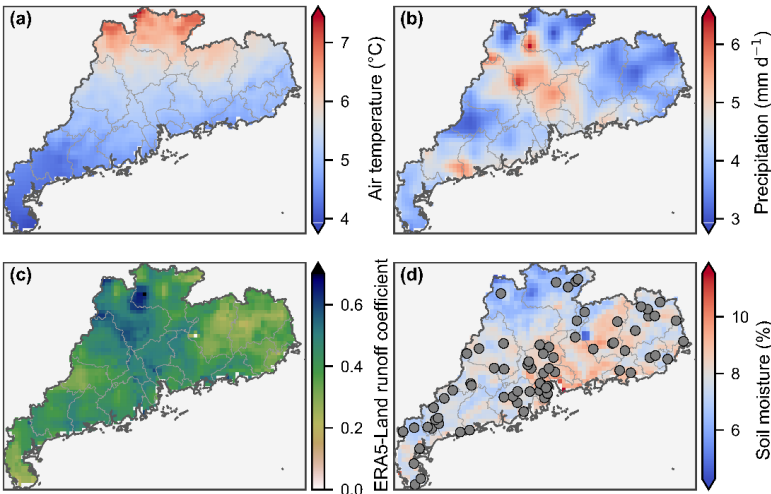


Figure 1. Study area and hydroclimatic setting of Guangdong Province. (a) Standard deviation of 2-m air temperature, (b) standard deviation of daily precipitation, and (d) standard deviation of surface soil moisture, calculated from ERA5-Land data for 1950–2024. (c) Runoff coefficient for 2010–2023, calculated as accumulated total runoff divided by accumulated precipitation (Liu et al., 2024). Black circles indicate the 73 soil-moisture monitoring sites.

2.1.2 In situ SM observational networks

This study utilized in situ volumetric SM data (unit: %) from 73 agricultural monitoring sites across Guangdong Province (Fig. 2d). The dataset integrates two complementary networks: 41 sites from the Guangdong Department of Agriculture’s observation system (2020–2023) and 31 sites from the China Meteorological Administration’s agrometeorological station network (2010–2020). All sites provide hourly SM measurements at eight depths: 10, 20, 30, 40, 50, 60, 80, and 100 cm. Among these, 10, 30, and 60 cm were selected as representative target depths for shallow-, intermediate-, and deep-layer SM forecasting, respectively. A summary of the dataset is provided in Table S1. Quality control protocols were applied to remove sensor artifacts: missing codes and null fields were systematically pruned, and any records showing absolute zero variance over a rolling 7-day window, indicative of data freeze or probe malfunction, were removed (Li et al., 2022b).

2.1.3 Meteorological and land surface database

We considered precipitation (m/day), relative humidity (%), 2-meter temperature (K), 2-meter dewpoint temperature (K), surface pressure (Pa), 10-meter northward and eastward wind components (m/s), potential evaporation (m), total evaporation (m), surface net solar radiation (J/m²), surface latent heat flux (J/m²), total runoff variable (mm/day), and forecast albedo (unitless) as predictors of ML models. They were sourced from the ERA5-Land reanalysis (Muñoz-Sabater et al., 2021). This product provides a globally consistent and spatially complete record at approximately 9 km resolution (0.1° × 0.1° grid). For each study site, daily time series were retrieved based



on the exact geographic coordinates and observation timestamps. For convenience, the Meteorological Variable (MV) input group (Table S2) included the above predictors. For an (h)-day forecast initialized on day (t), where ($h = 1, \dots, 14$), the target was the observed SM on day ($t+h$). All dynamic predictors were restricted to the preceding 7-day window from ($t-6$) to (t); no variables from days ($t+1$) to ($t+h$) were used. Day-of-year and the sequential day index were treated as deterministic calendar covariates known at the forecast initialization time. This design ensures that the reported skill represents genuine out-of-sample forecasting rather than retrospective prediction conditioned on future reanalysis forcing.

We utilized static variables as the candidate environmental factors for the SM sites. SoilGrids 250m global soil database (Poggio et al., 2021) was applied, including bulk density, silt content, sand content, and clay content at six depth intervals: 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm. These variables were extracted point-wise using the geographic coordinates of each monitoring site. For geographical variables, we used the record of altitude and precise location in the SM monitoring system.

2.2 Overall experimental design

Figure 2 summarizes a three-step framework for quantifying data-driven SM predictability, identifying reproducible lead-time-decay regimes, and examining their associations with local hydrological characteristics. The framework combines complementary model-based, time-series, event-scale, and environmental diagnostics rather than treating any single metric as a direct measure of hydrological causality.

Step 1: Establishing an empirical benchmark of SM predictability

We developed a stacking ensemble to quantify SM predictability at 1–14-day lead times and at 10, 30, and 60 cm depths. For each target depth and lead time, base learners used antecedent SM and hydro-meteorological covariates available at the forecast initialization time. Their out-of-fold predictions, together with the original physical predictors, were supplied to an XGBoost meta-learner. This leakage-free stacking design enabled the meta-learner to combine complementary predictive information across model classes. The resulting ensemble provides an empirical high-performance benchmark of data-driven SM predictability under the prescribed observational information set.

Step 2: Identifying and diagnosing predictability regimes

For each site, we characterized lead-time-dependent forecast degradation using its standardized 1–14-day normalized root-mean-square-error (NRMSE) trajectory. K-means clustering was then used to identify sites with rapid and gradual error growth, hereafter termed the fast-decay (FD) and slow-decay (SD) predictability regimes, respectively.

We assessed the physical consistency of the identified regimes using two complementary diagnostics. First, Sample Entropy (SampEn) was used to compare the temporal complexity of observed SM dynamics between FD and SD sites. Second, event-scale analyses quantified SM responses and forecast-skill changes following precipitation events. To examine model behaviour, SHAP values were calculated for held-out predictions and aggregated by regime, lead time, depth, and predictor group. These analyses reveal how the information used by the ensemble changes as predictability decays, rather than treating SHAP values as direct causal evidence.

Step 3: Linking predictability regimes to environmental characteristics

We evaluated the associations between predictability regimes and candidate environmental characteristics, including soil-profile properties, precipitation-extreme metrics, climatic variables, and geographical factors. Pre-specified subsets of these



variables were compared according to their ability to distinguish FD from SD sites using leave-one-site-out validation and the area under the receiver operating characteristic curve (ROC-AUC). This procedure identifies the environmental information that generalizes to sites not used to fit the classification relationship. We then used threshold analysis to identify joint soil-precipitation conditions associated with the FD regime. These conditions were synthesized into a Fast-Decay Index (FDI), which quantifies a site's continuous exposure to hydrological characteristics associated with rapid loss of forecastability. The FDI is interpreted as a physically informed candidate diagnostic for pre-modeling assessment, whose transferability is evaluated through the independent-site validation described above.

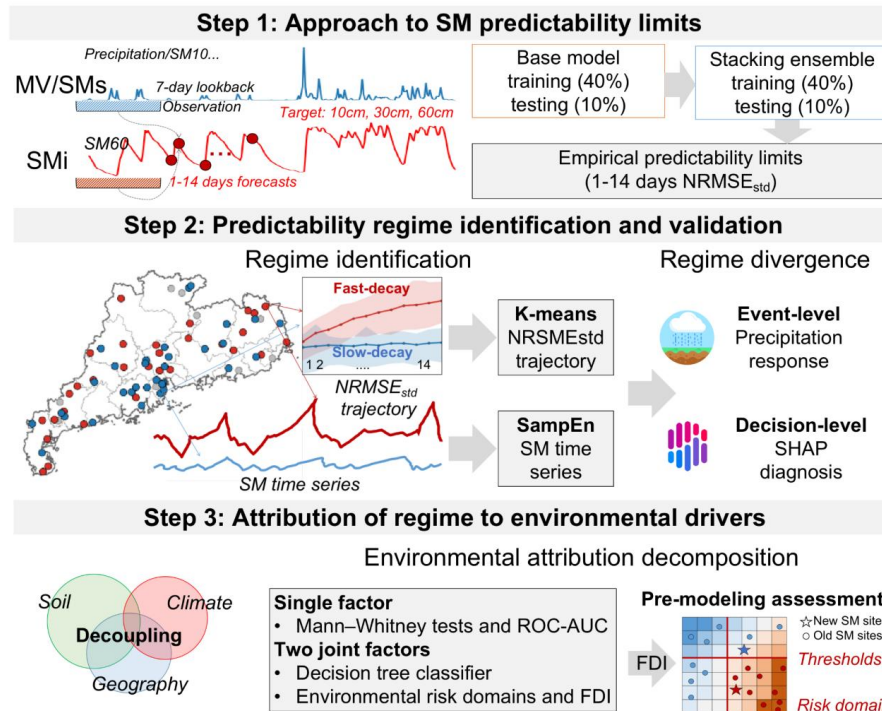


Figure 2 Integrated workflow for pre-modeling assessment and diagnostic attribution of soil moisture predictability. MV denotes meteorological variables; SMi denotes antecedent SM at the target depth; SMs denotes antecedent SM at other depths; NRMSEstd denotes RMSE normalized by the standard deviation of observations; SampEn denotes sample entropy; SHAP denotes Shapley additive explanations; FD denotes fast-decay regime; SD denotes slow-decay regime; FDI denotes fast-decay index; ROC-AUC denotes the area under the receiver operating characteristic curve.

2.3 Stacking ensemble forecasting

2.3.1 Direct forecasting task and base learners

For each site and target depth (10, 30, and 60 cm), we evaluated SM forecasts at lead times of 1-14 days. The predictor set at forecast initialization time t comprised the preceding 7-day histories ($t-6:t$) of: (i) SM at the target depth ($SM_{i<}$), (ii) SM at the other monitored depths (SMs), and (iii) dynamic hydro-meteorological and land-



surface covariates (MV). The prediction target was SM at $t + h$, where $h = 1, \dots, 14$. No SM or reanalysis-derived covariate from $t + 1$ to $t + h$ was supplied to any model.

We used a direct multi-horizon forecasting strategy. Each lead time was predicted from the same observed history, without using a forecast from an earlier lead time as an input to a later lead time. This design prevents the propagation of predicted-SM errors across the forecast horizon and quantifies the predictive information recoverable from the historical soil-water and hydro-meteorological state available at initialization. It does not assume that future atmospheric forcing is irrelevant; rather, it deliberately excludes unknown future forcing from the information set.

We constructed a library of 18 base learners spanning four algorithmic families: linear, instance-based, tree-based, and neural-network approaches. The complete list of models, feature preprocessing, and hyperparameter search spaces using grid search (LaValle et al., 2004) are given in Methods S1 and Table S3. Within each site-depth forecasting task, the first 40% of the chronological record was used for base-learner fitting and hyperparameter tuning. Tuning used expanding-window cross-validation, so that every validation fold followed its corresponding training fold in time. The subsequent 10% of the record was used for chronological validation and configuration selection. After selection, each base learner was refitted using the first 50% of the record.

2.3.2 Stacking ensemble optimization

The selected base learners were used to generate out-of-sample predictions for the subsequent 40% of each site record. These predictions, denoted by Preds, were combined with the original antecedent SM and MV predictors in an XGBoost meta-learner (Qian et al., 2025). Because the base learners used to generate Preds were fitted only on the preceding 50% of the record, the meta-learner never received in-sample base-learner predictions. This chronological blending design prevents target leakage during ensemble construction.

We compared pre-specified meta-learner feature configurations, including Preds alone and Preds combined with antecedent SM and MV predictors (Table S4). We also evaluated the best single base learner and a simple averaging ensemble as reference benchmarks (Methods S1). The XGBoost meta-learner was tuned by expanding-window cross-validation within its dedicated 40% training period. All transformations, including imputation, scaling, and feature selection where applicable, were fitted within the relevant training data and then applied unchanged to subsequent validation or test data. The final 10% of each chronological record was held out from all forecast-model training, tuning, and model-selection steps and was used only for final forecast evaluation. The resulting held-out skill profiles were subsequently used as outcomes in the separate cross-site regime and environmental analyses (Sections 2.4 and 2.5). The stacking ensemble is interpreted as an empirical high-performance benchmark of data-driven SM predictability under the prescribed historical information set and evaluated model library.

2.3.3 Forecast skill evaluation

Forecast performance was evaluated separately for every site, target depth, and lead time using the final held-out period. Root-mean-square error (RMSE) was the primary absolute-error metric. To compare errors across sites and depths with different SM variability, we used the standard-deviation-normalized RMSE ($\text{NRMSE}_{\text{std}}$):



$$NRMSE_{std} = \frac{1}{\sigma_{obs}} \times \left[\frac{\sum (y_{i,true} - y_{i,pred})^2}{n} \right]^{1/2} \quad (1)$$

where $y_{i,true}$ is the actual value of the i -th observation, $y_{i,pred}$ is the corresponding predicted value, and n is the sample size and σ_{obs} represents the standard deviation of the observations at a specific targeted depth.

Bias was also calculated with positive values denoting over-prediction. Range-normalized RMSE and mean absolute percentage error (MAPE) were reported as supplementary robustness metrics; because MAPE can become unstable under very low SM, it was not used to define predictability regimes. Their full definitions, together with complete site-, depth-, lead-, and model-specific score tables, are provided in Methods S2 and Tables S6-S8.

2.4 Profile-scale predictability regimes and process diagnostics

2.4.1 Unsupervised clustering sites for regime identification

We used $NRMSE_{std}$ as the core performance metric across all lead times as the quantitative predictability limits. This specific metric robustly represents the general predictability consistency reflected by alternative evaluation criteria (Fig. S4). For each site, we quantified the $NRMSE_{std}$ growth rate as the metric growth slope over the 1- to 14-day forecast horizon. To improve regime identification, we first normalized each $NRMSE_{std}$ trajectory using z-score transformation to remove the influence of absolute magnitudes, and then applied K-means clustering to classify regimes based on both the magnitude and slope of the trajectories across three depths.

The optimal number of clusters was identified as two, effectively capturing the dominant regimes of SM predictability (Fig. S5). The first cluster captures a slow-decay (SD) predictability regime, representing sites characterized by relatively stable error trajectories and limited skill loss as lead time increases. Conversely, the second cluster defines a fast-decay (FD) predictability regime, which comprises sites showing rapid error growth and a sharper decline in predictive skill across the 1-14 day horizon.

2.4.2 Temporal-complexity and event-scale diagnostics

Sample Entropy (SampEn) was calculated from the in situ SM series as a complementary diagnostic of temporal complexity. SampEn parameterization, record-length matching, and robustness analyses are described in Methods S3 and robustness tests are provided in Methods S3 and Fig. S16. We compared SampEn between the FD and SD regimes to assess whether their different forecast-skill trajectories were accompanied by systematic differences in the temporal organization of observed SM variability. SampEn was not treated as an independent causal validation of the clusters, but as convergent evidence on the character of local SM dynamics.

We further performed an event-scale soil-water response analysis to investigate whether FD and SD sites differed in their observed redistribution under precipitation (Wiekenkamp et al., 2016; Singh et al., 2021). A precipitation event was defined as one or more consecutive days with daily precipitation ≥ 1 mm/day; only events with cumulative precipitation ≥ 10 mm were retained. For each event, antecedent-to-peak storage changes were calculated from the event start to the event end plus a 2-day post-event window. Storage changes in the shallow (0–30 cm), deep (30–60 cm), and integrated (0–60 cm) layers were approximated from available in situ SM depths using thickness weighting. Two metrics were evaluated: the deep-layer response ratio (30–60 cm storage change / precipitation) and the vertical redistribution ratio (30–60 cm / 0–30 cm storage change).



2.4.3 Interpretation of ensemble behaviors

We used Shapley Additive Explanations (SHAP; Lundberg and Lee, 2017) to describe how the final XGBoost meta-learner used its input information in the two regimes. SHAP values were calculated only for the final held-out predictions and were summarized by regime, depth, lead time, and feature group. Feature groups included the base-model predictions, antecedent target-depth SM, cross-depth SM, and hydro-meteorological covariates. We report mean absolute SHAP values to characterize model reliance on each information source. These attributions describe the ensemble's predictive strategy; they are not interpreted as causal effects of individual predictors or as direct evidence that a predictor physically causes predictability loss.

2.5 Environmental associations and construction of the Fast-Decay Index

2.5.1 Candidate site descriptors

We characterized each site using pre-specified soil, hydroclimatic, geographic, and data-availability descriptors (Table 1). Soil matrix descriptors were extracted from SoilGrids 2.0 at the six standard depth intervals (0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm) (Poggio et al., 2021). Soil matrix was represented through both raw measurements, including sand, silt, and clay contents as well as bulk density at six standard depth intervals (0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm) (Poggio et al., 2021), and derived indices, including Thickness-Weighted Silt-dominant (TWS) fraction (Wankmüller et al., 2024), texture mutation (Li and Shao, 2013), field capacity (Bean et al., 2018; Kukal and Irmak, 2023), clay gradient (Obour et al., 2019), and bulk-density variability (Poggio et al., 2021). The TWS was calculated as:

$$TWS_i = \frac{\sum_{k=1}^K d_k I_{i,k}}{\sum_{k=1}^K d_k} \quad (2)$$

where i denotes the monitoring site, k denotes the SoilGrids depth interval and K is the total number of intervals; d_k is the thickness of layer k , and $I_{i,k}$ is an indicator variable equal to 1 when silt is the dominant texture fraction in layer k and 0 otherwise. In this study, $K=6$, corresponding to the 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm SoilGrids depths. Thus, TWS ranges from 0 to 1, with larger values indicating a greater thickness-weighted proportion of silt-dominant layers within the 0–200 cm soil profile. Climatic factors were summarized using statistical moments, including the mean, minimum, maximum, and standard deviation of key meteorological variables over the study period. Geographic context was defined by longitude, latitude, and altitude, capturing the topographic and coastal-inland gradients that reflect regional moisture fluxes and energy balance. Finally, data availability is defined as observation record length, which presents the total number of valid SM samples included to evaluate the impact of information volume on model performance. These descriptors were used to explain site-to-site variation in predictability regimes; they were not supplied as static predictors to the site-specific forecasting models.

Table 1. Candidate descriptors used to examine profile-scale soil-moisture predictability regimes.

Category	Examples	Roles in the analysis	Source
Soil matrix	Sand, silt, clay, bulk density, TWS, texture mutation, field capacity, clay gradient	Characterize water retention, storage, and vertical heterogeneity	SoilGrids 2.0; Poggio et al. (2021)
Climatic	Precipitation and	Characterize recurrent	ERA5-Land;



factors	temperature means, variability, and extremes	atmospheric disturbance and energy conditions	Muñoz-Sabater et al. (2021)
Geographic context	Latitude, longitude, altitude	Describe broad topographic and coastal-inland setting	Station metadata and DEM
Data availability	Record length	Describe the length of available data for training	This study

2.5.2 Environmental discrimination analysis

We first evaluated univariate associations between candidate descriptors and FD-SD membership using regime-wise distributional comparisons, Mann-Whitney U tests (Mann & Whitney, 1947), and Receiver Operating Characteristic-Area Under the Curve (ROC-AUC). The AUC scores are as defined:

$$AUC = \int_0^1 TPR(FPR) dFPR \quad (3)$$

where, $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FP}{FP+TN}$, and TP, FN, FP, and TN denote true positives, false negatives, false positives, and true negatives, respectively. Because the FD and SD regime sample sizes were not always equal across threshold-domain tests, BACC was used alongside AUC to account for class imbalance:

$$BACC = \frac{TPR+TNR}{2} \quad (4)$$

where $TNR = \frac{TN}{TN+FP}$ is the true-negative rate. BACC ranges from 0.5 to 1.

We next evaluated pre-specified low-dimensional combinations of soil and climatic descriptors using shallow decision-tree classifiers and class-weighted logistic regression. For decision tree, we set the model at the settings of maximum depth = min(3, number of input features) and minimum leaf size = 4, under stratified K-fold cross-validation, to ensure the small-sample conditions of this study (n = 73 sites).

Logistic regression was fitted after median imputation and standardization, with class weighting used to address class imbalance under the K-fold cross-validation to optimize the classifier. The AUC values were used to assess the ability of each factor combination to discriminate regime membership. Cross-validation results under both random and stratified K-fold configurations for all two-factor combinations are provided in Table S5. This K-fold cross-validation was conducted using different numbers of folds (<5) with the two-class count (Kohavi, 1995).

2.5.3 Threshold analysis and the Fast-Decay Index

We used threshold-based quadrant searches to identify the FD regime domains (Tyagi et al., 2022; Lisonbee et al., 2021). Candidate thresholds were drawn from observed unique values for discrete variables or from a 10th–90th percentile grid for continuous variables. Candidate domains were retained only when at least five training sites occurred both inside and outside the domain. The identified domains were validated based on three criteria: their BACC for two regimes, their statistical significance via Fisher’s exact test, and the physical divergence of their error-growth dynamics via group-wise statistical comparisons. Based on the identified FD regime domain, we defined two dimensionless threshold-relative margins:

$$u = \frac{|P_{max}-P_{thres}|}{sd(P_{max})}, \quad v = \frac{|TWS-TWS_{thres}|}{sd(TWS)} \quad (5)$$

where P_{max} and TWS is the site-level maximum daily precipitation and TWS, and



441 $sd(P_{max})$ and $sd(TWS)$ are the corresponding standard deviations estimated from the
442 training sites, and $P_{thres.}$ and $TWS_{thres.}$ are the identified thresholds. Positive values of
443 both margins indicate high precipitation disturbance and limited silt-dominant profile
444 buffering, respectively. We combined the margins using a smooth-minimum function
445 as:

$$446 \quad FDI = -\frac{1}{\beta} \ln \ln (e^{-\beta u} + e^{-\beta v}), \quad \beta = 1 \quad (6)$$

447 where β controls the smoothness of the minimum. The FDI is high only when both
448 required conditions are met, because it is governed primarily by the weaker of the two
449 margins. It is dimensionless and uses soil-profile and historical precipitation
450 information that can be assembled before fitting a site-specific SM forecasting model.
451 We therefore interpret it as a physically informed candidate for pre-modeling
452 assessment.

453

454 **3. Results**

455 **3.1 Forecast-skill benchmark and spatial heterogeneity of SM predictability**

456 On the final held-out period, the optimized stacking ensemble was the best-performing
457 configuration within the evaluated model library across the 1–14-day forecast horizon.
458 Fig. S6 shows that its advantage was retained across the East, PRD, North, and West
459 subregions (Fig. S3), while Fig. S7 compares it with the strongest single base learner
460 LightGBM (LGBM), a simple averaging ensemble, and alternative stacking-input
461 configurations. We therefore use the stacking ensemble as an empirical high-
462 performance benchmark for the subsequent analysis of site-level SM predictability.

463 The stacking ensemble consistently achieved lower $NRMSE_{std}$ than LGBM, the
464 strongest individual base learner, across target depths and lead times (Fig. 3a–c).
465 Relative reductions in $NRMSE_{std}$ at 1- and 7-day lead times were 47% and 22% at 10
466 cm, 30% and 21% at 30 cm, and 58% and 36% at 60 cm, respectively. The density
467 estimates also indicate that the stacking ensemble reproduced the observed SM
468 distribution more closely than LGBM at both lead times, including the central mode
469 and distributional spread. These gains demonstrate the value of combining
470 complementary model architectures and provide a consistent basis for comparing
471 forecast-skill trajectories among sites.

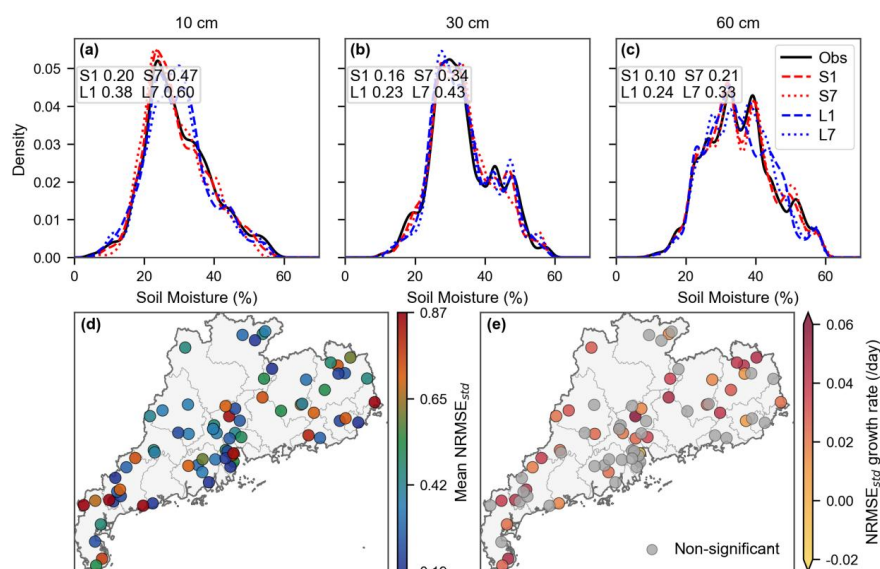


Figure 3. Soil moisture forecast skill of the stacking ensemble across depths and its spatial heterogeneity across Guangdong Province. (a–c) Kernel-density estimates of observed and forecast volumetric soil moisture at (a) 10 cm, (b) 30 cm, and (c) 60 cm. The distributions, constructed from the held-out samples, are shown for observations (Obs), the stacking ensemble (S), and the strongest single base learner, LightGBM (L), at 1- and 7-day lead times; for example, S7 denotes the stacking forecast at a 7-day lead time. Insets report the corresponding standard-deviation-normalized root-mean-square error (NRMSE_{std}). (d) Site-level mean NRMSE_{std} of the stacking ensemble, averaged over the three target depths and all 14 lead times. (e) Site-level NRMSE_{std} growth rate, calculated from the linear NRMSE_{std}–lead-time relationship after averaging across the three target depths. Grey circles denote sites with a nominal linear-trend p value ≥ 0.05.

However, spatial analysis of the stacking ensemble shows remarkable heterogeneity. We averaged NRMSE_{std} across all depths and lead times (Fig. 3d) and estimated the NRMSE_{std} growth rates for the three target depths (Fig. 3e). Sites in the northeast, characterized by low runoff fluctuation (Fig. 1c) and weak precipitation disturbances (Fig. 1d), showed relatively low mean NRMSE_{std} and high growth rates. Nevertheless, neither metric aligned systematically with regional temperature (Fig. 1a) or precipitation (Fig. 1b) disturbances. In addition, the mean values and the decay rate did not show comprehensive consistency (Fig. 3d-e). The resulting predictability landscape was spatially fragmented, with only weak correspondence to the broad regional patterns of air-temperature and precipitation variability shown in Fig. 1a-b. This descriptive comparison does not by itself identify a physical cause of the fragmentation. Instead, it motivates the profile-scale regime analysis and the subsequent tests of whether local soil properties and rainfall disturbances are associated with contrasting forecast-skill trajectories.

3.2 Distinct predictability regimes and their temporal-complexity signatures

We clustered the sites by K-means method based on the 1-14-day NRMSE_{std} trajectories, which combined three target depths and 14 lead times. Among the



candidate solutions with two to five clusters, the two-cluster solution had the highest silhouette score (Fig. S5). Fig. 4a shows the spatial distribution of recognized regimes and Fig. 4b visualizes the grouped time-wise $\text{NRMSE}_{\text{std}}$ trajectories into two regimes. The clustering partitioned the sites into two distinct regimes: fast-decay regime (FD, 31 sites) and slow-decay regime (SD, 32 sites). The two regimes were intermingled at fine spatial scales across the province, and the only exception was the PRD, where nearly all sites belonged to the high-predictability regime, implying that predictability is controlled by local land-surface and soil features rather than by large-scale climatic zones.

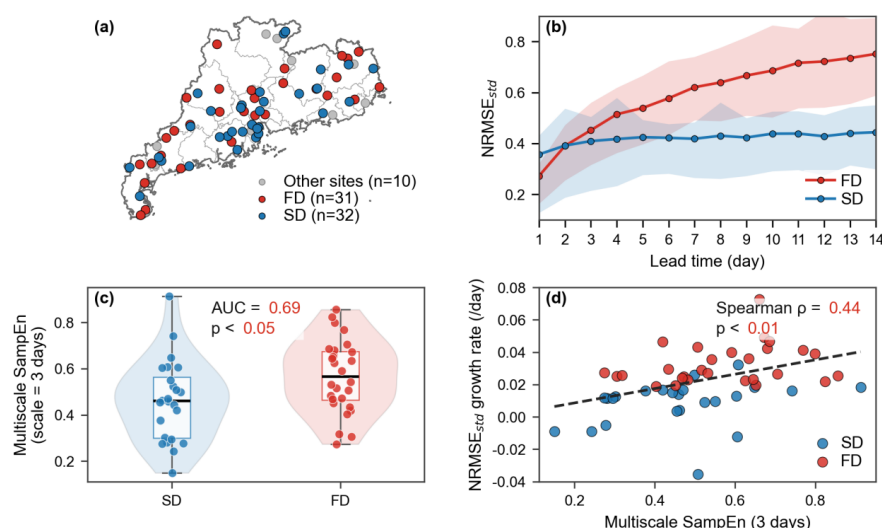


Figure 4 Predictability regime classification and temporal-complexity diagnostics. (a) Spatial distribution of fast-decay (FD; red) and slow-decay (SD; blue) sites across Guangdong Province. Grey points denote sites excluded from clustering because a complete 1–14-day held-out forecast-skill profile was unavailable. The three counts must sum to the total number of monitoring sites. (b) Regime-wise trajectories of depth-averaged standard-deviation-normalized root-mean-square error ($\text{NRMSE}_{\text{std}}$) across 1–14-day lead times. Solid lines show site means and shaded bands show interquartile ranges. (c) Distributions of multiscale SampEn (scale = 3 days) for FD and SD sites, with AUC and p-value from the Mann-Whitney U test. (d) Association between the site-level $\text{NRMSE}_{\text{std}}$ growth rate and multiscale SampEn at the 3-day scale. Only sites with multiscale SampEn available shown.

The observed SM series also showed a moderate, model-independent separation between the two regimes. SampEn was higher at FD than at SD sites (two-sided Mann–Whitney U test; ROC-AUC = 0.69; Fig. 4c), indicating that FD sites tended to have less temporally regular SM dynamics at the evaluated scale. The positive association between SampEn and the $\text{NRMSE}_{\text{std}}$ growth rate (Spearman's $\rho = 0.44$; Fig. 4d) further shows that sites with more complex observed SM trajectories tended to lose forecast skill more rapidly with lead time. These results confirm that the two predictability regimes are not model-dependent but reflect intrinsic differences in the underlying SM dynamics.



536 **3.3 Hydrological-response and ensemble-behaviour diagnostics across**
537 **predictability regimes**

538 **3.3.1 Contrasting event-scale soil-water responses to precipitation disturbances**

539 The two regimes exhibited contrasting model behaviour in response to precipitation
540 disturbances. Figs 5a and 5b present two representative sites that illustrate this contrast.
541 The Maoming and Yangjiang sites are geographically close and experienced
542 precipitation disturbances with similar timing and magnitude. Under similar
543 precipitation forcing, their contrasting soil-water responses can be attributed to
544 differences in soil hydraulic properties and profile structure. At the site with lower silt
545 dominance (Maoming; Fig. 5a), limited capillary buffering and higher unsaturated
546 hydraulic conductivity allow precipitation pulses to propagate rapidly into deeper layers,
547 producing a strong subsurface response. In contrast, the silt-rich profile at Yangjiang
548 (Fig. 5b) retains water within the soil matrix through stronger capillary forces, thereby
549 damping wetting-front propagation and preserving antecedent soil-moisture memory.

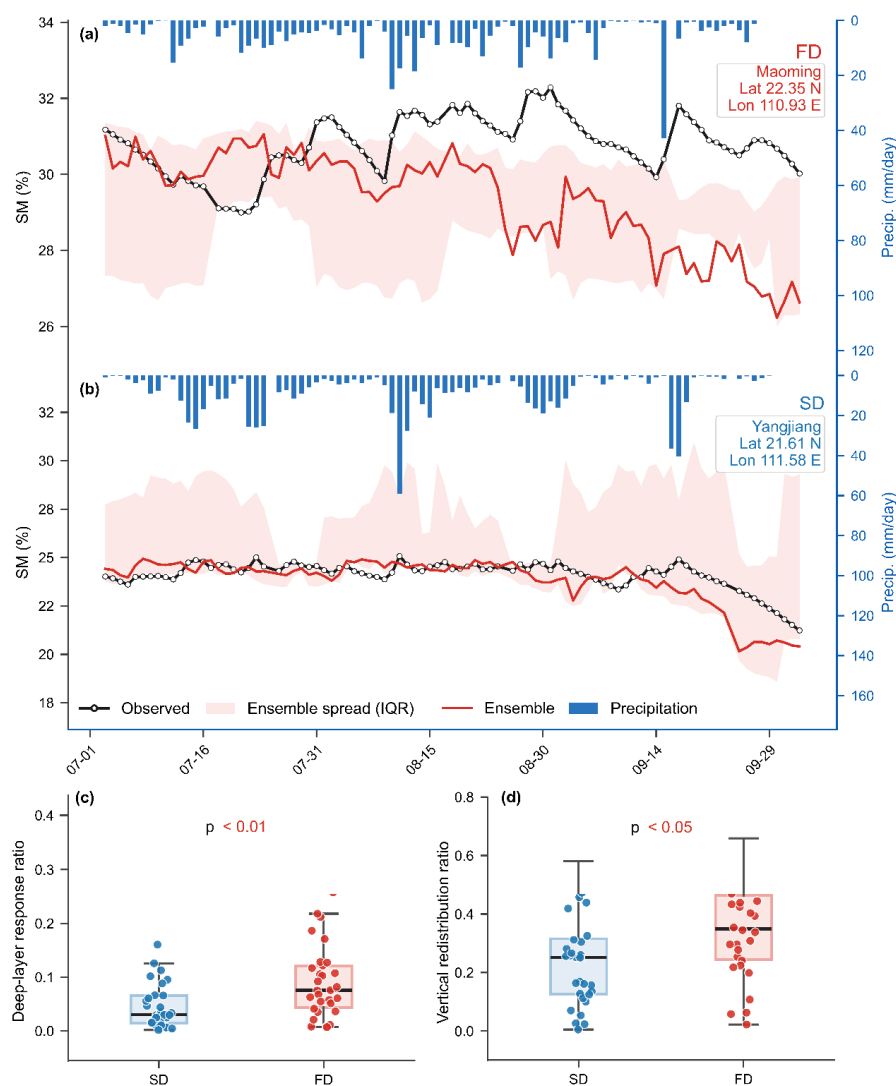


Figure 5. Contrasting observed soil-water responses and stacking-ensemble forecast behaviour during precipitation disturbances in FD and SD regimes. (a, b) Observed and stacking-ensemble-predicted SM at 30 cm depth for illustrative (a) FD (Maoming; 22.35°N, 110.93°E) and (b) SD (Yangjiang; 21.61°N, 111.58°E) sites during the final test period. The forecast lead time shown in these panels must be stated explicitly ($h = 14$ days). Daily precipitation (Precip., bars) is shown as observed diagnostic context. The red line denotes the stacking-ensemble forecast and the pink band denotes the interquartile range of the 18 base-learner predictions. (c, d) Box plots of the (c) deep-layer response ratio (30–60 cm storage change / precipitation) and (d) vertical redistribution ratio (30–60 cm / 0–30 cm storage change) at FD and SD sites; each point denotes a site, and p-values are from Mann–Whitney U tests between FD and SD regimes.



This divergence was also evident across the study domain, where the two regimes differed markedly in both the deep-layer response ratio and the vertical redistribution ratio (Figs. 5c and 5d). FD sites exhibited a significantly stronger deep-layer (30–60 cm) response (0.08 versus 0.03; $p = 0.005$; Fig. 5c) and a higher vertical redistribution ratio (0.35 versus 0.25; Fig. 5d). These results indicate that FD sites experience more pronounced vertical redistribution and deeper penetration of precipitation inputs, which may contribute to a more rapid loss of forecast skill following precipitation events. Supplementary Fig. S8 provides event-scale diagnostics of the deep-layer response and vertical redistribution ratios across all sites, showing consistent response patterns.

3.3.2 Divergent predictive strategies of the stacking meta-model

Figure 6 characterizes how the stacking meta-model allocated predictive attribution among its input features across the two predictability regimes. For each prediction at each site, the relative contribution of a predictor was calculated as the absolute value of its SHAP value divided by the sum of absolute SHAP values across all meta-model predictors. The resulting contributions therefore represent relative attribution shares under the fitted feature representation. Mean contributions of the three predictor groups including raw antecedent MV, Preds, and antecedent SM states ($SM_i + SM_s$) are shown in Figs. 6a, 6d, and 6g and summarized in Table S6.

Across all three target depths, the SD regime received a larger relative SHAP share from raw antecedent MV, whereas the FD regime received a larger share from base-model predictions. This difference should be interpreted conditionally: the base-model predictions were themselves generated from the same antecedent MV and SM information used as raw inputs to the meta-model. The relative contribution share of raw antecedent MV remained higher in the SD regime across the 1–14-day lead times, and the separation between regimes became more pronounced with increasing target depth (Figs. 6b, 6e, and 6h). Within the base-model-prediction group, the relative contribution share of the Bayesian-ridge (BY) prediction increased at FD sites, particularly at deeper depths, whereas the relative share of the LGBM prediction decreased with lead time (Figs. 6c, 6f, and 6i).

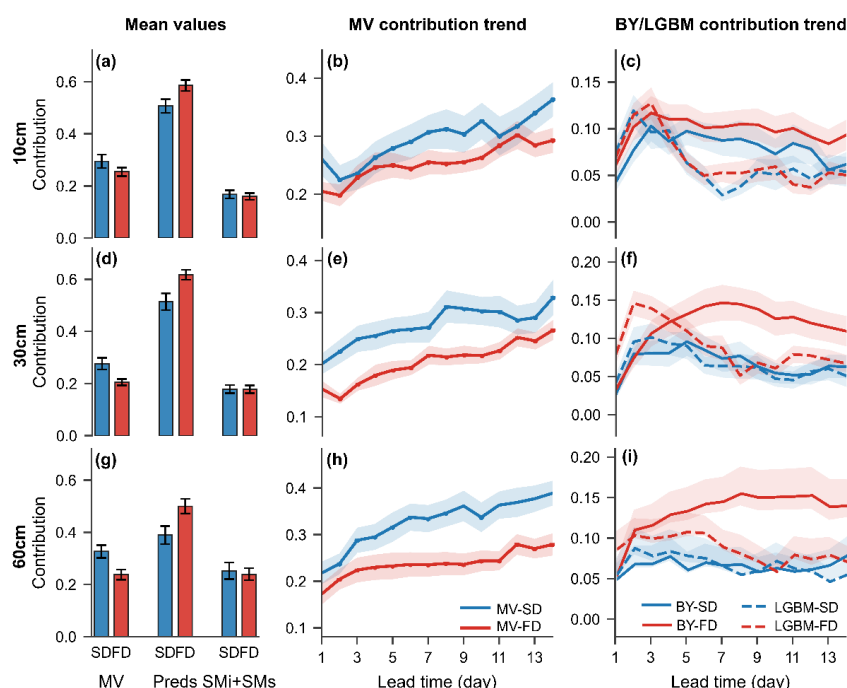


Figure 6 Regime-dependent predictor contribution patterns in the stacking meta-model diagnosed by SHAP. (a, d, g) Mean relative SHAP contribution proportions of raw antecedent meteorological variables (MV), base-model predictions (Preds), and antecedent SM states (SMi + SMs) for the FD and SD regimes at target depths of 10, 30, and 60 cm, respectively; error bars indicate standard errors across sites. Contribution proportions were calculated from normalized absolute SHAP values for each held-out prediction. (b, e, h) Lead-time evolution of the mean relative contribution proportion of MV for the FD and SD regimes at 10, 30, and 60 cm, respectively; shaded bands indicate interquartile ranges across sites. (c, f, i) Lead-time evolution of the relative contribution proportions assigned to the linear Bayesian-ridge (BY) and tree-based LightGBM (LGBM) base-learner predictions for the FD and SD regimes at 10, 30, and 60 cm, respectively; shaded bands indicate interquartile ranges across sites.

Overall, the results show regime-dependent attribution patterns in the stacking meta-model. At FD sites, a larger relative attribution share was assigned to base-model predictions, particularly to the BY prediction feature at greater depths, whereas raw antecedent MV received a smaller share. It might be because raw predictors and base-model outputs are correlated, these patterns describe conditional model attribution rather than independent contributions from distinct information sources. They are consistent with the contrasting hydrological responses identified above.

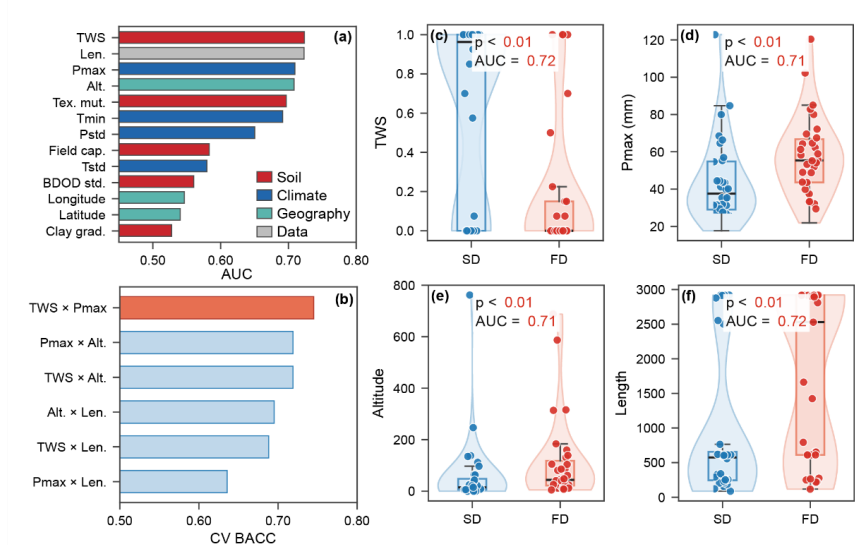
3.4 Joint environmental associations with predictability regimes

3.4.1 Single-variable screening and two-variable discrimination

Fig. 7a summarizes the ability of individual site descriptors to discriminate between the FD and SD regimes. Among the environmental descriptors, TWS was one of the strongest individual discriminators (AUC=0.72, $p<0.01$). Nevertheless, its distribution overlapped substantially between regimes, including at both low and high values of



623 TWS (Fig. 7c). TWS alone therefore did not provide a deterministic separation of the
624 two regimes. Record length (Len.) showed a similarly high AUC (0.72, $p < 0.01$), with
625 median record lengths of 2529 and 573 days in the two regimes (Fig. 7f). This result
626 identifies data availability as a potential confounder rather than a hydrological driver of
627 regime membership (Kratzert et al., 2019). Although sites with long records occurred
628 in both regimes, record length was retained in the subsequent sensitivity and adjusted
629 analyses. Among climatic descriptors, Pmax showed the strongest regime separation,
630 with a median of 55.4 mm/day at FD sites compared with 37.6 mm/day at SD sites
631 (AUC = 0.71, $p < 0.01$). Altitude also had moderate univariate discriminatory ability,
632 but its association with regime membership was no longer significant after accounting
633 for Pmax and TWS (partial Spearman's $\rho=0.19$; Table S7). Collectively, the overlap
634 among single-variable distributions indicates that no individual descriptor can
635 adequately distinguish the two regimes. Distributional comparisons for the remaining
636 candidate descriptors are provided in Fig. S9.



637 Figure 7. Single-variable and two-variable environmental discrimination of the FD and
638 SD predictability regimes. (a) Single-variable ROC-AUC values for candidate site
639 descriptors grouped into soil properties, hydroclimatic conditions, geography, and
640 data availability; FD was treated as the positive class. (b) Leave-one-site-out cross-validated
641 balanced accuracy (CV BACC) of the six best-performing two-variable combinations.
642 Variable-pair selection and classifier fitting were conducted using training sites only
643 within each fold. (c–f) Distributions of selected descriptors in the FD and SD regimes:
644 (c) thickness-weighted silt-dominant fraction (TWS), (d) maximum daily precipitation
645 (Pmax), (e) altitude, and (f) observation-record length. Individual sites are overlaid
646 as points. AUC values and Mann–Whitney U-test p values are reported for each descriptor.
647
648

649 Two-variable screening showed that combinations of TWS with precipitation statistics
650 consistently yielded the strongest cross-validated discrimination. In particular, the
651 combination of TWS and Pmax, and that of TWS and precipitation variability (Pstd),
652 achieved the highest leave-one-site-out cross-validated AUC and BACC among the
653 evaluated linear and decision-tree classifiers (Table S8). Variable pairs without TWS
654 generally performed less well, indicating that TWS was repeatedly included in the most



informative two-variable combinations. Restricting the comparison to pairs among the four leading individual descriptors (TWS, Pmax, altitude (Alt.), and Len.) produced the same ranking (Fig. 7b): the [TWS, Pmax] combination achieved the highest BACC (0.74), followed by [Pmax, altitude] and [TWS, altitude] (both 0.72), whereas combinations involving record length ranked lower. We therefore selected TWS and Pmax for threshold analysis because this pair combined the strongest internally cross-validated discrimination with a physically interpretable representation of soil-profile texture and precipitation extremes. Its association with regime membership remained after accounting for record length in sensitivity analyses (Table S9).

3.4.2 Cross-validated precipitation–soil domain and Fast-Decay Index

The nested threshold analysis identified a joint precipitation-soil domain characterized by higher Pmax and lower TWS (representative thresholds: Pmax > 28 mm/day and TWS < 0.7; Fig. S10). The observed proportion of FD sites was 73.5% within this domain, compared with 17.9% outside it (Fig. 8a). Many FD sites also had very low TWS values (<0.25). The inverse spatial pattern of silt and sand content shown in Fig. S11 provides regional context for this pattern, but does not establish an independent causal role of sand content. The cross-validated FDI was significantly higher at FD than at SD sites (Fig. 8b) and achieved an out-of-site ROC-AUC of 0.82. This performance exceeded that of either Pmax or TWS alone (AUC = 0.71 and 0.72, respectively). Cross-validated FDI was also positively associated with Day 1–14 error growth ($\rho = 0.46$; Fig. 8c), showing that stronger joint exposure to higher precipitation extremes and lower TWS corresponded to more rapid deterioration of forecast skill across held-out sites. Sites in the upper FDI tertile had a higher frequency of FD membership than those in the lower tertile (76.2% versus 14.3%; odds ratio = 19.2; Fisher's exact $p < 0.001$). The association between FDI and error growth remained significant after accounting for record length and altitude (Table S10).

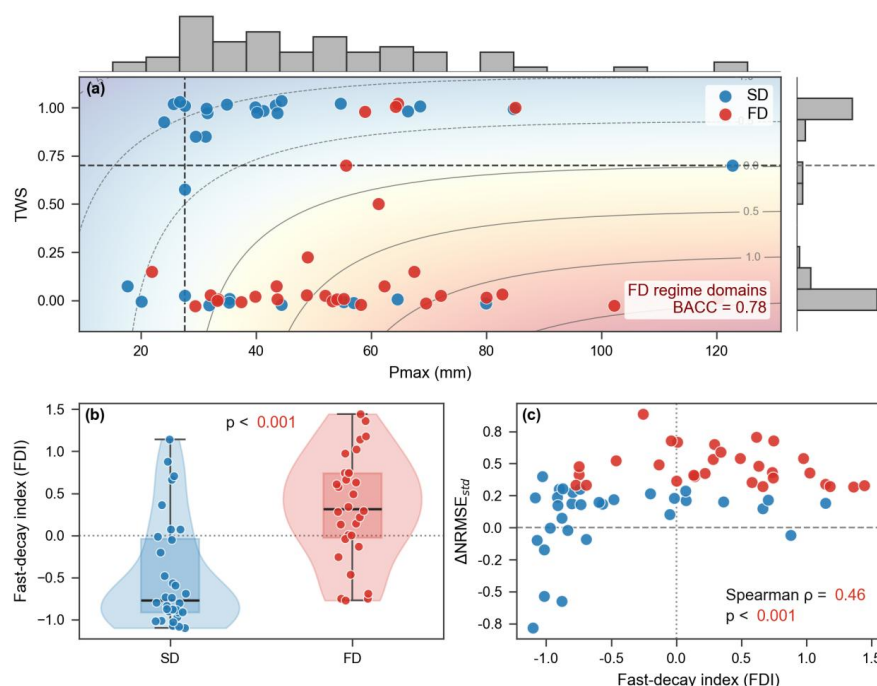


Figure 8. Cross-validated precipitation–soil threshold domain and Fast-Decay Index (FDI). (a) Two-variate distribution of Pmax and TWS for SD and FD sites. Dashed lines show the representative thresholds, Pmax=28 mm/day and TWS=0.7, identified by the nested threshold analysis. Background shading and contours show the continuous FDI field for visualization, and marginal histograms show pooled site distributions. All reported discrimination metrics and site-level FDI scores were calculated using thresholds selected from training sites only within each leave-one-site-out fold. (b) Distribution of cross-validated FDI scores for FD and SD sites; the p value is from a Mann–Whitney U test. (c) Relationship between cross-validated FDI and the increase in NRMSE_{std} from Day 1 to Day 14, across all sites.

Together, these results identify the FDI as a physically informed, internally cross-validated candidate indicator for pre-modeling assessment of SM predictability within the Guangdong monitoring network. Independent evaluation in a separate monitoring network or region is required before broader operational application.

4. Discussion

This study demonstrates that the spatial divergence in data-driven SM predictability is primarily dictated by localized hydrological characteristics, governed by the interaction between TWS and Pmax. Using a common forecasting design across the monitoring network, we identified two profile-scale predictability regimes with contrasting lead-time error-growth trajectories. Our findings highlight local hydrological processes as the primary control on forecast performance, rather than studies on ML architectures (Fang et al., 2017), feature selection (Saboori et al., 2025), or data quality (Jeung et al., 2026). Using site-specific stacking ensembles to establish empirical predictability benchmark (Fig. 3), we identified two distinct forecast skill regimes (Fig. 4) driven by



contrasting model behaviors during precipitation disturbances (Figs. 5-6). Environmental diagnostic analysis confirms that hydrological controls govern these regime dynamics (Fig. 7). Importantly, the joint threshold identified here enables a practical pre-modeling assessment tool to evaluate site suitability for data-driven forecasting (Fig. 8). The following sections discuss the scope of the empirical forecast-skill benchmark (Sect. 4.1), the robustness and hydrological interpretation of the identified regimes (Sect. 4.2), and the extent to which the empirical patterns are consistent with simulations from a physics-based SM model (Sect. 4.3).

4.1 Scope and reliability of the empirical forecast-skill benchmark

The stacking ensemble was designed to reduce dependence on a single architecture when estimating site-level forecast skill. It used meteorological variables (MV), target- and non-target-depth SM states (SM_i and SM_s), day of year, and a daily sequence index within a 7-day lookback window, consistent with previous SM forecasting studies (Pan et al., 2019; Li et al., 2022a; Xu et al., 2024; Wu et al., 2025). The base learners were selected to span broad categories of model architectures, ensuring that the ensemble captured a wide range of inductive biases. Linear models such as BY served as robust base learners and are known to perform well when SM dynamics follow smooth, quasi-linear trajectories (Pan et al., 2019). Tree-based ensembles such as RF and LGBM were also included for their ability to model nonlinear interactions and their strong empirical track record in SM prediction (Chen et al., 2024; Xu et al., 2024). The base learners spanned linear, tree-based, and deep-learning approaches. Bayesian ridge (BY) provides a robust linear baseline (Pan et al., 2019), whereas RF and LGBM represent nonlinear interactions (Chen et al., 2024; Xu et al., 2024). LSTM and an attention-augmented LSTM were included to capture temporal dependencies (Fang et al., 2019; Datta and Faroughi, 2023).

Within the evaluated library, greater architectural complexity did not necessarily improve held-out performance: the attention-augmented LSTM and EDLSTM did not outperform the standard LSTM (Fig. S12). We therefore used stacking to integrate diverse base learners. The XGBoost meta-model combined base-model predictions with the original antecedent predictors (Tahmouresi et al., 2024), and this augmented representation outperformed the strongest individual learner, a simple averaging ensemble, and alternative stacking-input configurations (Figs. S6 and S7). Although raw predictors and base-model outputs share the same historical information, the latter provide learned transformations that can complement the original inputs during precipitation disturbances (Figs. 5a–b).

The optimized ensemble achieved a site-averaged RMSE of 1.2% at Day 1 and 2.7% at Day 14 across target depths (Fig. S13), broadly comparable with recent 1–14-day SM forecasts based on ConvLSTM, hybrid deep-learning, and regional approaches (Xu et al., 2023; Li et al., 2025; Rabiei et al., 2025; Xiong et al., 2025). Direct cross-study comparisons remain limited by differences in data and evaluation protocols. Nevertheless, the ensemble consistently performed best among the tested configurations; its NRMSEstd trajectories (Fig. S4) therefore provide an empirical forecast-skill benchmark within the evaluated model and information space. Additional distributional and regime-stratified SHAP comparisons are shown in Figs. S14 and S15. Because future meteorological forcing was withheld, this benchmark reflects forecastability recoverable from antecedent states rather than the upper limit of an operational system using future weather forecasts.

4.2 Robustness and hydrological interpretation of the predictability regimes



The classification of regimes was robust to the main analytical choices. The two-cluster solution was supported by the clustering diagnostics (Fig. S5), and the regime assignments remained consistent when clustering was repeated using baseline-adjusted error-growth profiles. SampEn, calculated directly from the observed SM series, was higher at FD sites and positively associated with NRMSE_{std} growth rate (Fig. 4c-d); this pattern was robust across parameter settings (Fig. S16). The two approaches differ in both design (unsupervised versus model-agnostic) and basis (predictive performance versus SM dynamics). Their consistency indicates that the regime divergence reflects genuine differences in local soil water behavior, not methodological choices.

The event-scale analysis provides a hydrological interpretation of this distinction. FD sites showed stronger deep-layer responses and higher vertical redistribution ratios following precipitation events (Fig. 5c-d), consistent with more rapid redistribution of rainfall inputs and a shorter recoverable memory in the antecedent 7-day history. Similar links between soil texture, drainage, and SM memory have been reported previously (Martínez-Fernández et al., 2021; McColl et al., 2017). SHAP-based interpretations (Fig. 6) further confirm that this weak hydrological coupling introduces severe inter-event non-stationarity in the meteorological forcing-SM relationship. As a result, the meta-model struggles to learn stable diagnostic features at FD sites; the frequent disruption of nonlinear input-state trajectories forces the model to default to robust linear base-learners (BY) to mitigate error growth, a strategy that becomes paramount at deeper layers and longer lead times where dynamics mimic stochastic processes, consistent with the current findings (Pal et al., 2016). In sharp contrast, the intact physical continuity and memory retention at SD sites enable the persistent exploitation of nonlinear architectures (LGBM) for forcing-state mapping, securing high-fidelity predictability.

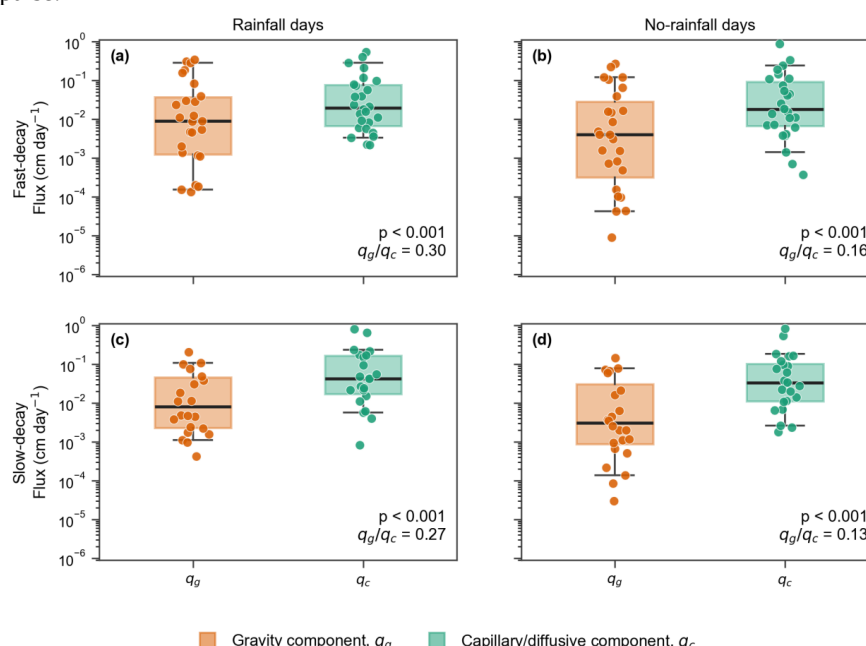
The predictability regimes identified from the joint effect of TWS and P_{max} proved robust under multiple lines of evaluation. The environmental controls from single-factor to two-factor combinations analysis were based on consistent results emerging across independent validation methods (Tables S6-S8). Notably, data record length, which previous studies have associated with predictability limits (Jeung et al., 2026), could not fully account for the regime divergence in our study domain (Fig. 7; Fig. S17). In total, the convergence of these empirical diagnostics cross-validates the robust physical foundation of our regime framework. Despite spatial sampling limitations arising from the predominance of loam-textured soils across the monitoring network (Fig. S11), the emergent trend linking joint soil-precipitation thresholds to predictability degradation exhibits strict consistency. The strong significant association between the FDI and NRMSE_{std} (Fig. 8c) indicates that proximity to the joint precipitation-soil FD domain corresponds to a continuous increase in forecast-error growth.

4.3 Process-model consistency with localized hydrological controls on SM predictability

To assess whether the observed regime differences are consistent with physically simulated soil-water dynamics, we solved the one-dimensional Richards equation at each site and diagnosed the partitioning of vertical fluxes under contrasting meteorological conditions (Methods S1). The Richards equation is used here as process-based diagnosis. We decomposed the simulated vertical flux into gravitational (q_g) and capillary/diffusive (q_c) components and compared their relative magnitudes on rainfall and no-rainfall days. A higher gravity-to-capillary flux magnitude ratio (q_g/q_c) indicates a larger relative role of gravity-driven downward transport, where as a



809 lower ratio indicates stronger control by matric-potential gradients.
 810 Fig. 9 shows contrasting simulated flux partitioning between the two regimes. On
 811 precipitation days, FD sites exhibited a substantially higher q_g/q_c than SD sites, whereas
 812 the difference was smaller under no-precipitation conditions. In contrast,
 813 capillary/diffusive fluxes remained relatively more important at SD sites under both
 814 weather conditions. These simulated patterns are consistent with the event-scale
 815 observations: FD sites showed stronger deep-layer responses and vertical redistribution
 816 following precipitation (Fig. 5c-d), while SD sites retained a more gradual soil-water
 817 response. By contrast, FD-domain sites, characterized by lower TWS and higher Pmax,
 818 showed a marked hydrodynamic transition during rainfall. Under no-rainfall conditions,
 819 their flux partitioning was broadly similar to that of SD sites. During intense rainfall,
 820 however, the q_g/q_c increased substantially above its baseline level and exceeded the
 821 response at SD sites. In these low-silt profiles, strong boundary forcing can rapidly
 822 increase soil wetness, weaken the matric-potential gradient, and increase the relative
 823 contribution of gravity-driven downward transport. Capillary control consequently
 824 becomes less dominant, and infiltration shifts towards a gravity-dominated downward
 825 pulse.



826 **Figure 9. Partitioning of gravity-driven and capillary-diffusive fluxes reveals the**
 827 **hydrodynamic mechanism underlying predictability-regime divergence.** (a, b)
 828 Fast-decay (FD) and (c, d) slow-decay (SD) regime flux magnitude (cm day⁻¹; log scale)
 829 of the gravity component (q_g , orange) and the capillary/diffusive component (q_c , teal)
 830 of the Richards equation on (a, c) precipitation days and (b, d) no-precipitation days. P-
 831 value, and the median gravity-to-diffusion ratio (q_g/q_c) are annotated in each panel.
 832
 833

834 This rainfall response can reduce the persistence of antecedent SM by rapidly
 835 redistributing water through the profile and decoupling post-event states from pre-event
 836 trajectories. The resulting inter-event non-stationarity is consistent with the meta-model
 837 assigning greater relative importance to robust linear base-learner predictions and less



to nonlinear LGBM predictions at FD sites (Fig. 6). Together, the Richards-equation diagnostics and the data-driven analyses suggest that the observed forecast-skill divergence is strongly associated with threshold-like precipitation-soil controls on vertical infiltration and SM-memory persistence, rather than arising solely from deficiencies of a particular statistical model.

4.4 Limitations

Several limitations should be considered. First, the study is restricted to 73 stations in a humid subtropical region of southern China; the TWS-Pmax thresholds and FDI therefore require independent evaluation in arid, temperate, and other hydroclimatic settings. Soil properties were extracted pointwise from SoilGrids and may not represent sub-grid heterogeneity at individual monitoring sites. The identified regimes should also be tested with additional monitoring networks and forecasting architectures. Moreover, because the present forecasts exclude meteorological forcing after initialization, the conclusions apply to predictability from antecedent information and may differ for operational systems using numerical weather predictions.

Second, the inferred redistribution mechanism is based on event-scale SM responses and Richards-equation simulations, rather than direct observations of infiltration rates, hydraulic gradients, or deep drainage. Independent hydraulic measurements or tracer-based field experiments would provide stronger tests of the proposed rainfall-response mechanism. Vegetation dynamics (Seneviratne et al., 2010), irrigation (Zha et al., 2025), and land-use change (te Wierik et al., 2021) may also contribute to unexplained site-level variability.

Finally, record length differed systematically between regimes and may affect model training and the precision of skill estimates. Although adjusted and stratified analyses indicated that record length did not fully explain the TWS-Pmax association, they cannot eliminate this confounding effect. Equal-length resampling and network-specific analyses are therefore necessary to isolate data-availability effects more comprehensively.

4.5 Implications

Within the Guangdong monitoring network, the FDI provides a candidate pre-modeling screening diagnostic based on site descriptors that can be assembled before fitting a site-specific SM forecasting model. It identifies soil–precipitation conditions associated with faster forecast-skill decay, rather than guaranteeing poor or high skill at an individual site. Independent validation in other networks and hydroclimatic regions is required before operational use.

The results nevertheless suggest a route towards regime-aware model development. Rather than prescribing a single optimal architecture, the FDI could guide targeted comparisons of model configurations, uncertainty treatment, and observation requirements across contrasting soil–precipitation settings. In data-sparse regions, monitoring design could similarly complement geometric spacing by ensuring coverage of key environmental gradients associated with different soil-water responses (Zhu et al., 2020).

More broadly, the study illustrates the potential of process-aware machine learning, in which physically interpretable landscape descriptors guide model development and evaluation (Reichstein et al., 2019; Slater et al., 2023). Such information can reduce purely trial-and-error model selection, but it does not replace empirical validation of forecasting performance at new sites.



Conclusions

This study assessed the spatial heterogeneity of data-driven SM predictability across 73 monitoring sites in Guangdong Province. Using a site-specific stacking ensemble as an empirical forecast-skill benchmark, we identified fast- and slow-decay regimes with contrasting lead-time error-growth trajectories. Higher sample entropy at FD sites provided complementary evidence of less temporally regular SM dynamics. Nested leave-one-site-out analyses showed that FD membership was associated with the joint occurrence of lower thickness-weighted silt dominance ($TWS < 0.7$) and higher maximum daily precipitation ($P_{max} > 28$ mm/day). Event-scale observations and Richards-equation diagnostics were consistent with this association: FD-domain sites showed stronger gravity-relative vertical transport during rainfall, deeper redistribution, and more rapid loss of antecedent SM persistence. These conditions were also accompanied by a greater relative contribution of linear base-learner predictions in the stacking meta-model. The resulting FDI provides an internally cross-validated candidate diagnostic for pre-modeling assessment within the Guangdong monitoring network. It can help identify sites where forecasts based on antecedent information are more likely to lose skill rapidly and where additional process representation, observations, or uncertainty treatment may be valuable. Independent validation across networks and hydroclimatic regions is needed before broader operational use. Overall, the study provides a process-aware framework for linking local soil–precipitation conditions to data-driven SM forecastability.

Acknowledgments

The study was partially supported by the National Natural Science Foundation of China (42375144, 4227515 and 42205149), Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM902), Guangdong Basic and Applied Basic Research Foundation (2021B0301030007) and China Meteorological Administration Youth Innovation Team (CMA2024QN01). F.H. acknowledges a PhD scholarship from the China Scholarship Council (202306380183). F.H., G.C. and M.P. acknowledge the Horizon project AI4PEX (grant agreement 101137682). Z.L. was supported by the China Scholarship Council (CSC, No. 202506350059).

References

- Ahmad, S. K., Kumar, S. V., Draper, C., and Reichle, R. H.: Challenges in unifying physically based and machine learning simulations through differentiable modeling: A land surface case study, *Geophys. Res. Lett.*, 52, e2024GL112893, <https://doi.org/10.1029/2024GL112893>, 2025.
- Bean, E. Z., Huffaker, R. G., and Migliaccio, K. W.: Estimating field capacity from volumetric soil water content time series using automated processing algorithms, *Vadose Zone J.*, 17, 180073, <https://doi.org/10.2136/vzj2018.04.0073>, 2018.
- Bennett, A., Nijssen, B., Ou, G., Clark, M. P., and Nearing, G. S.: Quantifying process connectivity with transfer entropy in hydrologic models, *Water Resour. Res.*, 55, 4613–4629, <https://doi.org/10.1029/2018WR024555>, 2019.
- Beven, K. and Germann, P.: Macropores and water flow in soils revisited, *Water Resour. Res.*, 49, 3071–3092, <https://doi.org/10.1002/wrcr.20156>, 2013.
- Breiman, L.: Stacked regressions, *Mach. Learn.*, 24, 49–64, <https://doi.org/10.1007/BF00117832>, 1996.



- 936 Brooks, J. R., Barnard, H. R., Coulombe, R., and McDonnell, J. J.: Ecohydrologic
937 separation of water between trees and streams in a Mediterranean climate, *Nat.*
938 *Geosci.*, 3, 100-104, <https://doi.org/10.1038/ngeo722>, 2010.
- 939 Chen, P.-Y., Chen, C.-C., Kang, C., Liu, J.-W., and Li, Y.-H.: Soil water content
940 prediction across seasons using random forest based on precipitation-related data,
941 *Comput. Electron. Agric.*, 230, 109802,
942 <https://doi.org/10.1016/j.compag.2024.109802>, 2025.
- 943 Chou, C. M.: Applying multiscale entropy to the complexity analysis of rainfall-
944 runoff relationships, *Entropy*, 14, 945-957, <https://doi.org/10.3390/e14050945>, 2012.
- 945 Chrysanthopoulos, E. and Kallioras, A.: Temporal and geographic extrapolation of
946 soil moisture using machine learning algorithms, *Catena*, 257, 109156,
947 <https://doi.org/10.1016/j.catena.2025.109156>, 2025.
- 948 Costa, M., Goldberger, A. L., and Peng, C.-K.: Multiscale entropy analysis of
949 complex physiologic time series, *Phys. Rev. Lett.*, 89, 068102,
950 <https://doi.org/10.1103/PhysRevLett.89.068102>, 2002.
- 951 Costa, M., Goldberger, A. L., and Peng, C.-K.: Multiscale entropy analysis of
952 biological signals, *Phys. Rev. E*, 71, 021906,
953 <https://doi.org/10.1103/PhysRevE.71.021906>, 2005.
- 954 Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G.,
955 Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C.
956 A., and Yang, Z.-L.: The Common Land Model, *Bull. Am. Meteorol. Soc.*, 84, 1013-
957 1024, <https://doi.org/10.1175/BAMS-84-8-1013>, 2003.
- 958 Datta, P. and Faroughi, S. A.: A multihead LSTM technique for prognostic prediction
959 of soil moisture, *Geoderma*, 433, 116452,
960 <https://doi.org/10.1016/j.geoderma.2023.116452>, 2023.
- 961 Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L.,
962 Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl,
963 G., Bogaen, H., Brocca, L., Calvet, J.-C., Camarero, J. J., Capello, G., Choi, M.,
964 Cosh, M. C., van de Giesen, N., Hajdu, I., Ikonen, J., Jensen, K. H., Kanniah, K. D.,
965 de Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A.,
966 Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J.
967 P., Osenga, E., Palecki, M. A., Pellarin, T., Petropoulos, G. P., Pfeil, I., Powers, J.,
968 Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T.,
969 Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J. P.,
970 Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van
971 Oevelen, P., Wagner, W., Scipal, K., Drusch, M., and Sabia, R.: The International
972 Soil Moisture Network: serving Earth system science for over a decade, *Hydrol. Earth*
973 *Syst. Sci.*, 25, 5749-5804, <https://doi.org/10.5194/hess-25-5749-2021>, 2021.
- 974 Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., and Wagner, W.:
975 Estimating root mean square errors in remotely sensed soil moisture over continental
976 scale domains, *Remote Sens. Environ.*, 137, 288-298,
977 <https://doi.org/10.1016/j.rse.2013.06.013>, 2013.
- 978 Du, J., Kimball, J. S., Jencso, K., He, L., Jones, L. A., Reichle, R. H., and Schaefer,
979 G. L.: Machine-learning based multi-layer soil moisture forecasts, *Water Resour.*
980 *Res.*, 60, e2023WR036973, <https://doi.org/10.1029/2023WR036973>, 2024.
- 981 Esit, M., Kumar, S., Pandey, A., Lawrence, D. M., Rangwala, I., and Yeager, S.:
982 Seasonal to multi-year soil moisture drought forecasting, *npj Clim. Atmos. Sci.*, 4, 16,
983 <https://doi.org/10.1038/s41612-021-00172-z>, 2021.
- 984 Fang, K., Shen, C., Kifer, D., and Yang, X.: Prolongation of SMAP to
985 spatiotemporally seamless coverage of continental U.S. using a deep learning neural



- 986 network, *Geophys. Res. Lett.*, 44, 11030-11039,
987 <https://doi.org/10.1002/2017GL075619>, 2017.
- 988 Fang, K., Pan, M., and Shen, C.: The value of SMAP for long-term soil moisture
989 estimation with the help of deep learning, *IEEE Trans. Geosci. Remote Sens.*, 57,
990 2221-2233, <https://doi.org/10.1109/TGRS.2018.2872131>, 2019.
- 991 Ford, T. W. and Quiring, S. M.: Comparison of contemporary in situ, model, and
992 satellite remote sensing soil moisture with a focus on drought monitoring, *Water*
993 *Resour. Res.*, 55, 1565-1582, <https://doi.org/10.1029/2018WR024039>, 2019.
- 994 Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Posselt, D.
995 J., Stephens, G. L., van den Heever, S. C., Varble, A. C., and Zuidema, P.: The future
996 of earth system prediction: Advances in model-data fusion, *Sci. Adv.*, 8, eabn3488,
997 <https://doi.org/10.1126/sciadv.abn3488>, 2022.
- 998 Hendrickx, M. G. A., Vanderborght, J., Janssens, P., Laloy, E., Bombeke, S.,
999 Matthyssen, E., Waverijn, A., and Diels, J.: Field-scale soil moisture predictions in
1000 real time using in situ sensor measurements in an inverse modeling framework:
1001 SWIM2, *Water Resour. Res.*, 62, e2025WR041324,
1002 <https://doi.org/10.1029/2025WR041324>, 2026.
- 1003 Huang, X., Shi, Z. H., Zhu, H. D., Zhang, H. Y., Ai, L., and Yin, W.: Soil moisture
1004 dynamics within soil profiles and associated environmental controls, *Catena*, 136,
1005 189-196, <https://doi.org/10.1016/j.catena.2015.01.014>, 2016.
- 1006 Jeung, M., Her, Y., Baek, S.-S., and Yoon, K.: Sensitivity of hydrological machine
1007 learning prediction accuracy to information quantity and quality, *Hydrol. Earth Syst.*
1008 *Sci.*, 30, 1077-1095, <https://doi.org/10.5194/hess-30-1077-2026>, 2026.
- 1009 Katipoğlu, O. M., Kartal, V., Akiner, M. E., Terzioğlu, Z. Ö., and Kılıç, Z.: Prediction
1010 of soil moisture via feature selection, model optimization, and climate data
1011 integration, *Phys. Chem. Earth*, 142, 104250,
1012 <https://doi.org/10.1016/j.pce.2025.104250>, 2026.
- 1013 Kivi, M., Vergopolan, N., and Dokoohaki, H.: A comprehensive assessment of in situ
1014 and remote sensing soil moisture data assimilation in the APSIM model for improving
1015 agricultural forecasting across the US Midwest, *Hydrol. Earth Syst. Sci.*, 27, 1173-
1016 1199, <https://doi.org/10.5194/hess-27-1173-2023>, 2023.
- 1017 Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and
1018 model selection, *Proceedings of the 14th International Joint Conference on Artificial*
1019 *Intelligence (IJCAI)*, 2, 1137-1143, 1995.
- 1020 Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.:
1021 Towards learning universal, regional, and local hydrological behaviors via machine
1022 learning applied to large-sample datasets, *Hydrol. Earth Syst. Sci.*, 23, 5089-5110,
1023 <https://doi.org/10.5194/hess-23-5089-2019>, 2019.
- 1024 Krishnamurthy, P. K., Fisher, J., Choularton, R., and Kareiva, P.: Anticipating
1025 drought-related food security changes, *Nat. Sustain.*, 5, 956-964,
1026 <https://doi.org/10.1038/s41893-022-00962-0>, 2022.
- 1027 Kukal, M. S. and Irmak, S.: Can limits of plant available water be inferred from soil
1028 moisture distributions?, *Agric. Environ. Lett.*, 8, e20113,
1029 <https://doi.org/10.1002/ael2.20113>, 2023.
- 1030 LaValle, S. M., Branicky, M. S., and Lindemann, S. R.: On the relationship between
1031 classical grid search and probabilistic roadmaps, *Int. J. Robot. Res.*, 23, 673-692,
1032 <https://doi.org/10.1177/0278364904045318>, 2004.
- 1033 Li, C., Lei, Y., He, W., and Dai, Z.: Land-ocean interaction in modern delta formation
1034 and development: A case study of the Pearl River Delta, China, *Sci. China Ser. B*, 44,
1035 63-71, 2001.



- 1036 Li, B., Sprenger, M., Wyatt, B. M., Giménez, D., Hirmas, D. R., Ajami, H.,
1037 Wiekenkamp, I., Groh, J., Nimmo, J. R., Amato, M. T., Singh, N. K., Crompton, O.,
1038 Araki, R., Xu, T., and Sullivan, P. L.: Ubiquity and causes of soil water preferential
1039 flow across 17 ecoregions, *Geophys. Res. Lett.*, 52, e2025GL118045,
1040 <https://doi.org/10.1029/2025GL118045>, 2025.
1041 Li, D. and Shao, M.: Simulating the vertical transition of soil textural layers in north-
1042 western China with a Markov chain model, *Soil Res.*, 51, 182-192,
1043 <https://doi.org/10.1071/SR12332>, 2013.
1044 Li, Q., Li, Z., Shangguan, W., Wang, X., Li, L., and Yu, F.: Improving soil moisture
1045 prediction using a novel encoder-decoder model with residual learning, *Comput.*
1046 *Electron. Agric.*, 195, 106816, <https://doi.org/10.1016/j.compag.2022.106816>, 2022a.
1047 Li, Q., Shi, G., Shangguan, W., Nourani, V., Li, J., Li, L., Huang, F., Zhang, Y.,
1048 Wang, C., Wang, D., Qiu, J., Lu, X., and Dai, Y.: A 1 km daily soil moisture dataset
1049 over China using in situ measurement and machine learning, *Earth Syst. Sci. Data*, 14,
1050 5267-5286, <https://doi.org/10.5194/essd-14-5267-2022>, 2022b.
1051 Li, X., Zhong, Z., Wang, J., Li, Q., Zhou, X., Yan, S., Zhu, J., and Chen, X.:
1052 Advancing agricultural drought level prediction in Guangdong utilizing ERA5-Land
1053 and SMAP-L3 data, *Water*, 17, 3564, <https://doi.org/10.3390/w17243564>, 2025.
1054 Lisonbee, J., Woloszyn, M., and Skumanich, M.: Making sense of flash drought:
1055 Definitions, indicators, and where we go from here, *J. Appl. Serv. Climatol.*, 2021, 1-
1056 19, <https://doi.org/10.46275/JOASC.2021.02.001>, 2021.
1057 Liu, K.-b., Shen, C., and Louie, K.-s.: A 1,000-year history of typhoon landfalls in
1058 Guangdong, Southern China, reconstructed from Chinese historical documentary
1059 records, *Ann. Assoc. Am. Geogr.*, 91, 453-464, [https://doi.org/10.1111/0004-](https://doi.org/10.1111/0004-5608.00253)
1060 [5608.00253](https://doi.org/10.1111/0004-5608.00253), 2001.
1061 Liu, L., Yi, Y., Jiang, H., Ran, Y., and Chen, D.: ERA5-Land overestimates runoff
1062 coefficient but underestimates runoff recession rate in the central Tibetan permafrost
1063 region, *J. Hydrol. Reg. Stud.*, 53, 101792, <https://doi.org/10.1016/j.ejrh.2024.101792>,
1064 2024.
1065 Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R.,
1066 Himmelfarb, J., Bansal, N., and Lee, S. I.: From local explanations to global
1067 understanding with explainable AI for trees, *Nat. Mach. Intell.*, 2, 56-67,
1068 <https://doi.org/10.1038/s42256-019-0138-9>, 2020.
1069 Mann, H. B. and Whitney, D. R.: On a test of whether one of two random variables is
1070 stochastically larger than the other, *Ann. Math. Stat.*, 18, 50-60,
1071 <https://doi.org/10.1214/aoms/1177730491>, 1947.
1072 Martínez-de la Torre, A. and Miguez-Macho, G.: Groundwater influence on soil
1073 moisture memory and land-atmosphere fluxes in the Iberian Peninsula, *Hydrol. Earth*
1074 *Syst. Sci.*, 23, 4909-4932, <https://doi.org/10.5194/hess-23-4909-2019>, 2019.
1075 Martínez-Fernández, J., González-Zamora, A., and Almendra-Martín, L.: Soil
1076 moisture memory and soil properties: An analysis with the stored precipitation
1077 fraction, *J. Hydrol.*, 593, 125622, <https://doi.org/10.1016/j.jhydrol.2020.125622>,
1078 2021.
1079 McColl, K. A., Wang, W., Peng, B., Akbar, R., Short Gianotti, D. J., Lu, H., Pan, M.,
1080 and Entekhabi, D.: Global characterization of surface soil moisture drydowns,
1081 *Geophys. Res. Lett.*, 44, 3682-3690, <https://doi.org/10.1002/2017GL072819>, 2017.
1082 Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G.,
1083 Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B.,
1084 Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and
1085 Thépaut, J.-N.: ERA5-Land: A state-of-the-art global reanalysis dataset for land



- 1086 applications, *Earth Syst. Sci. Data*, 13, 4349-4383, [https://doi.org/10.5194/essd-13-](https://doi.org/10.5194/essd-13-4349-2021)
1087 4349-2021, 2021.
- 1088 Nicolai-Shaw, N., Gudmundsson, L., Hirschi, M., and Seneviratne, S. I.: Long-term
1089 predictability of soil moisture dynamics at the global scale: Persistence versus large-
1090 scale drivers, *Geophys. Res. Lett.*, 43, 8554-8562,
1091 <https://doi.org/10.1002/2016GL069847>, 2016.
- 1092 Obour, P. B., Keller, T., Lamandé, M., and Munkholm, L. J.: Pore structure
1093 characteristics and soil workability along a clay gradient, *Geoderma*, 337, 1186-1195,
1094 <https://doi.org/10.1016/j.geoderma.2018.11.032>, 2019.
- 1095 Pal, M., Maity, R., and Dey, S.: Statistical modelling of vertical soil moisture profile:
1096 Coupling of memory and forcing, *Water Resour. Manag.*, 30, 1973-1986,
1097 <https://doi.org/10.1007/s11269-016-1263-4>, 2016.
- 1098 Pan, J., Shangguan, W., Li, L., Yuan, H., Zhang, S., Lu, X., Wei, N., and Dai, Y.:
1099 Using data-driven methods to explore the predictability of surface soil moisture with
1100 FLUXNET site data, *Hydrol. Process.*, 33, 2978-2996,
1101 <https://doi.org/10.1002/hyp.13540>, 2019.
- 1102 Philip, J. R.: The theory of infiltration: 1. The infiltration equation and its solution,
1103 *Soil Sci.*, 83, 345-357, <https://doi.org/10.1097/00010694-195705000-00002>, 1957.
- 1104 Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro,
1105 E., and Rossiter, D.: SoilGrids 2.0: Producing soil information for the globe with
1106 quantified spatial uncertainty, *SOIL*, 7, 217-240, [https://doi.org/10.5194/soil-7-217-](https://doi.org/10.5194/soil-7-217-2021)
1107 2021, 2021.
- 1108 Qian, X., Wang, B., Chen, J., Fan, Y., Mo, R., Xu, C., Liu, W., Liu, J., and Zhong, P.
1109 A.: An explainable ensemble deep learning model for long-term streamflow
1110 forecasting under multiple uncertainties, *J. Hydrol.*, 662, 133968,
1111 <https://doi.org/10.1016/j.jhydrol.2025.133968>, 2025.
- 1112 Rabiei, S., Babaeian, E., and Grunwald, S.: Deep learning-based short- and mid-term
1113 surface and subsurface soil moisture projections from remote sensing and digital soil
1114 maps, *Remote Sens.*, 17, 3219, <https://doi.org/10.3390/rs17183219>, 2025.
- 1115 Rahmati, M., Amelung, W., Brogi, C., Dari, J., Flammini, A., Bogen, H., Brocca, L.,
1116 Chen, H., Groh, J., Koster, R. D., McColl, K. A., Montzka, C., Moradi, S., Rahi, A.,
1117 Sharghi, F. S., and Vereecken, H.: Soil moisture memory: State-of-the-art and the way
1118 forward, *Rev. Geophys.*, 62, e2023RG000828,
1119 <https://doi.org/10.1029/2023RG000828>, 2024.
- 1120 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N.,
1121 and Prabhat: Deep learning and process understanding for data-driven Earth system
1122 science, *Nature*, 566, 195-204, <https://doi.org/10.1038/s41586-019-0912-1>, 2019.
- 1123 Richards, L. A.: Capillary conduction of liquids through porous mediums, *Physics*, 1,
1124 318-333, <https://doi.org/10.1063/1.1745010>, 1931.
- 1125 Richman, J. S. and Moorman, J. R.: Physiological time-series analysis using
1126 approximate entropy and sample entropy, *Am. J. Physiol.-Heart Circ. Physiol.*, 278,
1127 H2039-H2049, <https://doi.org/10.1152/ajpheart.2000.278.6.H2039>, 2000.
- 1128 Roy, A., Narvekar, P., Murtugudde, R., Shinde, V., and Ghosh, S.: Short and medium
1129 range irrigation scheduling using stochastic simulation-optimization framework with
1130 farm-scale ecohydrological model and weather forecasts, *Water Resour. Res.*, 57,
1131 e2020WR029004, <https://doi.org/10.1029/2020WR029004>, 2021.
- 1132 Saboori, M., Ghag, K. S., Panchanathan, A., Patro, E. R., and Haghighi, A. T.:
1133 Assessing feature importance for forecasting soil moisture in subarctic regions using
1134 gridded historical and forecasted climate data, *Geoderma*, 458, 117304,
1135 <https://doi.org/10.1016/j.geoderma.2025.117304>, 2025.



- 1136 Savenije, H. H. G.: HESS Opinions: Linking Darcy's equation to the linear reservoir,
1137 Hydrol. Earth Syst. Sci., 22, 1911-1916, <https://doi.org/10.5194/hess-22-1911-2018>,
1138 2018.
- 1139 Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I.,
1140 Orlowsky, B., and Teuling, A. J.: Investigating soil moisture-climate interactions in a
1141 changing climate: A review, Earth-Sci. Rev., 99, 125-161,
1142 <https://doi.org/10.1016/j.earscirev.2010.02.004>, 2010.
- 1143 Shapley, L. S.: A value for n-person games, in: Contributions to the Theory of Games,
1144 Volume II, edited by: Kuhn, H. W. and Tucker, A. W., Annals of Mathematics
1145 Studies, Vol. 28, Princeton University Press, 307-318,
1146 <https://doi.org/10.1515/9781400881970-018>, 1953.
- 1147 Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-
1148 Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J.,
1149 Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., Rahmani, F., Song,
1150 Y., Beck, H. E., Bindas, T., Dwivedi, D., Fang, K., Höge, M., Rackauckas, C.,
1151 Mohanty, B., Roy, T., Xu, C., and Lawson, K.: Differentiable modelling to unify
1152 machine learning and physical models for geosciences, Nat. Rev. Earth Environ., 4,
1153 552-567, <https://doi.org/10.1038/s43017-023-00450-9>, 2023.
- 1154 Singh, N. K., Emanuel, R. E., McGlynn, B. L., and Miniati, C. F.: Soil moisture
1155 responses to rainfall: Implications for runoff generation, Water Resour. Res., 57,
1156 e2020WR028827, <https://doi.org/10.1029/2020WR028827>, 2021.
- 1157 Singla, S., Céron, J.-P., Martin, E., Regimbeau, F., Déqué, M., Habets, F., and Vidal,
1158 J.-P.: Predictability of soil moisture and river flows over France for the spring season,
1159 Hydrol. Earth Syst. Sci., 16, 201-216, <https://doi.org/10.5194/hess-16-201-2012>,
1160 2012.
- 1161 Slater, L. J., Arnal, L., Boucher, M.-A., Chang, A. Y.-Y., Moulds, S., Murphy, C.,
1162 Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A.,
1163 and Zappa, M.: Hybrid forecasting: Blending climate predictions with AI models,
1164 Hydrol. Earth Syst. Sci., 27, 1865-1889, <https://doi.org/10.5194/hess-27-1865-2023>,
1165 2023.
- 1166 Tahmouresi, M. S., Niksokhan, M. H., and Ehsani, A. H.: Enhancing spatial
1167 resolution of satellite soil moisture data through stacking ensemble learning
1168 techniques, Sci. Rep., 14, 25454, <https://doi.org/10.1038/s41598-024-77050-0>, 2024.
- 1169 te Wierik, S. A., Cammeraat, E. L. H., Gupta, J., and Artzy-Randrup, Y. A.:
1170 Reviewing the impact of land use and land-use change on moisture recycling and
1171 precipitation patterns, Water Resour. Res., 57, e2020WR029234,
1172 <https://doi.org/10.1029/2020WR029234>, 2021.
- 1173 Tefera, G. W., Ray, R. L., Jackson, R., Deegala, B., and Akenzua, O.: Integrating
1174 machine learning models with ground sensors to enhance soil moisture prediction in
1175 agroecosystems of Texas, Comput. Electron. Agric., 242, 111358,
1176 <https://doi.org/10.1016/j.compag.2025.111358>, 2026.
- 1177 Tosan, M., Nourani, V., Kisi, O., and Dastourani, M.: Evolution of ensemble machine
1178 learning approaches in water resources management: A review, Earth Sci. Inform., 18,
1179 416, <https://doi.org/10.1007/s12145-025-01911-5>, 2025.
- 1180 Trambly, Y., Villarini, G., El Khalki, E. M., Gründemann, G., and Hughes, D.:
1181 Evaluation of the drivers responsible for flooding in Africa, Water Resour. Res., 57,
1182 e2021WR029595, <https://doi.org/10.1029/2021WR029595>, 2021.
- 1183 Tyagi, S., Zhang, X., Saraswat, D., Sahany, S., Mishra, S. K., and Niyogi, D.: Flash
1184 drought: Review of concept, prediction and the potential for machine learning, deep



- 1185 learning methods, *Earth's Future*, 10, e2022EF002723,
1186 <https://doi.org/10.1029/2022EF002723>, 2022.
- 1187 van Genuchten, M. Th.: A closed-form equation for predicting the hydraulic
1188 conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, 44, 892-898,
1189 <https://doi.org/10.2136/sssaj1980.03615995004400050002x>, 1980.
- 1190 Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka,
1191 C., Nemes, A., Pachepsky, Y. A., Padarian, J., Schaap, M. G., Tóth, B., Verhoef, A.,
1192 Vanderborght, J., van der Ploeg, M. J., Weihermüller, L., Zacharias, S., Zhang, Y.,
1193 and Vereecken, H.: Pedotransfer functions in Earth system science: Challenges and
1194 perspectives, *Rev. Geophys.*, 55, 1199-1256, <https://doi.org/10.1002/2017RG000581>,
1195 2017.
- 1196 Vereecken, H., Amelung, W., Bauke, S. L., Boga, H., Brüggemann, N., Montzka,
1197 C., Vanderborght, J., Bechtold, M., Blöschl, G., Carminati, A., Javaux, M., Konings,
1198 A. G., Kusche, J., Neuweiler, I., Or, D., Steele-Dunne, S., Verhoef, A., Young, M.,
1199 and Zhang, Y.: Soil hydrology in the Earth system, *Nat. Rev. Earth Environ.*, 3, 573-
1200 587, <https://doi.org/10.1038/s43017-022-00324-6>, 2022.
- 1201 Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., and
1202 Wood, E. F.: SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the
1203 conterminous US, *Sci. Data*, 8, 264, <https://doi.org/10.1038/s41597-021-01050-2>,
1204 2021.
- 1205 Wang, K., Niu, J., Li, T., and Zhou, Y.: Facing water stress in a changing climate: A
1206 case study of drought risk analysis under future climate projections in the Xi River
1207 Basin, China, *Front. Earth Sci.*, 8, 86, <https://doi.org/10.3389/feart.2020.00086>, 2020.
- 1208 Wang, Y., Shi, L., Hu, Y., Hu, X., Song, W., and Wang, L.: A comprehensive study
1209 of deep learning for soil moisture prediction, *Hydrol. Earth Syst. Sci.*, 28, 917-943,
1210 <https://doi.org/10.5194/hess-28-917-2024>, 2024.
- 1211 Wang, L., Shi, L., Reimers, C., Wang, Y., He, L., Wang, Y., Reichstein, M., and
1212 Jiang, S.: A self-supervised deep learning model for enhanced generalization in soil
1213 moisture prediction, *J. Hydrol.*, 662, 133974,
1214 <https://doi.org/10.1016/j.jhydrol.2025.133974>, 2025.
- 1215 Wankmüller, F. J. P., Delval, L., Lehmann, P., Baur, M. J., Cecere, A., Wolf, S., Or,
1216 D., Javaux, M., and Carminati, A.: Global influence of soil texture on ecosystem
1217 water limitation, *Nature*, 635, 631-638, <https://doi.org/10.1038/s41586-024-08089-2>,
1218 2024.
- 1219 Wiekenkamp, I., Huisman, J. A., Boga, H. R., Lin, H. S., and Vereecken, H.:
1220 Spatial and temporal occurrence of preferential flow in a forested headwater
1221 catchment, *J. Hydrol.*, 534, 139-149, <https://doi.org/10.1016/j.jhydrol.2015.12.050>,
1222 2016.
- 1223 Wilson, D. J., Western, A. W., and Grayson, R. B.: Identifying and quantifying
1224 sources of variability in temporal and spatial soil moisture observations, *Water*
1225 *Resour. Res.*, 40, W02507, <https://doi.org/10.1029/2003WR002306>, 2004.
- 1226 Wolpert, D. H.: Stacked generalization, *Neural Networks*, 5, 241-259,
1227 [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1), 1992.
- 1228 Wösten, J. H. M., Pachepsky, Y. A., and Rawls, W. J.: Pedotransfer functions:
1229 Bridging the gap between available basic soil data and missing soil hydraulic
1230 characteristics, *J. Hydrol.*, 251, 123-150, [https://doi.org/10.1016/S0022-1694\(01\)00464-4](https://doi.org/10.1016/S0022-1694(01)00464-4), 2001.
- 1231
1232 Wu, T., Xu, L., Lv, Y., Cai, R., Pan, Z., Zhang, X., Zhang, X., and Chen, N.:
1233 Integrating causal inference with ConvLSTM networks for spatiotemporal forecasting



1234 of root zone soil moisture, *J. Hydrol.*, 659, 133246,
1235 <https://doi.org/10.1016/j.jhydrol.2025.133246>, 2025.
1236 Wyatt, B. M., Ochsner, T. E., Krueger, E. S., and Jones, E. T.: In-situ soil moisture
1237 data improve seasonal streamflow forecast accuracy in rainfall-dominated watersheds,
1238 *J. Hydrol.*, 590, 125404, <https://doi.org/10.1016/j.jhydrol.2020.125404>, 2020.
1239 Xiong, Z., Zhang, Z., Gui, H., Hu, S., Chen, X., Gao, L., He, Y., Qiu, J., and Xin, Q.:
1240 A meteorology-driven Transformer network to predict soil moisture for agriculture
1241 drought forecasting, *IEEE Trans. Geosci. Remote Sens.*, 63, 4405818,
1242 <https://doi.org/10.1109/TGRS.2025.3543611>, 2025.
1243 Xu, L., Yu, H., Chen, Z., Du, W., Chen, N., and Huang, M.: Hybrid deep learning and
1244 S2S model for improved sub-seasonal surface and root-zone soil moisture forecasting,
1245 *Remote Sens.*, 15, 3410, <https://doi.org/10.3390/rs15133410>, 2023.
1246 Xu, L., Zhang, X., Wu, T., Yu, H., Du, W., Zhang, C., and Chen, N.: Global
1247 prediction of flash drought using machine learning, *Geophys. Res. Lett.*, 51,
1248 e2024GL111134, <https://doi.org/10.1029/2024GL111134>, 2024.
1249 Xu, Q., Shi, Y., Bamber, J. L., Tuo, Y., Ludwig, R., and Zhu, X. X.: Physics-aware
1250 machine learning revolutionizes scientific paradigm for process-based modeling in
1251 hydrology, *Earth-Sci. Rev.*, 271, 105276,
1252 <https://doi.org/10.1016/j.earscirev.2025.105276>, 2025.
1253 Yan, S., Weng, B., Dong, Z., Yan, D., and Fu, Q.: Simulation of soil moisture and
1254 drought prediction in middle reaches of the Yellow River based on machine learning,
1255 *Agric. Water Manag.*, 323, 110068, <https://doi.org/10.1016/j.agwat.2025.110068>,
1256 2026.
1257 Zha, X., Jia, S., Han, Y., Zhu, W., and Lv, A.: Enhancing soil moisture prediction in
1258 drought-prone agricultural regions using remote sensing and machine learning
1259 approaches, *Remote Sens.*, 17, 181, <https://doi.org/10.3390/rs17020181>, 2025.
1260 Zhang, Y., Hao, Z., Feng, S., Zhang, X., Xu, Y., and Hao, F.: Agricultural drought
1261 prediction in China based on drought propagation and large-scale drivers, *Agric.*
1262 *Water Manag.*, 255, 107028, <https://doi.org/10.1016/j.agwat.2021.107028>, 2021.
1263 Zhu, Q., Lin, H., and Doolittle, J.: Soil moisture sensor network design for
1264 hydrological applications, *Hydrol. Earth Syst. Sci.*, 24, 2577-2594,
1265 <https://doi.org/10.5194/hess-24-2577-2020>, 2020.