Journal of Hydrology 667 (2026) 134945

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Improved flash drought forecasting and attribution: A spatial-temporal causality-aware deep learning approach

Sijie Tang [a,b,c,d,e], Shuo Wang [a,b,*], Jiping Jiang [c,d,e], Yi Zheng [c,d,e]

[a] State Key Laboratory of Climate Resilience for Coastal Cities, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[b] Otto Poon Research Institute for Climate-Resilient Infrastructure and Research Institute for Land and Space, The Hong Kong Polytechnic University, Hong Kong, China
[c] School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
[d] Shenzhen Municipal Engineering Lab of Environmental IoT Technologies, Southern University of Science and Technology, Shenzhen 518055, China
[e] State Environmental Protection Key Laboratory of Integrated Surface Water-Groundwater Pollution Control, Southern University of Science and Technology, Shenzhen 518055, China

ARTICLE INFO

This manuscript was handled by Dr. A. Bardossy, Editor-in-Chief, with the assistance of Fijohn Chang, Associate Editor

Keywords:
Soil moisture
Flash drought
Deep learning
Causality

ABSTRACT

Flash droughts pose significant challenges to water resource management and agricultural sustainability, making it imperative to improve their predictability to mitigate potential risks. This study presents a novel deep learning framework that integrates a spatial–temporal causality-aware (STC) module into a CNN-LSTM hybrid architecture to enhance flash drought prediction in China's Greater Bay Area (GBA). Ablation experiments demonstrate that the causality module enhances model generalization (GA = 0.90) and performance (NSE = 0.83), and substantially increases the accuracy of flash drought onset prediction (F1 score = 0.33) compared to baseline models. Explainable Artificial Intelligence (AI) analyses further reveal that incorporating causality strengthens the predictive contributions of key flash drought drivers, including soil moisture memory, downward longwave radiation, and precipitation. Especially, it reveals new insights into drought drivers: downward longwave radiation emerges as a critical yet previously underrecognized predictor of soil moisture variability in humid subtropical climates. Additionally, this study distinguishes the mechanisms underlying slow and flash droughts, highlighting the dominant role of initial soil moisture and persistent shortwave radiation in slow droughts, versus rapid energy imbalances and longwave radiation in flash droughts. Further findings suggest that anthropogenic activities in China's GBA intensify the complexity of drought mechanisms, increasing both prediction difficulty and regional vulnerability to hydrological extremes. The proposed framework and insights provide a foundation for developing more effective flash drought risk management and adaptation strategies in humid subtropical regions.

1. Introduction

Recent studies have increasingly highlighted flash drought (FD) as a growing concern due to their rapid onset and severe impacts (Senay et al., 2008). These events can develop quickly under extreme atmospheric conditions, such as elevated temperatures, strong winds, and low humidity, and are often triggered by insufficient precipitation or increased evaporative demand. As a result, they can cause significant disruptions to ecosystems and agricultural systems (Crausbay et al., 2017; Otkin et al., 2015). Global-scale analyses have shown that FDs are becoming more frequent and intense under the influence of climate change (Christian et al., 2021; Christian et al., 2023), and South China has been identified as a regional hotspot for FD occurrence (Qing et al., 2022; Yuan et al., 2019). Several recent events have underscored the vulnerability of this region. For example, in the summer of 2013, a FD affected 13 provinces, damaging over two million hectares of crops (Yuan et al., 2015). In late summer 2019, another extreme FD persisted through the autumn, impacting 2.35 million hectares of crops and resulting in direct economic losses of 15.08 billion Chinese Yuan (Wang et al., 2020). The most severe event on record occurred in the summer of

* Corresponding author at: State Key Laboratory of Climate Resilience for Coastal Cities, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China.
E-mail address: shuo.s.wang@polyu.edu.hk (S. Wang).

https://doi.org/10.1016/j.jhydrol.2026.134945
Received 23 September 2025; Received in revised form 12 December 2025; Accepted 10 January 2026
Available online 12 January 2026

2022, when the Yangtze River Basin experienced its worst FD since meteorological records began in 1961 (Liu et al., 2023). This event severely affected water resources, hydropower generation, agricultural production, ecosystems, and public health, with an estimated impact on 52.5 million people and 60.9 million hectares of crops, causing direct economic losses of 51.28 billion Chinese Yuan (Lu et al., 2023; Wang and Yuan, 2023).

Given their rapid development and severe consequences, FDs demand increased awareness, proactive adaptation strategies, early warning systems, and effective emergency response (Otkin et al., 2022). However, their sudden onset poses significant challenges for timely prediction and response, complicating decision-making for resource managers and stakeholders (Anderson et al., 2013). Despite the urgency, relatively few studies have specifically addressed FD prediction. Traditional approaches often rely on drought indices, physical process-based models, or coupled systems that use meteorological and hydrological variables (Tyagi et al., 2022). Index-based methods typically assess the interaction between soil moisture and climate drivers using indicators such as the Standardized Precipitation Index or the Standardized Precipitation Evapotranspiration Index to track drought development and identify FD events (AghaKouchak, 2014; Chen et al., 2019; Ho et al., 2021; Liang and Yuan, 2021; Yaseen et al., 2025; Yevjevich, 1967). However, these methods often struggle to capture the nonlinear and dynamic interactions among multiple drivers. In addition, physically based models, including coupled models like Global Climate Models and the Variable Infiltration Capacity model, have been used to simulate soil moisture under future climate scenarios (Mishra et al., 2021). Their performance is limited by process simplifications, uncertainties in model parameters, and the coarse spatial and temporal resolution of input data (Prodhan et al., 2021; Yuan et al., 2020). As a result, they are often inadequate for predicting the rapid onset of FDs, which are driven by complex and fast-evolving physical mechanisms (Huntingford et al., 2019; Yuan et al., 2020).

Machine learning and deep learning techniques have demonstrated considerable potential in water environments (Ahmadianfar et al., 2022; Samadi-koucheksaraee et al., 2019; Samadi-Koucheksaraee et al., 2022; Shirvani-Hosseini et al., 2022), water resource management (Ahmadianfar and Zamani, 2020), and traditional long-term drought prediction (Jamei et al., 2023; Samadi-Koucheksaraee and Chu, 2024), owing to their superior predictive performance. However, their application to FD forecasting remains relatively limited (Feng et al., 2024; Xu et al., 2024). Classical machine learning models (e.g., Artificial Neural Networks, Support Vector Machines, and Random Forests) have been effective in handling multicollinearity and capturing nonlinear relationships among drought-related variables (Park et al., 2016; Tufaner and Özbeyaz, 2020; Zhu et al., 2021). Building on these strengths, deep learning methods like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) offer further advantages. These models are better suited for learning complex spatial–temporal dependencies and are less prone to overfitting in the presence of lagged or high-dimensional data. As a result, they provide improved performance for drought prediction tasks that involve multi-scale and multi-dimensional inputs—features that are particularly relevant to the dynamic nature of FDs (Dikshit et al., 2021; Xiao et al., 2019; Xiao et al., 2024).

Despite their effectiveness in prediction and classification, data-driven machine learning and deep learning models are not inherently designed to uncover or quantify the internal dependencies within complex systems (Runge et al., 2019). To address this limitation, causal inference methods serve as a valuable complement, offering a pathway to improve theoretical understanding by representing the system through an interpretable, potentially deterministic mathematical structure (Reichstein et al., 2019). Previous studies have shown that integrating physical insights from dynamical systems into data-driven models can substantially enhance predictive accuracy and model generalization (Cai et al., 2022; Hoedt et al., 2021; Jiang et al., 2020; Zhao et al., 2019). In addition, improvements in model accuracy can reinforce the reliability of post hoc interpretation using eXplainable Artificial Intelligence (XAI) techniques (Molnar, 2020). XAI has recently received growing attention in geoscience applications, where it has been used to uncover insights into flooding mechanisms (Jiang et al., 2024), ecosystem response (Li et al., 2022), and FD dynamics (Feng et al., 2024).

This study proposes a spatial–temporal causality-aware module integrated into a traditional CNN-LSTM framework to improve soil moisture and FD prediction in China's Greater Bay Area (GBA), a core region of South China. Based on this architecture, we developed a Spatial-Temporal Causality-aware CNN-LSTM (STC-CNNLSTM) model that leverages daily spatiotemporal maps of meteorological drivers to forecast soil moisture. To assess the added value of the causality module, we compared model performance with existing models, with a particular focus on forecasting FD events. We further applied an XAI technique, Expected Gradients, to interpret model predictions and identify the key drivers influencing the onset of both flash and slow droughts. In addition, we examined how the complexity of drought onset correlates with model performance and human activity patterns in the region. This causality-enhanced framework fundamentally differs from existing physics-informed machine learning approaches, such as physics-constrained networks or hybrid dynamical–statistical models. By embedding causal strengths directly into the recurrent network through LSTM cell state initialization, it not only improves predictive accuracy but also enhances understanding of the mechanisms driving drought onset in complex urbanized environments.

2. Model development

To improve FD forecasting and better understand its drivers, we developed the Spatial-Temporal Causality-aware CNN-LSTM (STC-CNNLSTM) model (Fig. 1a). This model predicts soil moisture for a central pixel with a two-week lead time, based on a 14-day time series of ten meteorological variables from the surrounding 3×3 grid.

2.1. Model architecture

Recent research highlights that FDs are shaped by multiscale interactions among land, atmosphere, and ocean systems, influenced by both local indicators and broader climatic drivers (Liang and Yuan, 2021; Nguyen et al., 2021). To address this complexity, we adopted a CNN-LSTM architecture, which is well suited for capturing spatial and temporal patterns critical to soil moisture dynamics. However, conventional deep learning models often struggle to generalize beyond their training conditions, particularly when forecasting extreme events (Plésiat et al., 2024; Trok et al., 2024). This limitation is partly due to the absence of embedded prior knowledge, such as physical principles, process understanding, and causal relationships, in purely data-driven approaches. To overcome this, we enhanced the CNN-LSTM model by incorporating attention mechanisms and a causality module, aiming to improve predictive performance and provide meaningful insights to support early warning and risk management of FDs (Cai et al., 2022; Hoedt et al., 2021; Jiang et al., 2020; Zhao et al., 2019).

The input data is organized as a four-dimensional array with the shape (sequence $\times$ channel $\times$ height $\times$ width), where "sequence" represents the length of the time series, "channel" represents the number of climatic and meteorological drivers, and "height" and "width" define the spatial extent, set as a 3×3 grid in this study. The data is first processed by an attention-based CNN module to extract spatial features from each time step independently, allowing for parallel computation (Plésiat et al., 2024). This is followed by an average pooling layer, which compresses the spatial dimensions into a single point and embeds spatial features into the channel dimension (Zhang et al., 2021). The resulting representation is then passed to a causality-informed LSTM module, which extracts temporal patterns across the time series for all drivers

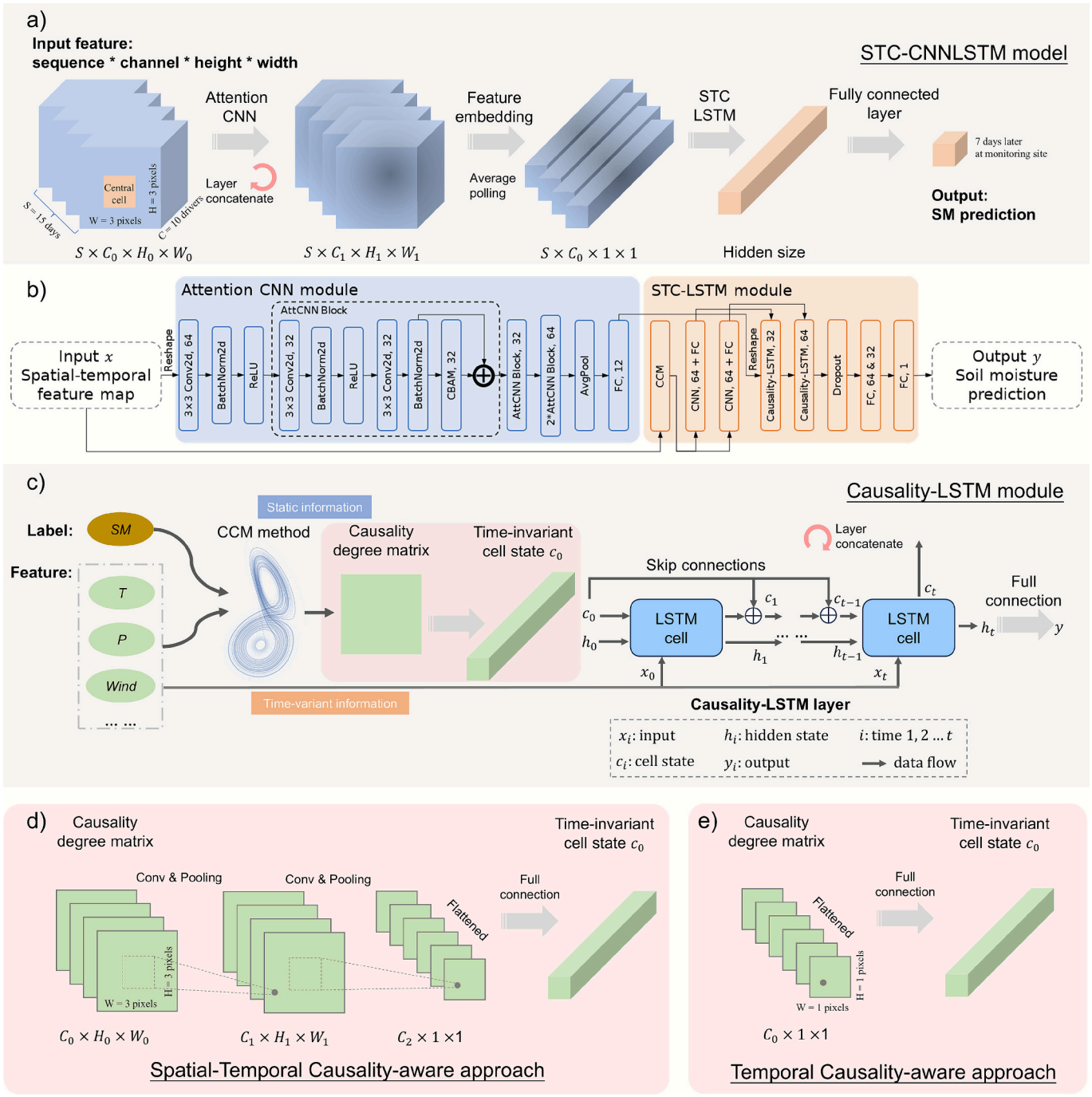

Fig. 1. (a) Overview of the proposed STC-CNNLSTM (Spatial-Temporal Causality-aware CNN-LSTM) model, which predicts soil moisture using a sequence of attention-based CNN and STC-LSTM modules applied to spatial–temporal meteorological driver maps. (b) Architecture and data flow of the STC-CNNLSTM model. (c) Structure and implementation of the Causality LSTM module, which incorporates causality information through two approaches. (d) The Spatial-Temporal Causality-aware approach, used in the STC-CNNLSTM model. (e) The Temporal Causality-aware approach, applied in the TC-CNNLSTM model for ablation analysis.

(Mousavi et al., 2020). Finally, the last hidden state of the LSTM is used to achieve soil moisture prediction through a fully connected layer.

2.2. Attention CNN module

The Attention CNN module is designed to extract spatial features from meteorological input data. In this study, we adopt the classic Residual Network (ResNet) architecture, which is well known for its ability to stabilize training and improve convergence in deep neural networks (He et al., 2016). ResNet is used to capture spatial patterns in surrounding climatic and meteorological variables that influence soil moisture at the target location (Venkatesan and Li, 2017). To enhance the model's ability to focus on relevant spatial features, we incorporate an attention mechanism inspired by human visual perception. Attention mechanisms enable models to emphasize informative input regions while down-weighting irrelevant or noisy signals, thereby enhancing robustness and interpretability in complex and heterogeneous environments (Vaswani et al., 2017). In this study, we integrate ResNet with attention modules, forming the AttCNN block (Fig. 1b), which enhances spatial feature extraction for FD prediction.

The attention mechanism follows the structure of the Convolutional Block Attention Module (CBAM) (Woo et al., 2018), as illustrated in

Fig. S1. CBAM provides lightweight channel- and spatial-attention mechanisms that enable the network to adaptively highlight informative variables and grid-cell regions while suppressing less relevant features. This is important for FD prediction because early signals often manifest as subtle atmospheric and soil-moisture anomalies embedded within heterogeneous hydro-meteorological inputs. The channel-attention module reweights different predictor variables (e.g., radiation fluxes, temperature, wind speed), which differ substantially in physical meaning and temporal variability, while the spatial-attention module emphasizes regions undergoing rapid moisture depletion or atmospheric stress. Unlike transformer-based multi-head attention, CBAM relies on simple attention maps that modulate convolutional features without adding substantial computational complexity. This targeted feature refinement improves the model's sensitivity to weak pre-onset signals and supports more robust identification of FD onset.

Given an input feature map $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$, CBAM sequentially computes attention maps along two dimensions (i.e., channel and spatial) and applies them to refine the feature representation. The result is a spatially informed output that better supports downstream temporal modeling and prediction tasks:

$$\mathbf{F}' = \mathbf{M_C}(\mathbf{F}) \odot \mathbf{F},$$

$$\mathbf{F}'' = \mathbf{M_S}(\mathbf{F}') \odot \mathbf{F}', \tag{1}$$

where $\mathbf{M_C}$ represents the channel attention map, $\mathbf{M_S}$ is the spatial attention map, and $\odot$ denotes element-wise multiplication. Channel attention focuses on identifying 'what' is meaningful across the input channels. As illustrated in Fig. S2b, it can be computed as follows:

$$\mathbf{M_C}(\mathbf{F}) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)) \tag{2}$$

where σ represents the sigmoid function, and MLP denoted the shared multi-layer perceptron with one hidden layer.

In contrast to channel attention, spatial attention identifies "where" informative parts are located, complementing the channel attention (Fig. S2c). It is computed as follows:

$$\mathbf{M_S}(\mathbf{F}) = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)]) \tag{3}$$

where σ denotes the sigmoid function, and $f^{7\times7}$ represents a convolution operation with the filter size of 7×7.

Previous studies have indicates that this additional attention architecture enhances both classification and regression performance across various models (Chen et al., 2021; Mohammadi Foumani et al., 2024). As shown in Fig. 1b, the basic AttCNN block in this study consists of two convolutional layers followed by a CBAM, and four AttCNN blocks form the backbone of the Attention CNN module. Notably, the convolutional layers do not process temporal information in this context. Therefore, the temporal dimension S of the original input $\boldsymbol{x} \in \mathbb{R}^{n\times S\times C\times H\times W}$ is flatten into batch size n, to generate the input $\boldsymbol{x}' \in \mathbb{R}^{nS\times C\times H\times W}$ for the convolutional layers. After average pooling, which reduces the spatial information to a single point, another reshaping operation restores the time dimension. The ResAtt module's output is thus formatted to match the input size ($\boldsymbol{x}'' \in \mathbb{R}^{n\times S\times C}$) for the subsequent recurrent module.

2.3. Causality LSTM module

The Causality LSTM module is designed to extract temporal information while embedding causal relationships directly into the cell state of a standard LSTM, as illustrated in Fig. 1c. Previous studies have shown that incorporating domain knowledge or physical insights from dynamical systems can significantly enhance the performance and generalizability of data-driven models (Cai et al., 2022; Hoedt et al., 2021; Jiang et al., 2020; Zhao et al., 2019).

In this framework, inputs to the Causality LSTM are divided into two types: (1) time-varying forcing variables and (2) static causal information. The forcing variables, such as temperature, precipitation, and pressure, serve as the sequential input to the LSTM. Static causal information is derived from Convergent Cross Mapping (CCM), which quantifies the causal strength between meteorological drivers and soil moisture. This causality degree matrix is used to construct the initial cell state c_0, enabling the LSTM to begin with prior knowledge about the system's causal structure.

The construction of c_0 depends on the type of causality information used. If the causality degree matrix contains only temporal information at the target location (i.e., local pixel), fully connected layers are used to generate c_0 (Fig. 1e). If spatial–temporal causality is included, incorporating all pixels within the model's spatial horizon, a CNN is used to produce c_0 (Fig. 1d). The full model, STC-CNNLSTM, adopts the spatial–temporal causality-aware approach, while the simplified TC-CNNLSTM, used in ablation studies, employs only temporal causality. In addition to initializing the LSTM, the cell state c_0 is skip-connected to all subsequent cell states in the sequence. This design ensures that static causal information continues to influence the temporal modeling throughout the entire sequence.

The CCM-based spatial–temporal causality integration differs fundamentally from existing physics-informed machine learning frameworks such as physics-constrained networks or hybrid dynamical–statistical models. Physics-constrained networks typically incorporate governing equations into the loss function to penalize non-physical solutions (Raissi et al., 2017), whereas hybrid approaches blend outputs from process-based models and data-driven predictors (Qian et al., 2020). By contrast, our framework does not require explicit process equations; instead, it leverages CCM to empirically identify causal relationships among climate drivers and soil moisture. These causal signals are then embedded directly into the recurrent network by initializing the LSTM cell state.

In standard LSTM models, initial cell states c_0 are typically initialized to zero, which forces the model to learn temporal dependencies entirely from data, resulting in high initial loss and slower convergence. Prior work has shown that learning these initial states as trainable parameters can improve model performance (Pitis, 2016). Therefore, by seeding c_0 with CCM-derived causal priors, the network begins with a structured representation of the relative influence of climate drivers, effectively biasing the learning process toward physically interpretable interactions. The skip connections (He et al., 2016) from the causality-informed c_0 to subsequent memory states act as a mechanism to preserve these static priors across all timesteps, rather than allowing them to decay as training progresses. This design functions as an internal causal regularizer: rather than externally constraining outputs, it shapes the trajectory of the hidden state dynamics from the outset.

The main advantage of this approach is that it embeds observationally grounded causal information directly into the network architecture, thereby providing a feature-level weighting strategy, guiding the LSTM to focus on the most influential drivers. It also improves model transparency and reduces risks of overfitting to spurious correlations. Such refinement is particularly important for capturing the sequential dynamics of soil moisture and enhances both the model's ability to learn spatial–temporal dependencies and the interpretability of its predictions (Lehmann et al., 2020).

Specifically, LSTM networks, first introduced by Hochreiter and Schmidhuber (1997), are well suited for learning long-term dependencies in sequence prediction tasks. This makes them particularly effective for modeling complex drought dynamics, such as the transition from meteorological drought to soil moisture drought, and for capturing the memory effects of meteorological drivers over extended time periods. A typical LSTM unit includes a memory cell and three gating mechanisms: the input gate, forget gate, and output gate (Fig. S3). The memory cell retains information across time steps, while the gates regulate the flow of information into, within, and out of the cell (Konapala et al., 2020). These gates are computed as follows:

$$\mathbf{I}_t = \sigma(\boldsymbol{x}_t \boldsymbol{w}_{\mathrm{xi}} + \boldsymbol{h}_{t-1} \boldsymbol{w}_{\mathrm{hi}} + \boldsymbol{b}_{\mathrm{i}}),$$

$$\mathbf{F}_t = \sigma\left(\boldsymbol{x}_t \boldsymbol{w}_{\mathrm{xf}} + \boldsymbol{h}_{t-1} \boldsymbol{w}_{\mathrm{hf}} + \boldsymbol{b}_{\mathrm{f}}\right),$$

$$\mathbf{O}_t = \sigma(\boldsymbol{x}_t \boldsymbol{w}_{\mathrm{xo}} + \boldsymbol{h}_{t-1} \boldsymbol{w}_{\mathrm{ho}} + \boldsymbol{b}_{\mathrm{o}}), \tag{4}$$

where $\mathbf{I}_t$ is the input gate, $\mathbf{F}_t$ is the forget gate, $\mathbf{O}_t$ is the output gate; $\boldsymbol{x}_t$ is the input, $\boldsymbol{h}_t$ represents the current hidden state; $\boldsymbol{w}_{\mathrm{xi}}$, $\boldsymbol{w}_{\mathrm{hi}}$, $\boldsymbol{w}_{\mathrm{xf}}$, $\boldsymbol{w}_{\mathrm{hf}}$, $\boldsymbol{w}_{\mathrm{xo}}$, $\boldsymbol{w}_{\mathrm{ho}}$ are weight parameters and $\boldsymbol{b}_{\mathrm{i}}$, $\boldsymbol{b}_{\mathrm{f}}$, $\boldsymbol{b}_{\mathrm{o}}$ are bias parameters. As depicted in Fig. 1c, the hidden state $\boldsymbol{h}_t$ signifies short-term memory, while the cell state $\boldsymbol{c}_t$ represents long-term memory. Both hidden state and cell state are passed to the cell of the next time step, and are controlled by an input node $\widetilde{\boldsymbol{c}}_t$:

$$\widetilde{\boldsymbol{c}}_t = tanh(\boldsymbol{x}_t \boldsymbol{w}_{\mathrm{xc}} + \boldsymbol{h}_{t-1} \boldsymbol{w}_{\mathrm{hc}} + \boldsymbol{b}_{\mathrm{c}}), \tag{5}$$

where $\boldsymbol{w}_{\mathrm{xc}}$ and $\boldsymbol{w}_{\mathrm{hc}}$ are weight parameters and $\boldsymbol{b}_{\mathrm{c}}$ is bias parameter. The input gate $\mathbf{I}_t$ governs how much we take new data into account via $\widetilde{\boldsymbol{c}}_t$ and the forget gate $\mathbf{F}_t$ addresses how much of the old cell state $\boldsymbol{c}_{t-1}$ we retain. In our modification, we aim for the cell state $\boldsymbol{c}_t$ capture and convey essential long-term information, specifically global causality. To achieve this, we use causality information to initialize the cell state $\boldsymbol{c}_0$, reinforcing this information in subsequent cell states through a skip connection. In mathematical expressions, these computations are performed as follows:

$$\boldsymbol{c}_0 = \mathbf{CNN}(\mathbf{CCM}(\boldsymbol{X}, \boldsymbol{Y})) \text{ or } \boldsymbol{c}_0 = \mathbf{MLP}(\mathbf{CCM}(\boldsymbol{X}, \boldsymbol{Y}))$$

$$\boldsymbol{c}_t = \mathbf{F}_t \odot (\boldsymbol{c}_{t-1} + \boldsymbol{c}_0) + \mathbf{I}_t \odot \widetilde{\boldsymbol{c}}_t,$$

$$\boldsymbol{h}_t = \mathbf{O}_t \odot \boldsymbol{tanh}(\boldsymbol{c}_t) \tag{6}$$

where $\mathbf{CCM}(\boldsymbol{X}, \boldsymbol{Y})$ represents the causality degree between input and output of training dataset, calculated via the CCM approach. The function CNN() represents the Spatial-Temporal Causality-aware approach using convolutional layers, while MLP() represents the Temporal Causality-aware approach with fully connected layers. And $\odot$ represents elementwise product operator. As a comparison, basic LSTM networks generally initiate hidden state $\boldsymbol{h}_0$ and cell state $\boldsymbol{c}_0$ with zero or random matrices, and cell state $\boldsymbol{c}_t$ is directly derived from $\boldsymbol{c}_{t-1}$ instead of $\boldsymbol{c}_0$.

3. Data and methods

3.1. Study area and dataset

The Guangdong–Hong Kong–Macao Greater Bay Area (GBA) is the most populous urban region in the world, with a total population of approximately 86 million and a land area of about 56,000 km^2. Centered around the Pearl River Delta, the GBA includes nine major cities such as Guangzhou, Shenzhen, Hong Kong, and Macao (Fig. S4). The region's climate ranges from humid subtropical to tropical in the southernmost areas, with an average annual precipitation of around 2,000 mm. Given its high population density, rapid urbanization, and frequent reports of FD events in the literature, South China, and China's GBA in particular, has been identified as a hotspot for FD risk. Therefore, China's GBA is selected as a representative study area for exploring FD prediction in the broader context of South China.

Produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5 is a state-of-the-art reanalysis dataset that provides global atmospheric and surface variables at high spatial and temporal resolutions. In this study, we selected a comprehensive set of 10 variables from the ERA5-Land dataset (Muñoz-Sabater et al., 2021) to achieve the prediction and analysis of soil moisture dynamics. These variables include 2-meter mean and maximum air temperatures (Temp, MaxT), precipitation (Prec), evaporation (Evap), surface pressure (PSFC), surface net solar radiation (SSR), surface net thermal radiation (STR), 10-meter wind speed (Wind), and soil temperature and water content (SoilT, SoilW). Daily data with a spatial resolution of 0.1° (~9 km) was obtained for the period from 1950 to 2023.

Notably, the dataset provides two vertical wind components, which were combined into a single scalar wind speed value, disregarding directional information. Input data were extracted from a 3 × 3 grid, with the central pixel as the prediction target. The neighborhood size was set to 3 × 3 based on the spatial characteristics of the GBA and the resolution of ERA5-Land. Because the domain contains extensive coastal boundaries, expanding to a 5 × 5 window would result in a large portion of ocean pixels, especially around key regions such as Shenzhen and Hong Kong (Fig. S4), and would fail our requirement that at least 50 % of the neighborhood consist of land pixels with valid observations. The 3 × 3 window is therefore the minimal configuration that preserves spatial heterogeneity while ensuring full coverage in this coastal region. For 3 × 3 neighborhoods that contain ocean pixels, missing values are replaced with the mean of the land pixels within the same window to maintain spatial consistency. Recognizing the importance of both forecasting FD onset and identifying causal transmission patterns, the model was trained on sequences of 14 consecutive days.

3.2. Flash drought definition

Flash drought (FD) events were identified from soil moisture observations based on a rapid onset intensification rate (Qing et al., 2022). As illustrated in Fig. 2, a flash drought is defined by a decline in soil moisture from the 40th percentile to the 20th percentile, with an average intensification rate of at least one percentile per day. Drought events that do not meet this threshold are classified as slow droughts. To ensure consistency in evaluation, soil moisture percentiles were calculated over the full observation period. This time-invariant percentile threshold aligns with our short-term soil moisture prediction framework. In contrast, moving-window or day-of-year percentile baselines rely on specific climatological context and would decouple drought identification from the model's short-term regression target. By allowing direct comparison of absolute soil moisture changes, the fixed-percentile definition provides a more coherent and appropriate basis for evaluating FD. The onset point of a drought is marked by the last day on which soil moisture exceeds the 40th percentile. The drought period begins when soil moisture first drops below the 20th percentile and ends when it subsequently rises above that threshold. The intensification rate during the onset period is mathematically defined as:

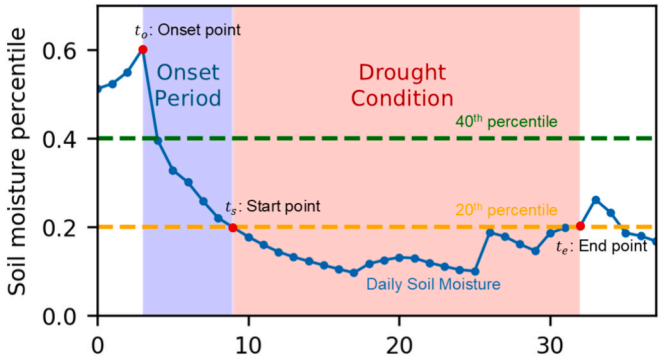

Fig. 2. Illustration of the flash drought definition using a real case from the study area. Soil moisture dropped from above the 40th percentile (onset point t_o) to below the 20th percentile (start point t_s) within six days and remained below the 20th percentile for more than 20 days until the end point t_e. The orange and green dashed lines indicate the 40th and 20th soil moisture percentiles, representing wet and dry conditions, respectively, calculated over the entire monitoring period. The purple shaded area marks the onset period, while the pink shaded area represents the duration of drought conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\text{Intensification rate} = \frac{p_{t_o} - p_{t_s}}{t_s - t_o} \tag{7}$$

where p_t indicates the soil moisture percentile at time t. If the computed intensification rate is not less than one percentile per day, the event is classified as a flash drought; otherwise, it is considered a slow drought.

3.3. Convergent cross mapping for causal discovery

In hydrology, several methods have been developed in recent decades to infer causal relationships from observational data. These approaches can be broadly classified into four categories: linear and nonlinear autoregressive models (such as Granger causality and Transfer Entropy), graph-based methods (for example, the Peter-Clark algorithm), and techniques based on time-delay embedding theory, including Convergent Cross Mapping (CCM). The first three groups generally assume that the underlying system behaves in a purely stochastic manner. In contrast, CCM is based on the idea that causal interactions arise from an underlying dynamical system (Sugihara et al., 2012). CCM was specifically developed to address nonlinear, state-dependent relationships. It approaches the problem from the perspective of dynamical systems, positing that interactions between variables are governed by an underlying deterministic framework. By applying Takens' theorem and nonlinear state-space reconstruction, CCM seeks to identify causal links that may not be evident through traditional statistical methods. This approach is particularly effective in cases where variables are inseparable and only weakly coupled. Furthermore, CCM can distinguish between direct interactions and apparent correlations that result from shared external drivers (Sugihara et al., 2012).

Land–atmosphere interactions play a central role in controlling Earth's surface water and energy budgets (Indu et al., 2022). These feedback mechanisms are often represented by systems of coupled differential equations, with soil moisture acting as a key variable that influences the partitioning of water and energy fluxes (Brubaker and Entekhabi, 1996). However, the relationship between soil moisture and meteorological variables is characterized by complex nonlinearity and weak coupling, which results in poor synchronization and makes causal detection challenging. In such cases, CCM has been shown to be a more effective approach for identifying causality (Wang et al., 2018). An explanation of CCM's underlying principles and its application in this study is provided in Text S1. Specifically, we conducted a sensitivity analysis on the CCM parameters to optimize causal signal retrieval. It is important to note that CCM requires sufficiently long and reliable time series for accurate state-space reconstruction. In this study, the ERA5-Land dataset provides more than 70 years of continuous daily records, ensuring robustness of the results. However, when applied to shorter or noisier observational datasets, CCM may be subject to increased uncertainty in causality detection.

In this study, we assessed the bidirectional causal relationships between meteorological variables and soil moisture within the surrounding 3 × 3 grid cells. The correlation coefficients and their statistical significance were combined to form a causality degree matrix, which was then input into convolutional layers to extract static patterns. This process enhances the representation of physical mechanisms within the cell state of LSTM units using skip connections (Fig. 1c).

To prevent information leakage, all CCM causality degree matrices were computed exclusively from the training-period data for both temporal and spatial split. For temporal split, CCM was derived for each pixel using only its corresponding training years, and the resulting matrices were applied unchanged during prediction on the test period. For spatial cross-validation, CCM matrices for test-region pixels were recomputed solely from their pre-forecast historical data. Although this is not implemented as a rolling window, the full pre-forecast history was used to ensure adequate data length for convergence, consistent with the long-term physical relationships that CCM is designed to capture. All forecasting and interpretability analyses were performed strictly on test data, ensuring that no future or test-period information influenced model training or CCM calculation. Calculations of CCM-based causality were conducted using the causal-ccm package in Python (Javier, 2021).

While CCM provides a valuable means of detecting potential causal influences in nonlinear systems, we acknowledge that its results do not fully preclude the influence of confounding variables or hidden drivers. In this study, CCM was employed as an empirical tool to complement physical understanding of hydro-meteorological processes, rather than as a definitive proof of causation. Compared with correlation or lagged analysis, CCM reduces, but does not entirely eliminate, the possibility of spurious associations. Extensive prior literature has examined the performance characteristics of CCM across a range of dynamical systems, showing that CCM may be less reliable under strong coupling but performs well under weakly coupled dynamics (Benkő et al., 2025). This is one of the primary reasons it is suitable for climate and meteorological variables. In the present study, to reduce the impact of potential confounding within reasonable scope, we adopted several safeguards: (i) use of physically interpretable drivers, (ii) CCM convergence diagnostics, and (iii) the use of bidirectional CCM agreement to filter spurious relationships. Importantly, the CCM-derived matrices are used only as a relative weighting signal for LSTM cell initialization, not as standalone causal claims. More comprehensive robustness tests (e.g., surrogate data experiments, expanded driver sets, or integration with graphical causal frameworks) could further strengthen causal inference. Future studies could incorporate such approaches to refine the identification of causal pathways leading to FD onset.

3.4. Data Preprocessing and model training

For each site, the time series was divided chronologically into training, validation, and test sets in a 70:10:20 ratio (Bishop and Nasrabadi, 2006; Xu and Goodacre, 2018). The validation set was used to monitor the training process and implement early stopping to prevent overfitting (Prechelt, 2002). To maintain sample independence despite the sequential structure of the data, input features with dimensions 14 × 13 × 3 × 3 (comprising 10 variables, 2 location labels, and 1 time label) and their corresponding targets (of length 1) were paired and then shuffled independently within each dataset (Han et al., 2019). Feature normalization was performed using the z-score method, defined as follows:

$$z = \frac{x - \mu}{\sigma} \tag{8}$$

where z is the normalized value, x is the raw value, μ is the mean, and σ is the standard deviation, both calculated from the training set (Grus, 2019). For wind speed and precipitation, which exhibited skewed distributions, log normalization was applied instead of z-score normalization.

Model architecture hyperparameters (e.g., the number of layers in the convolutional and recurrent modules) were determined by balancing model performance and complexity (Bengio, 2012; Goodfellow et al., 2016). Given limited computational resources, an exhaustive hyperparameter search was not feasible. Instead, we manually tuned the architecture through empirical testing, guided by established heuristics such as the ten-times rule and scaling laws (Bahri et al., 2024). Detailed parameter configurations for the STC-CNNLSTM model are provided in Table S1.

Training hyperparameters govern the learning dynamics of the model. Among these, the learning rate and batch size jointly exert a significant influence on convergence and training stability (He et al., 2019). We evaluated various combinations of batch sizes (ranging from 32 to 512) and learning rates (from 1e to 4 to 1e-1) to identify optimal settings (Balles et al., 2016). Based on convergence behavior and computational efficiency, a batch size of 512 and a learning rate of 1e-4 were selected. The Adam optimizer was employed for gradient descent, owing to its adaptive learning rate mechanism, which generally reduces

sensitivity to the initial learning rate compared to other optimizers (Kingma and Ba, 2014).

The number of training epochs was treated as a hyperparameter to mitigate overfitting. Early stopping was implemented during training, and the model achieving the best performance on the validation set was selected for subsequent evaluation and analysis. For visualization purposes, the model was trained for 5 epochs, with each epoch processing approximately 8.9 million samples. This training process required approximately 6 h on a single NVIDIA GeForce RTX 3060 12G GPU. To ensure the robustness and stability of the results, we conducted ten independent training runs with different random seeds, forming a multiple-model ensemble to account for predictive uncertainty (Mienye and Sun, 2022).

While the aforementioned dataset-partitioning scheme enables the assessment of the model's temporal generalization, its applicability to regions with distinct climatic conditions and soil properties remains insufficiently evaluated. To address this limitation and provide a more comprehensive assessment of spatial generalization, we employed a spatial Monte Carlo cross-validation approach (Xu and Liang, 2001), involving five independent random splits (70 % training, 15 % validation, 15 % test). Early stopping was again applied using the validation set. The average performance across the five test sets was used as a robust estimate of the model's spatial generalizability.

3.5. Ablation study and metrics

We conducted an ablation study to evaluate the effectiveness of the STC-CNNLSTM model. By replacing the spatial–temporal causality module with temporal causality module, we obtained the TC-CNNLSTM model. Further removing the causality module resulted in the CNN-LSTM model, which serves as the basic backbone. The LSTM model is a fundamental baseline model that focuses solely on the target pixel, disregarding spatial information from more distant grids.

Mean Squared Error (MSE) is the most widely used loss function for regression tasks. In this study, we utilized MSE as the loss function to train our models and as a criterion to assess their performance. MSE assigns greater weight to larger errors compared to smaller ones, meaning it penalizes outliers more severely. It is also referred to as the L2 norm or the Euclidean distance. MSE measures the average of the squared differences between the simulated values S_i and actual observations O_i, and can be calculated as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2 \tag{9}$$

where N is the total sample number. The L1 norm (mean absolute error, MAE) is more robust to outliers than MSE, measuring the average of the absolute differences between the simulations and observations:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|S_i - O_i| \tag{10}$$

The Nash–Sutcliffe Efficiency coefficient (NSE) is another widely used evaluation metric in geosciences, which quantifies how well the model simulations predict observed outcomes (Nash and Sutcliffe, 1970; Shahid et al., 2018). The NSE is defined as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2} \tag{11}$$

where $\overline{O}$ represents the average of observations. An NSE value of 1 indicates perfect model prediction, while a value of 0 suggests that the model's predictions are as accurate as simply using the mean of the observed data. Negative values indicate that the model performs worse than the mean predictor.

Generalizability is a crucial consideration for data-driven models, reflecting the model's ability to maintain strong performance when applied to new and unseen datasets. To quantitatively assess this aspect, the generalization ability (GA) metric is employed, which compares the model's performance on the test set to its performance on the training set (Gorgij et al., 2023). GA is defined as follows:

$$GA = \frac{RMSE_{train}}{RMSE_{test}} = \frac{\sqrt{MSE_{train}}}{\sqrt{MSE_{test}}} \tag{12}$$

where RMSE refers to Root Mean Square Error. If the GA value exceeds one, the model is considered undertrained or underfitting; conversely, if the GA value is less than one, the model is regarded as overtrained or overfitting. Generally, moderate overfitting is acceptable in deep learning models, whereas underfitting is not. Therefore, when the GA value is less than one, the closer it is to one, the better the model's generalization ability.

We expect the models to effectively identify FD events based on soil moisture simulations. Since drought event prediction is a classification task rather than a regression problem, four metrics are employed to evaluate classification performance: accuracy, precision, recall, and F1-score (Powers, 2020). These metrics are calculated using the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) in the context of binary classification. Accuracy is defined as the proportion of correct classifications (both positive and negative) among all predictions, and can be expressed as follows (Menditto et al., 2007):

$$\text{Accuracy} = \frac{\text{correct classifications}}{\text{total classifications}} = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

In a balanced dataset, where the sample sizes of all classes are comparable, accuracy can serve as a coarse measure of model quality. However, in cases involving extreme events, such as FDs, which represent only a small fraction of the total observations, a high accuracy score may be misleading due to class imbalance. In such scenarios, recall, also known as the true positive rate, becomes particularly important. Recall is defined as the proportion of actual positive cases that are correctly identified by the model, and is calculated as (Powers, 2020):

$$\text{Recall} = \frac{\text{correctly classified actual positives}}{\text{all actual positives}} = \frac{TP}{TP + FN} \tag{14}$$

Recall represents the probability of detection, indicating the fraction of FD events that are successfully identified by the models. Precision, on the other hand, quantifies the proportion of all predicted positive cases that are indeed true positives. Precision is mathematically defined as (Menditto et al., 2007):

$$\text{Precision} = \frac{\text{correctly classified actual positives}}{\text{all classified positives}} = \frac{TP}{TP + FP} \tag{15}$$

Precision can measure the fraction of actual FD events that are classified as FDs.

Precision improves as the number of false positives decreases, whereas recall improves as the number of false negatives decreases. However, false positives and false negatives often exhibit an inverse relationship, resulting in a trade-off between precision and recall. To balance this trade-off, the F1 score is commonly employed. The F1 score is defined as the harmonic mean of precision and recall, and is calculated as (Powers, 2020):

$$\frac{2}{\text{F1}} = \frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}$$

$$\text{F1} = 2^{*}\frac{\text{Precision [MULSGN] Recall}}{\text{Precision + Recall}} = \frac{2TP}{2TP + FP + FN} \tag{16}$$

This metric balances the importance of precision and recall, making it a more informative measure than accuracy, especially for class-imbalanced datasets. When precision and recall are similar in value, the F1 score will also be close to that value. Conversely, when there is a significant disparity between precision and recall, the F1 score will be

closer to the lower of the two. This property ensures that the F1 score penalizes models that perform well on one metric but poorly on the other, thus providing a more holistic assessment of classification performance.

3.6. XAI-based model interpretation

It is essential to employ Explainable Artificial Intelligence (XAI) techniques to gain insight into the implicit linear or nonlinear interactions between potential drivers and soil moisture, thereby enhancing our understanding of the mechanisms behind FDs. In this study, we utilized the Expected Gradients (EG) method, developed by Erion et al. (2021) as an extension of the Integrated Gradients (IG) method (Sundararajan et al., 2017). Similar to several other attribution methods, such as the Shapley value, EG aims to explain the difference between a model's current prediction and the prediction that the model would make with a baseline input.

EG was selected as the interpretability technique because it provides path-integrated, baseline-robust attributions suited for spatiotemporal regression tasks. Since the goal of the interpretability analysis is to obtain relative driver importance rather than to benchmark XAI methods, we focus on EG as a consistent, theoretically grounded attribution tool. A broader comparison with additional attribution techniques is beyond the scope of this study but represents a promising direction for future research.

For model f, the IG value for feature i is defined as:

$$\mathrm{IG}_i(x, x') := (x_i - x') \times \int_{\alpha=0}^{1} \frac{\partial f(x' + \alpha(x - x'))}{\partial x_i} d\alpha \tag{17}$$

where the local gradients are integrated over a simplified straight-line path ($\gamma(\alpha) = x' + \alpha(x - x')$ for $\alpha \in [0, 1]$) from the baseline input x' to the target input x. The target input x represents a specific input sample whose features are expected to be interpreted, while the baseline input x' is tricky to choose that correctly represents a lack of information. EG avoids an arbitrary choice of baseline by integrating over a dataset. For model f, the EG value for is defined as:

$$\mathrm{EG}_i(x) := \int_{x'} \left((x_i - x') \times \int_{\alpha=0}^{1} \frac{\partial f(x' + \alpha(x - x'))}{\partial x_i} d\alpha \right) p_D(x') dx' \tag{18}$$

where p_D is the density function of the underlying data distribution D. The EG method assumes that the baseline input x' follows the same data distribution D thus avoids specifying x'. As directly integrating over the training distribution is difficult, the integrals in Equation (19) can be instead reformulated as expectations:

$$\mathrm{EG}_i(x) := \mathbb{E}_{x' \sim D, \alpha \sim U(0,1)} \left[(x_i - x') \times \frac{\partial f(x' + \alpha(x - x'))}{\partial x_i} \right] \tag{19}$$

This expectation-based formulation is a sampling-based approximation: (1) draw samples of x' from the training dataset and α from the uniform distribution $U(0,1)$, (2) compute the value inside the expectation for each sample x' and (3) average the values over all samples. Eventually, the EG value $\mathrm{EG}_i(x)$ denotes the direct additive influence of feature i in the target input x to the model output (soil moisture in our case).

Given the spatial–temporal component of the Expected Gradients (EG) values for features in this study, we aggregated the EG values across spatial–temporal grids to represent the overall contribution of feature i. A positive EG value for a specific feature indicates its role in increasing the predicted value relative to the baseline prediction, while a negative EG value suggests the opposite. We established a cutoff threshold to determine whether a specific driving variable makes a significant contribution, selecting the 80th percentile of the absolute values of all EG values for feature i as the threshold. Features with absolute EG values exceeding this threshold are considered main drivers. Furthermore, the number of main drivers contributing to soil moisture is defined as the complexity of soil moisture dynamics. In this study, we calculated EG using the SHAP package (Lundberg, 2017) in Python.

4. Results and discussion

4.1. Changes in flash drought characteristics

We analyzed the FD frequency and trends in China's Greater Bay Area (GBA) from 1980 to 2023. Since the 1980 s, anthropogenic greenhouse gas emissions have contributed to rising temperatures and altered climate variability (Mouginot et al., 2019; Reid et al., 2016). As shown in Fig. 3, FD occurrences have increased across most of China's GBA, with approximately 74 % of the area experiencing higher annual FD frequency and 53 % showing statistically significant increases. While some regions exhibited significant decreases, the overall trend indicates an increase of 2.4 FD events per year for the entire study area (Fig. 3c). The southern China, particularly Guangzhou, face higher FD risks compared to the northern China (Yuan et al., 2019). The persistent rise in FD frequency, along with projections of continued increases throughout the century (Yuan et al., 2019), underscores the urgent need for effective FD prediction tools in this region.

4.2. Soil moisture prediction

Fig. 4a and 4b present the loss curves for the STC-CNNLSTM model. The validation loss initially declines along with the training loss during the first 150 iterations, but then plateaus while the training loss continues to decrease, indicating overfitting. We selected models with the best validation performance within five epochs (240 iterations) for further analysis. An ablation study assessed the impact of the causality module: replacing the CNN encoder with fully connected layers produced the TC-CNNLSTM (local causality only), omitting initial state training yielded the CNNLSTM, and further removal of the spatial module resulted in the baseline LSTM. While the STC-CNNLSTM showed lower training performance, it consistently outperformed the simplified models (TC-CNNLSTM, CNN-LSTM, and LSTM) on the test set, demonstrating the benefit of the causality module and broader spatial–temporal context (Fig. 4c–e). As shown in Fig. 4f, the STC-CNNLSTM achieved a mean generalizability score of 0.90, surpassing TC-CNNLSTM (0.54), CNNLSTM (0.41), and LSTM (0.63). The detailed metrics can be found in Table S2.

The results show that the baseline LSTM model partially captures temporal patterns in soil moisture dynamics, yielding reasonably reliable predictions, consistent with previous studies (Datta and Faroughi, 2023; Gao et al., 2021). Incorporating the spatial information extractor substantially improves model performance; however, this enhancement has minimal impact on NSE for the test set. In contrast, the training set NSE increases by 0.24, likely due to the increased model complexity introduced by the spatial module. Regularization techniques are therefore essential to mitigate overfitting and improve generalization, as reflected by a decrease in training NSE (from 0.90 with CNN-LSTM to 0.84 with TC-CNNLSTM) and an increase in test NSE (from 0.35 to 0.56). The TC-CNNLSTM demonstrates acceptable generalization (mean GA = 0.54) and notable improvements on unseen datasets.

Furthermore, the STC-CNNLSTM leverages additional convolutional layers to embed spatial–temporal causality, guiding LSTM cell states and achieving further gains in test NSE (+0.27) and GA (+0.36) over the TC-CNNLSTM. These results underscore the causality module's role in enhancing generalization as an implicit regularization mechanism. By strategically weighting feature dimensions, the causality module encourages LSTM layers to prioritize salient features (Hwang, 2020; Pitis, 2016; Wenke and Fleming, 2019). Overall, these findings highlight the value of integrating causality-based signals derived from physically interpretable relationships into data-driven models, particularly for complex dynamical systems such as soil moisture dynamics (Cai et al., 2022; Hoedt et al., 2021; Jiang et al., 2020; Zhao et al., 2019).

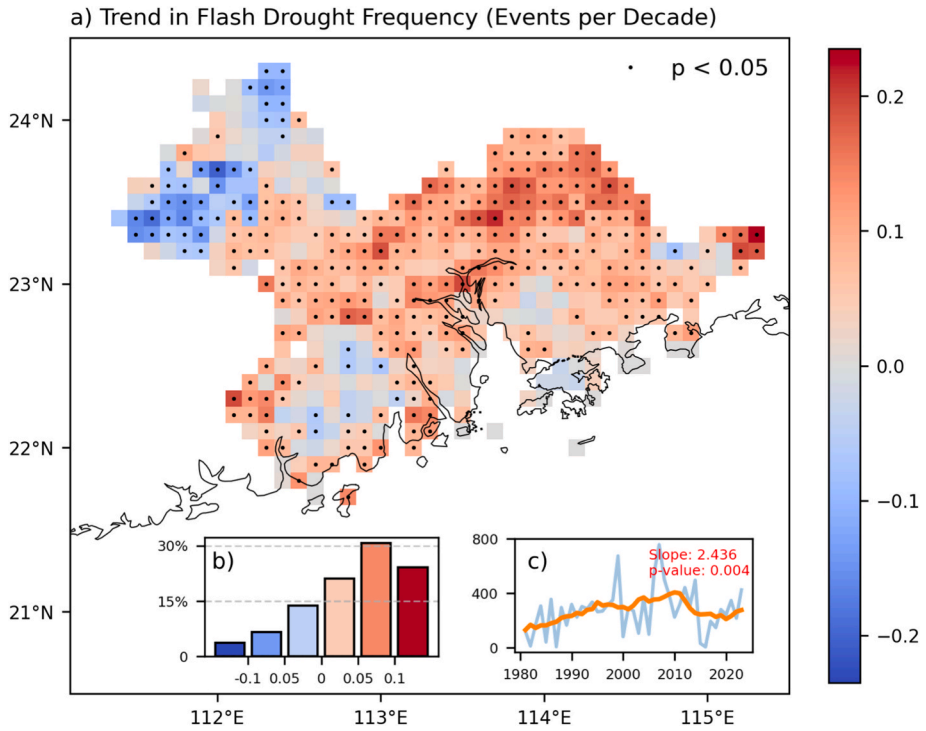

Fig. 3. Long-term increasing trend in flash drought events in China's GBA, 1980–2023. (a) Rate of change in decadal FD occurrence per grid cell, with dots denoting statistical significance; (b) Proportion of grid cells by change rate category; (c) Interannual variation in total FD events across the region, with the orange line indicating a 10-year rolling average. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

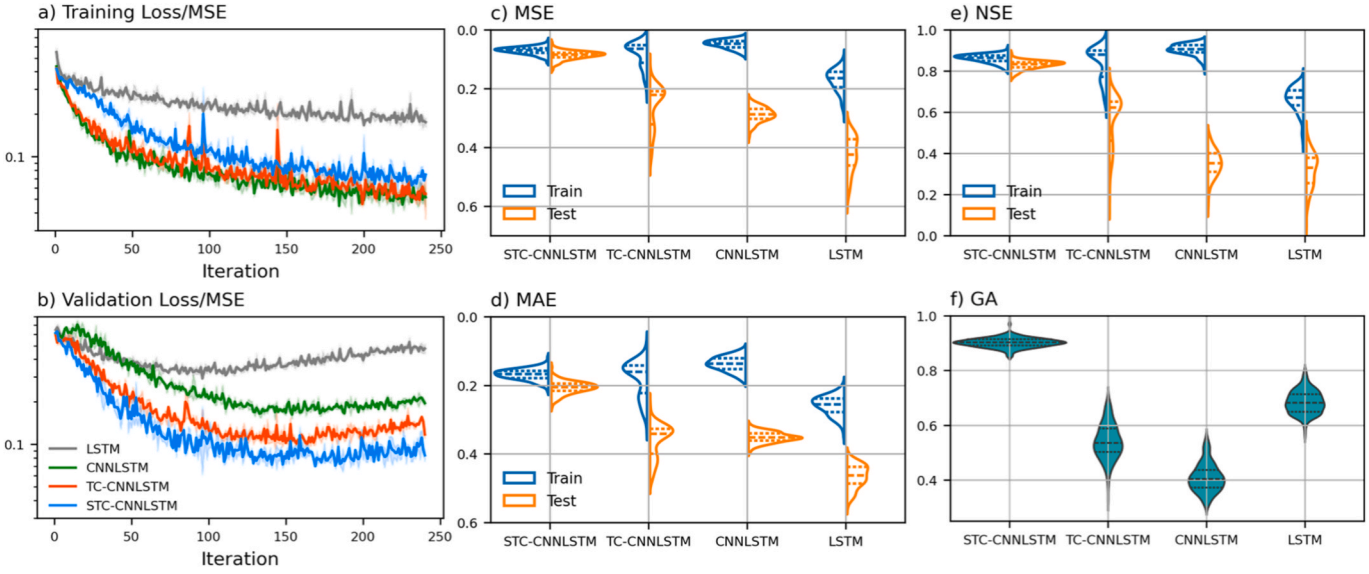

Fig. 4. Comparison of the STC-CNNLSTM model with TC-CNNLSTM, benchmark CNN-LSTM, and baseline LSTM across 10 training periods with different random seeds. The STC-CNNLSTM model outperforms the others on test set. (a, b) Training and validation loss curves from 10 runs; solid lines indicate session means with 95% confidence intervals. (c–e) Model performance metrics (MSE, MAE, and NSE) on training and test sets. (f) Generalization ability of all models, with violin plots showing kernel density estimates and quartiles (dashed lines).

Although the GBA is relatively small in area, it encompasses diverse underlying surface characteristics. To assess how internal heterogeneity influences model performance, we stratified the evaluation by land cover type. Using the MODIS Land Cover Classification dataset (MCD12C1_T1; Fig. S5), we compared model metrics across major land cover categories. The results (Fig. S6) show that the proposed STC-CNNLSTM consistently outperforms the other models across all land cover types. While performance improvements are generally comparable across categories, NSE variation is larger for land covers with smaller spatial extent, reflecting limited sample size. These findings confirm the stability and robustness of the proposed approach under contrasting surface conditions.

To more thoroughly assess spatial generalization, we performed spatial Monte Carlo cross-validation. Unlike temporal splits, which sample from the same spatial distribution, spatial cross-validation requires the model to predict at entirely unseen locations. This resulted in

greater fluctuations in validation loss during training (Fig. 5a) compared to the temporal scenario (Fig. 4b), reflecting increased spatial uncertainty and heterogeneity. Such variability is expected, given that geophysical factors influencing FDs (e.g., soil properties, land cover, and microclimate) can differ substantially across regions, complicating generalization. Overall, model performance declined slightly across all metrics under spatial cross-validation (Fig. 5b–e and Table S2), consistent with previous findings that spatial cross-validation provides a lower-bound estimate of model performance due to the challenge of spatial extrapolation (Wadoux et al., 2021). Nevertheless, the proposed STC-CNNLSTM achieved an average GA score of 0.72, which, although lower than in the temporal split scenario, remains substantially higher than the benchmark and baseline models. This demonstrates the model's robust capacity for spatial generalization.

4.3. Flash drought forecasts

While prior studies have primarily addressed soil moisture prediction with various models (Cai et al., 2019; Kornelsen and Coulibaly, 2014), few have explored their application for drought event detection and early warning. Data-driven models, especially those trained on large datasets, often emphasize frequent patterns in the data. As a result, they perform well under typical conditions but may struggle with rare phenomena such as FDs.

Fig. 6 presents the continuous soil moisture predictions used for drought event identification. The baseline LSTM exhibits a persistent upward bias and fails to track temporal variability, whereas the STC-CNNLSTM effectively captures both gradual soil moisture deficits and abrupt drops. Notably, some FDs occur when soil moisture remains near the drought threshold (20th percentile) rather than reaching extremely low levels, which makes accurate classification particularly challenging. The predictions from STC-CNNLSTM closely follow the observed dynamics. Both STC-CNNLSTM and TC-CNNLSTM successfully detected the FD that developed in the latter half of 2019, while CNN-LSTM misclassified it as a normal slow drought and the baseline LSTM captured only a shortened slow-drought period. For the short but intense soil moisture decline in early 2020, only STC-CNNLSTM correctly identified the FD, with all other models failing to classify it. The proposed model also captured the slow drought events from 2016 to 2019. However, during the severe fluctuation in late 2021, where FD closely followed a slow drought, STC-CNNLSTM merged the two events into a single prolonged slow drought. Minor deviations near the threshold can lead to substantial differences in drought categorization, underscoring the inherent difficulty of early-warning applications. Overall, the STC-CNNLSTM demonstrates clear advantages over TC-CNNLSTM, CNN-LSTM, and the baseline LSTM in capturing both slow and rapid drought developments.

As shown in Table 1, the classification accuracy of all four models closely aligns with their regression performance illustrated in Fig. 4. The STC-CNNLSTM model achieved the highest overall accuracy (0.91) in classifying all days in the test dataset across event categories. However, its advantage is modest, with the other models achieving accuracies of 0.90, 0.86, and 0.80, respectively. The predominance of the no drought category, which accounts for 80.89 % of test days, strongly influences the overall accuracy. Even the baseline LSTM model performs well in this category, reaching an F1 score of 0.91 when extreme soil moisture values are absent. The most pronounced differences among the models emerge in the classification of FD events.

Notably, the baseline LSTM model performs poorly in detecting the onset of FDs, with all three metrics near 0.02, indicating that nearly all (98 %) FD onset days are misclassified, and almost all (98 %) predicted onset days are false positives. Such unreliable predictions are inappropriate for early warning applications. In contrast, the STC-CNNLSTM model demonstrates substantially improved performance for FD events, correctly identifying approximately 28 % of onset days, with a positive predictive value of 47 % for this category. The TC-CNNLSTM exhibits lower classification performance than the full model, with the benchmark CNN-LSTM's performance falling between that of the TC-CNNLSTM and the baseline. Notably, differences in continuous soil moisture prediction do not translate linearly to the classification accuracy, as regression and classification tasks capture distinct aspects of model performance and are evaluated using different metrics (Chicco and Jurman, 2020).

The distribution of model predictions, as shown in Fig. S7, provides further insights into classification performance. Notably, the prediction distributions for the lowest two percentile groups (i.e., drought days with soil moisture below the 20th percentile) differ substantially among the four models. The two causality-informed models effectively avoid generating unreasonable outlier predictions in this range. Across all models, there is a tendency to overestimate low soil moisture values and underestimate high soil moisture values. However, both the error magnitude and variance are greater when predicting low soil moisture, underscoring a common limitation of these models in accurately capturing drought conditions compared to periods of high soil moisture.

The prediction lead time has a clear influence on model performance (Bommer et al., 2025). As shown in Table S3, skill decreases steadily as the forecast horizon extends from 1 to 4 weeks. At a 7-day lead time, classification and regression performance both improve relative to the 14-day baseline, with notably higher skill for FD onset (F1 score = 0.402) and FD condition (F1 score = 0.507), reflecting stronger short-term predictability of rapid soil-moisture changes. Regression performance also strengthens (NSE = 0.872). Beyond 14 days, classification skill declines sharply, particularly for FD onset (F1 score = 0.011), consistent with the limited sub-seasonal predictability of the atmospheric drivers that trigger flash droughts (Pendergrass et al., 2020). Regression metrics deteriorate more gradually (NSE from 0.829 at 14 days to 0.312 at 28 days), indicating that overall soil-moisture evolution remains predictable even as precise onset timing becomes harder to capture. Accordingly, a 2-week lead time is adopted as a practical compromise, balancing operational reaction time with acceptable

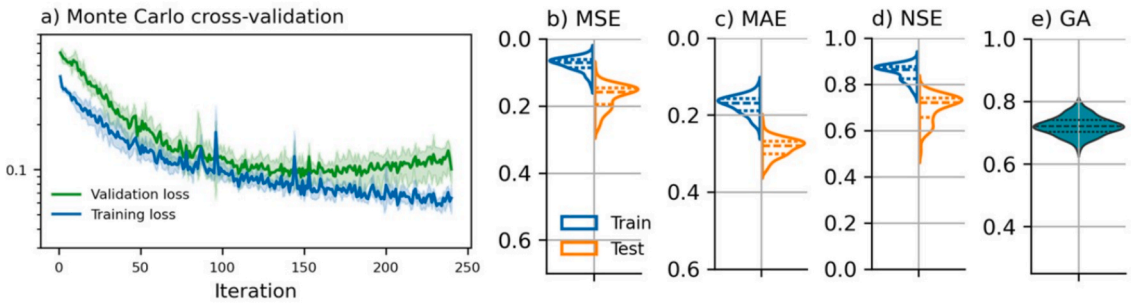

Fig. 5. (a) Training and validation loss curves from five random spatial resamplings; the solid lines represent the mean with a 95% confidence interval. (b–d) Model performance metrics (MSE, MAE, and NSE) on training and test sets. (e) Spatial generalization capability of the proposed model; dashed lines within the violin plots represent the quartiles of the kernel density estimation.

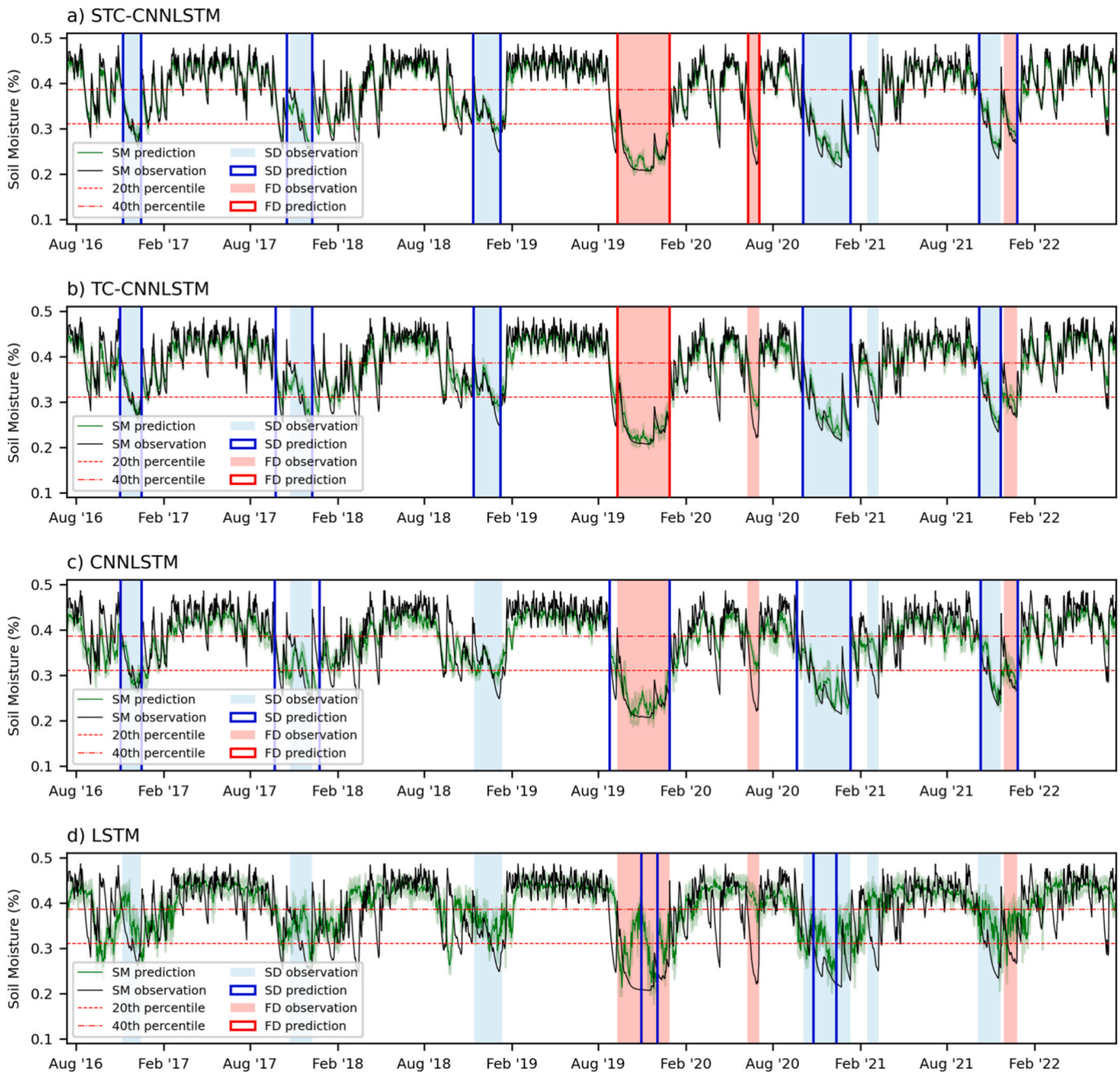

Fig. 6. Typical soil moisture dynamics and model predictions at a representative grid cell (23.0° N, 112.3° E) over six years in the test set. The proposed model better captures extreme soil moisture fluctuations compared with other methods.

predictive skill.

4.4. *Improvement of predictive performance through causality*

To investigate how spatial–temporal causality interacts with model learning across the feature dimension, we compared the CCM-based causality degree for each driver (Fig. S8) with feature-importance scores derived from expected gradients (Fig. S9). Although the overall distributions of causality degree and feature importance do not directly overlap, their relationship becomes evident when examined through model performance at individual pixel cells. Specifically, Fig. 7a–e show that drivers identified as important by the model (e.g., downward longwave radiation, precipitation, soil moisture memory) also exhibit a positive and significant association between causality degree and prediction skill (NSE). Correlation strengths vary across drivers (r = 0.216–0.538), reflecting their different physical roles in flash-drought evolution and the varying degree to which soil moisture responds to each driver across the landscape. In contrast, less influential drivers show no such relationship (Fig. S10).

These results support our hypothesis that feature importance modulates the coupling between causality and model performance. The causality degree represents the intrinsic strength of driver–response interactions (i.e., whether changes in a variable meaningfully propagate to soil moisture), while feature importance captures how sensitively the deep learning model relies on that driver. The coupling therefore emerges when (i) a driver physically affects soil moisture, and (ii) the model identifies this driver as useful for prediction. Fig. 7f quantifies this effect by correlating the causality–performance relationships from Fig. 7a–e and S8 with model-derived feature importance, showing a strong and significant correlation ($r = 0.724$, $p = 0.018$). This confirms

Table 1
Model performance on daily event classification in the test dataset. Days are categorized into five groups according to drought event occurrence and phase: ND (non-drought), SD (slow drought), and FD (flash drought). Performance metrics are summarized as mean ± standard deviation across 10 training periods with different random seeds.

Models	Metrics	ND	SD onset	SD condition	FD onset	FD condition
STC-CNNLSTM	Precision	0.973 ± 0.006	0.596 ± 0.010	0.662 ± 0.003	0.469 ± 0.044	0.572 ± 0.036
	Recall	0.979 ± 0.003	0.695 ± 0.046	0.735 ± 0.032	0.280 ± 0.032	0.382 ± 0.031
	F1-score	**0.976 ± 0.003**	**0.622 ± 0.026**	**0.679 ± 0.016**	**0.330 ± 0.036**	**0.428 ± 0.031**
	Accuracy	0.911 ± 0.005				
TC-CNNLSTM	Precision	0.969 ± 0.005	0.529 ± 0.021	0.616 ± 0.017	0.341 ± 0.041	0.435 ± 0.015
	Recall	0.974 ± 0.002	0.653 ± 0.007	0.722 ± 0.015	0.175 ± 0.035	0.260 ± 0.039
	F1-score	**0.971 ± 0.002**	**0.562 ± 0.015**	**0.646 ± 0.015**	**0.217 ± 0.038**	**0.299 ± 0.031**
	Accuracy	0.896 ± 0.004				
CNNLSTM	Precision	0.939 ± 0.002	0.421 ± 0.020	0.541 ± 0.007	0.174 ± 0.012	0.239 ± 0.020
	Recall	0.965 ± 0.005	0.512 ± 0.038	0.557 ± 0.050	0.066 ± 0.006	0.125 ± 0.029
	F1-score	**0.952 ± 0.002**	**0.441 ± 0.025**	**0.524 ± 0.022**	**0.090 ± 0.009**	**0.151 ± 0.032**
	Accuracy	0.860 ± 0.003				
LSTM	Precision	0.863 ± 0.012	0.190 ± 0.018	0.248 ± 0.017	0.020 ± 0.006	0.216 ± 0.016
	Recall	0.966 ± 0.009	0.120 ± 0.026	0.078 ± 0.021	0.018 ± 0.006	0.128 ± 0.052
	F1-score	**0.911 ± 0.003**	**0.132 ± 0.023**	**0.107 ± 0.025**	**0.017 ± 0.006**	**0.146 ± 0.038**
	Accuracy	0.801 ± 0.003				
Proportion		80.89 %	6.11 %	7.15 %	1.09 %	4.75 %

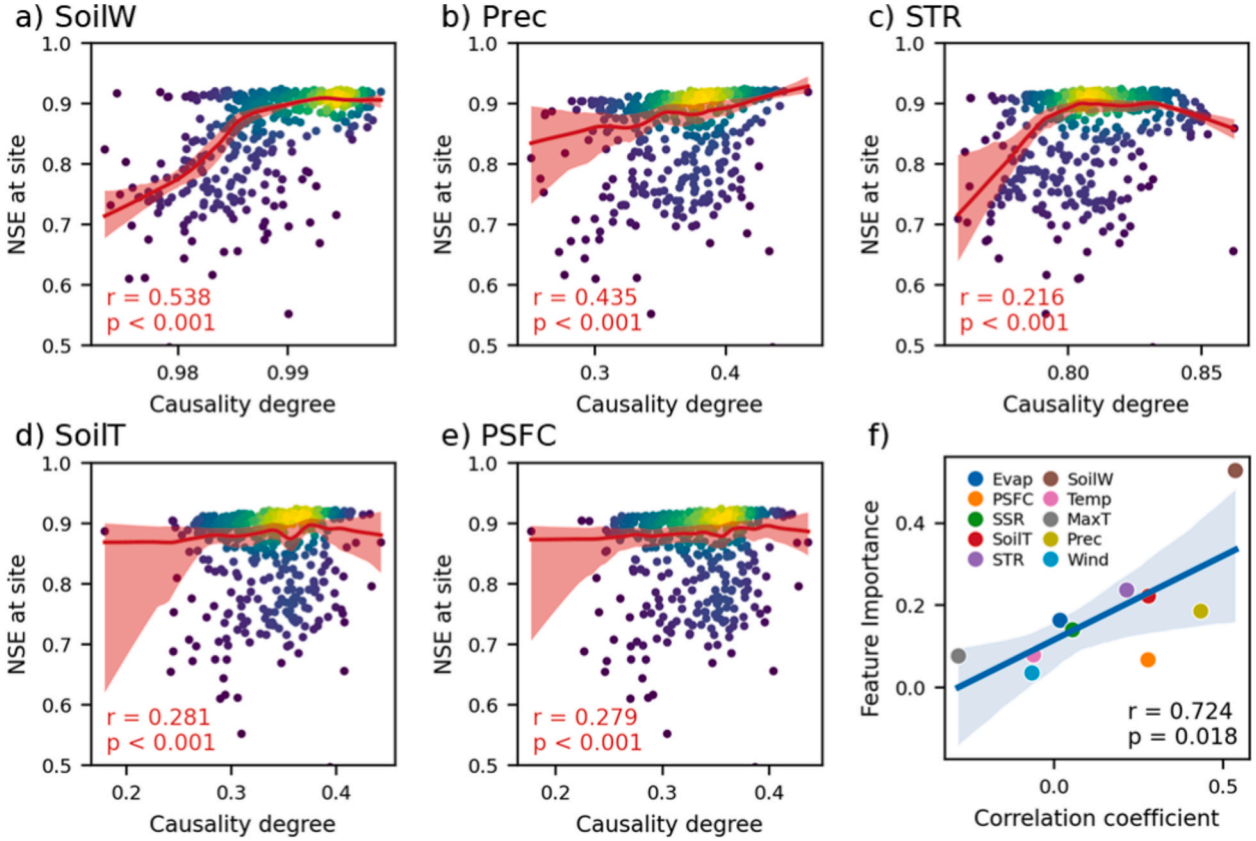

Fig. 7. Causality enhances model performance. (a–e) illustrate the relationship between causality degree (x-axis) and model performance, measured by NSE (y-axis), for the five primary driving features. Each point represents one pixel cell within the study area (n = 470). Color gradients represent local point density. Each panel reports the Pearson correlation coefficient and the corresponding p-value, with locally weighted scatterplot smoothing (LOWESS) of the points and 95 % confidence interval from 200 bootstraps. Panel (f) presents the strength of this correlation (x-axis) in relation to the feature importance (y-axis) of the drivers.

that improvements in predictive skill arise not simply from adding more information, but from integrating physically meaningful causal signals that align with the model's learned sensitivity.

This amplification effect helps explain why the proposed spatial–temporal causality module enhances predictive performance. Traditional deep learning models behave deterministically and do not explicitly account for whether input variables exert real physical influence on the target. By incorporating causal information, the model gains an additional dimension, analogous to providing context or confidence weighting, that increases robustness against noisy or weakly informative signals. In this way, causality guides the learning process toward physically meaningful patterns, enabling more reliable identification of rapid changes associated with flash drought onset.

4.5. Mechanisms behind flash drought onset

Leveraging the model's ability to accurately fit soil moisture and detect drought events, EG-based feature importance was further analyzed to explore the mechanisms underlying the onset of flash and slow droughts. Fig. 8 highlights the most influential features for soil moisture prediction. Despite notable uncertainty and bias in ERA5 soil moisture products (Kim et al., 2020; Yang et al., 2020; Zheng et al., 2024), antecedent soil moisture remains a critical predictor of future soil water content, reflecting the well-established concept of soil moisture memory (Rahmati et al., 2024). Additionally, soil temperature, a key indicator of soil conditions, shows a strong and consistent correlation with soil moisture, as confirmed by previous studies (Lakshmi et al., 2003; Riveros-Iregui et al., 2007) and its effective application in soil moisture modeling (Dong et al., 2016).

Precipitation and evaporation are the primary direct drivers of soil moisture through water input and loss, exerting a strong influence on soil moisture prediction (note: low Evap values indicate evaporation, while high values indicate condensation) (Jia et al., 2018; Rushton et al., 2006). Surface thermal radiation (STR) and surface solar radiation (SSR) also play significant roles in soil moisture dynamics, with STR emerging as more influential than SSR. This is consistent with the subtropical, humid climate of China's GBA, which receives an average annual precipitation of approximately 2,000 mm.

Our results show that STR has a strong positive association with soil moisture. Although high downward STR typically indicates less nocturnal heat loss and can increase nighttime evaporation (Heck et al., 2020; Padrón et al., 2020), evaporation is governed by multiple factors, including net surface energy balance, vapor pressure gradient, wind, and humidity (Condie and Webster, 1997; Wang and Bras, 2001). In this study, high STR often corresponds to cloudy, humid, and rainy conditions (Fig. 8c), which limits daytime evaporation due to reduced shortwave radiation under cloud cover (Fig. 8d). Conversely, low STR is generally associated with clear skies, allowing for greater incoming solar radiation, higher surface temperatures, and increased evaporation during the day, as well as higher vapor pressure deficits that enhance evaporative demand. These mechanisms underscore the prominent role of STR in soil moisture regulation in humid regions (Heck et al., 2020; Naud et al., 2013).

SSR, on the other hand, exhibits a more complex relationship with soil moisture. Lower SSR can both increase and decrease soil moisture (Fig. 8b). Notably, episodes of high SSR combined with heavy precipitation, which is common in China's GBA, can enhance soil moisture, as high SSR can promote atmospheric instability and storm development (Huang et al., 2019). Following convective storms, skies may clear rapidly, resulting in high SSR after rainfall events (Lin et al., 2023; Yang et al., 2024).

Overall, these findings highlight the greater importance of radiation variables over temperature (Temp, MaxT) and pressure (PSFC, Wind) indices in humid climates. They also suggest that models focusing solely on shortwave radiation may underestimate the role of radiative energy in soil moisture dynamics in such regions.

Fig. 9 presents the average feature importance scores for both drought types, aggregated across all historical events and grid cells. For slow droughts, which typically develop over weeks to months and are often driven by prolonged rainfall deficits and cumulative radiation stress (Gibson et al., 2020), the initial soil moisture status emerges as a key predictor. Dry antecedent soils increase the risk of slow drought persistence or onset, as strong soil moisture memory causes the system to respond sluggishly, with past soil moisture conditions dominating future evolution (He et al., 2023; Rahmati et al., 2024). Persistent incoming shortwave radiation (SSR) is another critical factor, gradually depleting soil moisture over time.

In contrast, FDs are characterized by abrupt declines in soil moisture, typically occurring under conditions of high radiation, clear skies, elevated soil temperatures, and high vapor pressure deficit (Otkin et al., 2018). STR plays a crucial role by modulating atmospheric thermal insulation and surface energy balance, particularly at night (Small and Kurc, 2003). Warmer soils accelerate evapotranspiration and drying, especially in the root zone. As negative soil moisture anomalies suppress evapotranspiration, a larger fraction of available energy is converted to sensible rather than latent heat flux, further increasing soil temperatures. This "drier-hotter" feedback, characterized by rapid energy exchange with the atmosphere, intensifies extreme high-temperature and

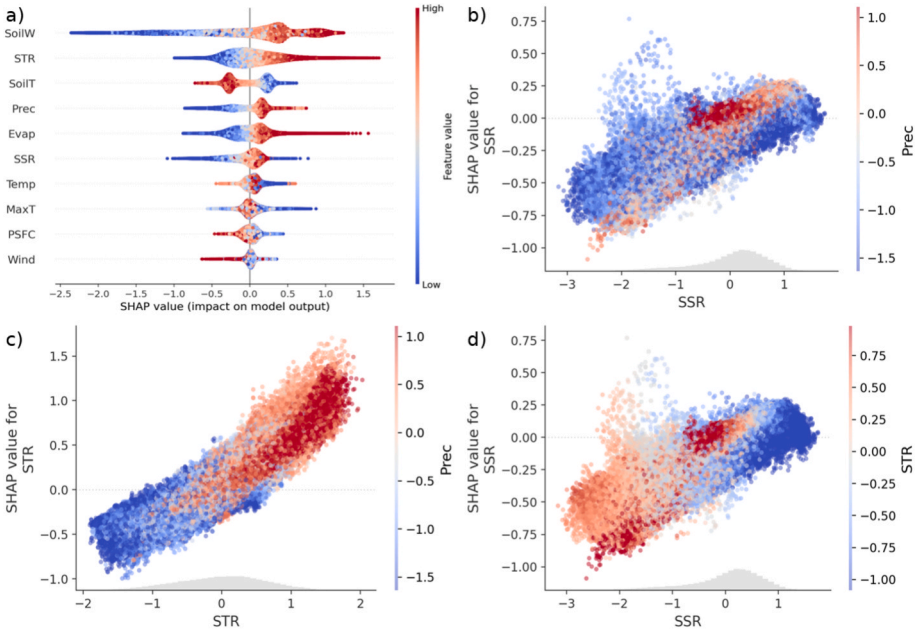

Fig. 8. SHAP-based feature importance of various drivers for soil moisture prediction. Each dot represents the model explanation for a single day. (a) displays the overall distribution of feature importance and feature values, with features ranked by the mean absolute SHAP value. (b–d) show dependence plots, illustrating the effect of individual features on model predictions. Dot colors indicate another feature, highlighting non-linear interaction effects among precipitation (Prec), surface solar radiation (SSR), and surface thermal radiation (STR). (Note: Positive vertical fluxes indicate downward direction.).

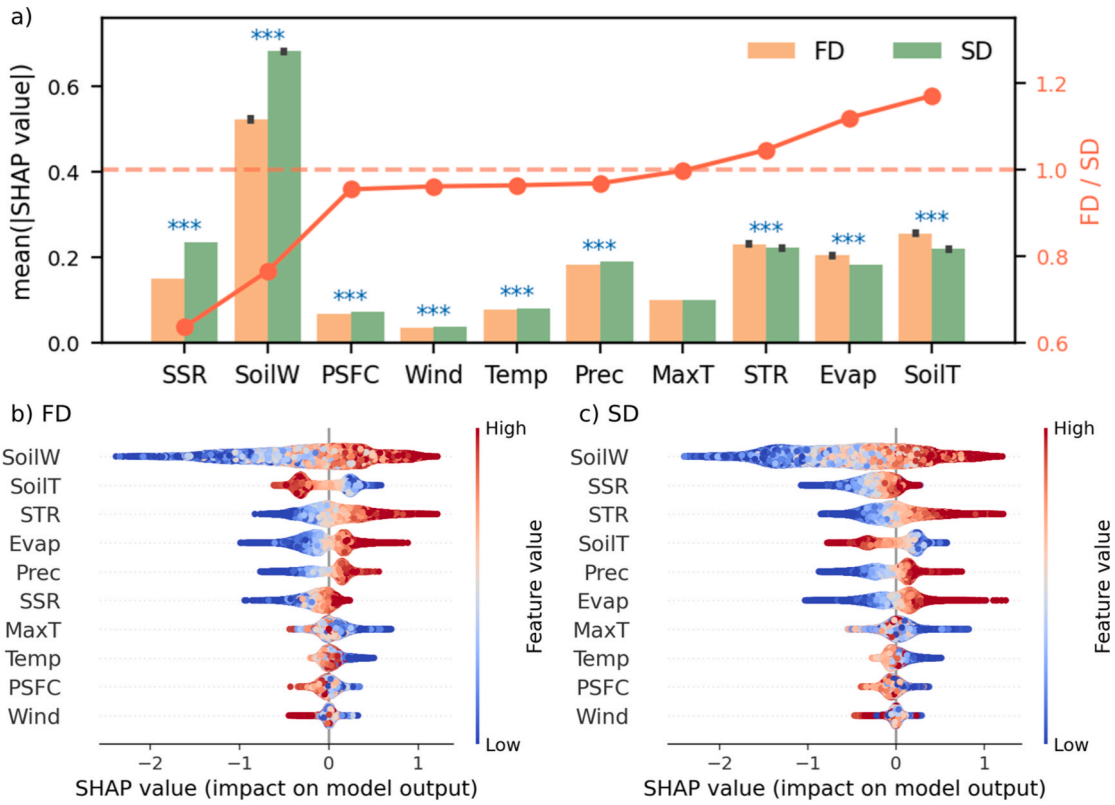

Fig. 9. (a) Feature importance of different drivers for the onset of slow and flash droughts. Dots above the orange dashed line indicate drivers with greater influence on flash drought onset, while those below indicate drivers more influential for slow drought onset. Error bars represent 95% confidence intervals across all events (*** denotes a significant difference at the 0.001 level). (b–c) SHAP-based feature importance of various drivers for soil moisture: overall distributions of feature importance versus feature value during flash drought and slow drought onset, respectively, with features ranked by the mean absolute SHAP value. Each dot represents the explanation for a single onset day. (Note: Positive vertical fluxes are directed downward.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

FD events (Ford and Quiring, 2014; Hauser et al., 2016; Miralles et al., 2014).

4.6. *Drought complexity*

Previous studies have highlighted the accelerating transition from slow to FDs based on global reanalysis and assimilation datasets (Qing et al., 2022; Yuan et al., 2023)—a trend that is also evident locally in China's GBA, as shown in Fig. 3. However, our analysis of drought complexity—defined as the number of primary drivers present during drought onset days—reveals a more heterogeneous spatial distribution across the region. Notably, areas with the highest complexity are concentrated in the core of China's GBA, particularly in the built-up zones surrounding the Pearl River estuary. Mechanistic analyses indicate that complex interactions among multiple drivers can compound drought impacts, thereby accelerating the onset process. Our results further demonstrate that higher drought complexity is associated with increased underestimation of drought events (Fig. 10b), consistent with findings from analogous research on flooding (Jiang et al., 2024).

Additional experiments incorporating population density data (Fig. S11) reinforce the link between human activity and drought complexity (Fig. 10c). Urban areas—characterized by impervious

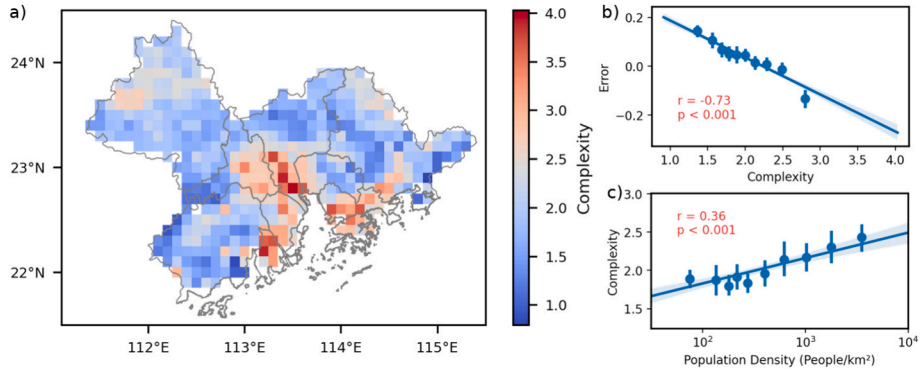

Fig. 10. (a) Spatial distribution of drought complexity, defined as the average number of primary drivers (80th percentile) present during historical drought onset days. (b–c) Relationships among drought complexity, population density, and prediction error, aggregated at the grid-cell level. Data are evenly divided into ten bins based on the x-axis variable; regression lines are fitted to the original data, with shaded areas representing the 95% confidence interval.

surfaces, urban heat island effects, and altered surface energy balances—exhibit significantly modified land–atmosphere interactions. These factors introduce multiple concurrent drivers affecting soil moisture, thereby increasing the complexity of droughts in these regions (Huang et al., 2024). Collectively, these results suggest that urbanization and anthropogenic surface modifications amplify the multidimensionality of drought mechanisms, increasing both the difficulty of drought prediction and the vulnerability of urban areas to hydrological extremes.

It should be noted that these findings are not sensitive to the specific threshold used to define drought complexity. We conducted additional analyses by varying the percentile cutoff from the 60th to the 90th percentiles. The results (Figs. S12–S14) show that, while the absolute number of main drivers decreases as the threshold increases, the relationships between predictive error and drought complexity, as well as between population density and drought complexity, remain stable. This confirms the robustness of our conclusions. Furthermore, replacing population density with land cover as an indicator of human activity yields consistent results (Fig. S15), reinforcing the link between anthropogenic influence and increased drought complexity.

4.7. Discussion

While more advanced regularization strategies (e.g., weight decay optimization, ensemble averaging) could further improve predictive accuracy, such techniques are beyond the scope of this study. To ensure fairness across model comparisons, we applied only standard methods—dropout layers and early stopping—consistently to all models, thereby confirming that the observed improvements are attributable to the proposed framework.

It should be noted that the improvements observed in the STC-CNNLSTM are likely attributable to the combined effect of incorporating CCM-derived causal information and the skip connection mechanism. The former provides physically interpretable guidance for initializing and updating cell states, while the latter facilitates efficient information flow across layers. Since our objective is to evaluate the integrated framework, rather than to disentangle the relative contributions of these components, we did not isolate CCM information as static features in a separate baseline. Additional baselines, deeper gating analyses, or attention-specific ablations could provide complementary insights, these represent separate research directions and constitute a promising direction for future work. Our current design and ablation tests are intended to support the practical contribution of embedding causal information directly into the recurrent state initialization for FD prediction.

While the proposed STC-CNNLSTM framework shows promising capability for flash-drought prediction, several limitations should be noted. In this study, ERA5-Land was selected as the target dataset due to its long-term temporal coverage (1950–present) and relatively fine spatial resolution (0.1°), both of which are critical for causal analysis and data-demanding deep learning models. Alternative products such as SMAP or GLDAS-Noah lack either sufficient temporal length or comparable spatial detail, and in situ observations are too sparse in the study area to support spatially distributed model evaluation. However, the ERA5-Land soil moisture product inherits uncertainties from model parameterizations and sparse observational coverage in certain regions. Soil moisture products are subject to retrieval biases related to vegetation, surface heterogeneity, and sensor limitations (Kim et al., 2020; Yang et al., 2020; Zheng et al., 2024), which may propagate into the model and affect causality detection. These limitations highlight the need for future work that combines reanalysis with in-situ observations, leverages higher-resolution or bias-corrected soil moisture datasets, and evaluates the proposed framework under diverse climatic and land-cover conditions. As the model is designed to reproduce ERA5-Land soil moisture dynamics rather than reconcile among different soil moisture sources, we did not conduct cross-product validation in this study, which would primarily reflect inherent inter-product disagreement rather than model performance.

Although the F1 score for FD onset prediction remains modest in absolute terms (0.33), it represents a more than fifteen-fold improvement compared to the baseline (0.02). This relative gain is particularly meaningful given that FDs are rare and extreme events that inherently challenge predictive models. The achieved precision of approximately 0.5 indicates that about half of the model-issued onset warnings correspond to actual events, which, from an operational perspective, can provide valuable early warning information. We further developed a simple decision-analytic framework that evaluates how each model performs under varying trade-offs between false alarms and missed detections. Based on the classification results of each model on the test dataset, we quantified the expected cost per decision as

$$E(\text{Cost}) = c_{FP}(1-p)FPR + c_{FN}pFNR \tag{20}$$

where p is the event prevalence, c_{FP} and c_{FN} represent the assumed costs of a false alarm and a missed detection, and FPR and FNR are the false-positive and false-negative rates. Because true economic losses of FDs are difficult to quantify, we varied the cost ratio c_{FP}/c_{FN} over a range and computed the corresponding normalized expected cost for each model. Fig. S17 provides a direct visualization of how the relative value of each model changes under different operational priorities, highlighting the advantage of the proposed STC-CNNLSTM model. Across the full range of cost ratios examined, STC-CNNLSTM consistently yielded the lowest expected cost.

Specifically, when the cost of a false alarm is small relative to the cost of missing a FD onset (cost ratio = 0.01), STC-CNNLSTM reduces the expected cost by 12.70 %, 22.89 %, and 27.05 % compared with the other models. This scenario reflects decision contexts in which avoiding missed detections is critical because missed onsets can lead to rapid crop stress, yield loss, and diminished capacity for timely irrigation adjustments. Even when false alarms are strongly penalized (cost ratio = 10), STC-CNNLSTM still achieves the lowest expected cost, providing reductions of 7.62 %, 3.87 %, and 61.08 % relative to the others. These results indicate that STC-CNNLSTM offers consistently favorable economic performance across diverse operational priorities, whether decision-makers aim to minimize missed onsets or unnecessary actions.

In practice, forecasts of extreme events are often combined with monitoring indicators and decision-support tools. The demonstrated ability of the proposed framework to anticipate FD onset makes it suitable for integration into operational early-warning platforms, where even short lead-time improvements can create critical windows for action given the rapid development of FDs. Enhanced early warnings could support drought-risk management measures such as irrigation scheduling, crop stress mitigation, and contingency water allocation in agriculturally intensive and rapidly urbanizing regions such as the GBA. Furthermore, the identification of downward longwave radiation and soil-moisture memory as key predictors provides actionable guidance for optimizing observational networks, highlighting the value of prioritizing radiation flux and soil monitoring in humid subtropical climates. Although further improvements are needed to increase recall and reduce false alarms, the present findings show that the proposed framework not only improves predictive skill but also enhances operational utility, which is an essential requirement for effective early-warning systems and adaptive agricultural strategies.

Despite the demonstrated advantage of the proposed framework, several limitations should be acknowledged. First, the model is trained exclusively on data from the GBA, and its performance relies on the region's specific hydro-meteorological conditions and land–atmosphere interactions. As a result, the framework is not guaranteed to transfer directly to other humid subtropical regions without retraining or fine-tuning, especially where radiation regimes, land-surface properties, or soil–vegetation feedback differ. Second, temporal generalization remains an inherent challenge under evolving climate conditions. Although our temporally ordered train–validation–test split and recent

event analyses indicate reasonable robustness, long-term non-stationarity (e.g., shifts in temperature baselines and precipitation patterns) may gradually reduce model skill. Therefore, operational deployment would require periodic model updating or online learning using the most recent observations to account for changing background climate states. Addressing these spatial and temporal transferability issues represents an important avenue for future research.

5. Conclusions

To facilitate effective FD risk management and adaptation strategies, this study introduces a novel deep learning framework—the STC-CNNLSTM model—to enhance soil moisture FD prediction. This framework utilizes a CNN-LSTM backbone to capture the dependencies of soil moisture on spatial–temporal meteorological variables. A novel CCM-based spatial–temporal causality module is integrated into the conventional LSTM architecture, enabling strategic weighting of feature dimensions by initiating and controlling cell states.

The study demonstrates that: (a) the proposed causality module acts as a regularization technique, improving both generalization (GA = 0.90) and predictive performance (NSE = 0.83 on the test set) over alternative models; (b) the STC-CNNLSTM model excels at capturing soil moisture extremes, enabling more effective FD forecasts, as shown by an F1 score of 0.33 for FD onset versus 0.02 for the baseline model; (c) XAI-based interpretation reveals that the degree of causality significantly enhances performance for the most important drivers, supporting the inclusion of the causality module; (d) key environmental variables, including soil moisture memory, radiation components (especially downward longwave radiation, STR), and precipitation, are critical predictors of soil moisture in the humid GBA; (e) slow droughts are primarily influenced by initial soil moisture and persistent shortwave radiation, while FDs are driven by rapid energy imbalances involving STR and hot, dry conditions; and (f) anthropogenic activities in China's GBA may amplify the multidimensionality of drought mechanisms, increasing both prediction difficulty and the vulnerability of these regions to hydrological extremes.

Research Data

The ERA5-Land reanalysis data are obtained from European Centre for Medium-Range Weather Forecasts (ECMWF) at https://cds.climate.copernicus.eu/.

All code in this study is in Python. The core code for model building and training can be found in a Zenodo repository at https://doi.org/10.5281/zenodo.16554347. The calculations for the CCM causality were performed using the causal-ccm package (Javier, 2021). Model realizations are based on PyTorch package (Paszke et al., 2019); the XAI-based model interpretation is performed using the SHAP package (Lundberg and Lee, 2017); all visualizations are completed by Matplotlib (Hunter, 2007) and Seaborn (Waskom, 2021) packages. All packages can be found via Python Package Index, the official repository of software for the Python programming language, at https://pypi.org/.

CRediT authorship contribution statement

Sijie Tang: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Shuo Wang:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Jiping Jiang:** Writing – review & editing, Supervision. **Yi Zheng:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Acknowledgments

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU/RGC 15232023), the Otto Poon Research Institute for Climate-Resilient Infrastructure (Project No. P0055919), and the Hong Kong Polytechnic University (Project No. P0045957). Additional support was provided by the High-level University Special Fund (Grant No. G030290001).

The computational resources in this study were supported by the Center for Computational Science and Engineering at Southern University of Science and Technology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2026.134945.

Data availability

Data will be made available on request.

References

AghaKouchak, A., 2014. A baseline probabilistic drought forecasting framework using standardized soil moisture index: application to the 2012 United States drought. Hydrol. Earth Syst. Sci. 18 (7), 2485–2492.

Ahmadianfar, I., Shirvani-Hosseini, S., Samadi-Koucheksaraee, A., Yaseen, Z.M., 2022. Surface water sodium (Na+) concentration prediction using hybrid weighted exponential regression model with gradient-based optimization. Environ. Sci. Pollut. Res. 29 (35), 53456–53481.

Ahmadianfar, I., Zamani, R., 2020. Assessment of the hedging policy on reservoir operation for future drought conditions under climate change. Clim. Change 159 (2), 253–268.

Anderson, M.C., Hain, C., Otkin, J., Zhan, X., Mo, K., Svoboda, M., Wardlow, B., Pimstein, A., 2013. An intercomparison of drought indicators based on thermal remote sensing and NLDAS-2 simulations with US Drought Monitor classifications. J. Hydrometeorol. 14 (4), 1035–1056.

Bahri, Y., Dyer, E., Kaplan, J., Lee, J., Sharma, U., 2024. Explaining neural scaling laws. Proc. Natl. Acad. Sci. 121 (27), e2311878121.

Balles, L., Romero, J. and Hennig, P. 2016. Coupling adaptive batch sizes with learning rates. arXiv preprint arXiv:1612.05086.

Bengio, Y. (2012) Neural networks: Tricks of the trade: Second edition, pp. 437-478, Springer.

Benkő, Z., Varga, B., Stippinger, M. and Somogyvári, Z. 2025. Detecting causality in the frequency domain with Cross-Mapping Coherence. Physica D: Nonlinear Phenomena, 134708.

Bishop, C.M. and Nasrabadi, N.M. (2006) Pattern Recognition and Machine Learning, Springer.

Bommer, P.L., Kretschmer, M., Spuler, F.R., Bykov, K. and Höhne, M.M.-C. 2025. Deep Learning Meets Teleconnections: Improving S2S Predictions for European Winter Weather. arXiv preprint arXiv:2504.07625.

Brubaker, K.L., Entekhabi, D., 1996. Analysis of feedback mechanisms in land-atmosphere interaction. Water Resour. Res. 32 (5), 1343–1357.

Cai, H., Liu, S., Shi, H., Zhou, Z., Jiang, S., Babovic, V., 2022. Toward improved lumped groundwater level predictions at catchment scale: Mutual integration of water balance mechanism and deep learning method. J. Hydrol. 613, 128495.

Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L., Xue, X., 2019. Research on soil moisture prediction model based on deep learning. PLoS One 14 (4), e0214508.

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., Miao, Y., 2021. Review of image classification algorithms based on convolutional neural networks. Remote Sens. (Basel) 13 (22), 4712.

Chen, L.G., Gottschalck, J., Hartman, A., Miskus, D., Tinker, R., Artusa, A., 2019. Flash drought characteristics based on US drought monitor. Atmos. 10 (9), 498.

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21, 1–13.

Christian, J.I., Basara, J.B., Hunt, E.D., Otkin, J.A., Furtado, J.C., Mishra, V., Xiao, X., Randall, R.M., 2021. Global distribution, trends, and drivers of flash drought occurrence. Nat. Commun. 12 (1), 6330.

Christian, J.I., Martin, E.R., Basara, J.B., Furtado, J.C., Otkin, J.A., Lowman, L.E., Hunt, E.D., Mishra, V., Xiao, X., 2023. Global projections of flash drought show increased risk in a warming climate. Commun. Earth Environ. 4 (1), 165.

Condie, S.A., Webster, I.T., 1997. The influence of wind stress, temperature, and humidity gradients on evaporation from reservoirs. Water Resour. Res. 33 (12), 2813–2822.

Crausbay, S.D., Ramirez, A.R., Carter, S.L., Cross, M.S., Hall, K.R., Bathke, D.J., Betancourt, J.L., Colt, S., Cravens, A.E., Dalton, M.S., 2017. Defining ecological drought for the twenty-first century. Bull. Am. Meteorol. Soc. 98 (12), 2543–2550.

Datta, P., Faroughi, S.A., 2023. A multihead LSTM technique for prognostic prediction of soil moisture. Geoderma 433, 116452.
Dikshit, A., Pradhan, B., Alamri, A.M., 2021. Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. Sci. Total Environ. 755, 142638.
Dong, J., Steele-Dunne, S.C., Ochsner, T.E., Van De Giesen, N., 2016. Determining soil moisture and soil properties in vegetated areas by assimilating soil temperatures. Water Resour. Res. 52 (6), 4280–4300.
Erion, G., Janizek, J.D., Sturmfels, P., Lundberg, S.M., Lee, S.-I., 2021. Improving performance of deep learning models with axiomatic attribution priors and expected gradients. Nat. Mach. Intell. 3 (7), 620–631.
Feng, J., Li, J., Xu, C.Y., Wang, Z., Zhang, Z., Wu, X., Lai, C., Zeng, Z., Tong, H., Jiang, S., 2024. Viewing soil moisture flash drought onset mechanism and their changes through XAI lens: a case study in eastern China. Water Resour. Res. 60 (6), e2023WR036297.
Ford, T.W., Quiring, S.M., 2014. In situ soil moisture coupled with extreme temperatures: a study based on the Oklahoma Mesonet. Geophys. Res. Lett. 41 (13), 4727–4734.
Gao, P., Qiu, H., Lan, Y., Wang, W., Chen, W., Han, X., Lu, J., 2021. Modeling for the prediction of soil moisture in litchi orchard with deep long short-term memory. Agriculture 12 (1), 25.
Gibson, A.J., Verdon-Kidd, D.C., Hancock, G.R., Willgoose, G., 2020. Catchment-scale drought: capturing the whole drought cycle using multiple indicators. Hydrol. Earth Syst. Sci. 24 (4), 1985–2002.
Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y. (2016) Deep Learning, MIT press Cambridge.
Gorgij, A.D., Askari, G., Taghipour, A., Jami, M., Mirfardi, M., 2023. Spatiotemporal forecasting of the groundwater quality for irrigation purposes, using deep learning method: long short-term memory (LSTM). Agric Water Manag 277, 108088.
Grus, J. (2019) Data science from scratch: first principles with python, O'Reilly Media.
Han, Z., Zhao, J., Leung, H., Ma, K.F., Wang, W., 2019. A review of deep learning models for time series prediction. IEEE Sens. J. 21 (6), 7833–7848.
Hauser, M., Orth, R., Seneviratne, S.I., 2016. Role of soil moisture versus recent climate change for the 2010 heat wave in western Russia. Geophys. Res. Lett. 43 (6), 2819–2826.
He, F., Liu, T. and Tao, D. 2019. Control batch size and learning rate to generalize well: Theoretical and empirical evidence. Advances in neural information processing systems 32.
He, K., Zhang, X., Ren, S. and Sun, J. 2016 Deep residual learning for image recognition, pp. 770-778.
He, Q., Lu, H., Yang, K., 2023. Soil moisture memory of land surface models utilized in major reanalyses differ significantly from SMAP observation. Earth's Future 11 (4), e2022EF003215.
Heck, K., Coltman, E., Schneider, J., Helmig, R., 2020. Influence of radiation on evaporation rates: a numerical analysis. Water Resour. Res. 56 (10), e2020WR027332.
Ho, S., Tian, L., Disse, M., Tuo, Y., 2021. A new approach to quantify propagation time from meteorological to hydrological drought. J. Hydrol. 603, 127056.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hoedt, P.-J., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G.S., Hochreiter, S. and Klambauer, G. 2021 Mc-lstm: Mass-conserving lstm, pp. 4275-4286, PMLR.
Huang, G., Li, Z., Li, X., Liang, S., Yang, K., Wang, D., Zhang, Y., 2019. Estimating surface solar irradiance from satellites: past, present, and future perspectives. Remote Sens. Environ. 233, 111371.
Huang, S., Wang, S., Gan, Y., Wang, C., Horton, D.E., Li, C., Zhang, X., Niyogi, D., Xia, J., Chen, N., 2024. Widespread global exacerbation of extreme drought induced by urbanization. Nat. Cities 1 (9), 597–609.
Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9 (03), 90–95.
Huntingford, C., Jeffers, E.S., Bonsall, M.B., Christensen, H.M., Lees, T., Yang, H., 2019. Machine learning and artificial intelligence to aid climate change research and preparedness. Environ. Res. Lett. 14 (12), 124007.
Hwang, J. 2020. Modeling Financial Time Series using LSTM with Trainable Initial Hidden States. arXiv preprint arXiv:2007.06848.
Indu, J., Nair, A.S., Pradhan, A., Mangla, R., Krishnan, S., Verma, K. and Huggannavar, V. (2022) Radar Remote Sensing, pp. 123-148, Elsevier.
Jamei, M., Ahmadianfar, I., Karbasi, M., Malik, A., Kisi, O., Yaseen, Z.M., 2023. Development of wavelet-based kalman online sequential extreme learning machine optimized with boruta-random forest for drought index forecasting. Eng. Appl. Artif. Intel. 117, 105545.
Javier, P.J.E. 2021 causal-ccm a Python implementation of Convergent Cross Mapping.
Jia, B., Liu, J., Xie, Z., Shi, C., 2018. Interannual variations and trends in remotely sensed and modeled soil moisture in China. J. Hydrometeorol. 19 (5), 831–847.
Jiang, S., Tarasova, L., Yu, G., Zscheischler, J., 2024. Compounding effects in flood drivers challenge estimates of extreme river floods. Sci. Adv. 10 (13), eadl4005.
Jiang, S., Zheng, Y., Solomatine, D., 2020. Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning. Geophys. Res. Lett. 47 (13), e2020GL088229.
Kim, H., Wigneron, J.-P., Kumar, S., Dong, J., Wagner, W., Cosh, M.H., Bosch, D.D., Collins, C.H., Starks, P.J., Seyfried, M., 2020. Global scale error assessments of soil moisture estimates from microwave-based active and passive satellites and land surface models over forest and mixed irrigated/dryland agriculture regions. Remote Sens. Environ. 251, 112052.
Kingma, D.P. and Ba, J. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Konapala, G., Kao, S.-C., Painter, S.L., Lu, D., 2020. Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. Environ. Res. Lett. 15 (10), 104022.
Kornelsen, K.C., Coulibaly, P., 2014. Root-zone soil moisture estimation using data-driven methods. Water Resour. Res. 50 (4), 2946–2962.
Lakshmi, V., Jackson, T.J., Zehrfuhs, D., 2003. Soil moisture–temperature relationships: results from two field experiments. Hydrol. Process. 17 (15), 3041–3057.
Lehmann, J., Kretschmer, M., Schauberger, B., Wechsung, F., 2020. Potential for early forecast of Moroccan wheat yields based on climatic drivers. Geophys. Res. Lett. 47 (12), e2020GL087516.
Li, W., Migliavacca, M., Forkel, M., Denissen, J.M., Reichstein, M., Yang, H., Duveiller, G., Weber, U., Orth, R., 2022. Widespread increasing vegetation sensitivity to soil moisture. Nat. Commun. 13 (1), 3959.
Liang, M., Yuan, X., 2021. Critical role of soil moisture memory in predicting the 2012 Central United States flash drought. Front. Earth Sci. 9, 615969.
Lin, G., Wang, Z., Ziegler, C., Hu, X.-M., Xue, M., Geerts, B., Chu, Y., 2023. A comparison of convective storm inflow moisture variability between the great plains and the southeastern United States using multiplatform field campaign observations. J. Atmos. Oceanic Tech. 40 (5), 539–556.
Liu, Y., Yuan, S., Zhu, Y., Ren, L., Chen, R., Zhu, X., Xia, R., 2023. The patterns, magnitude, and drivers of unprecedented 2022 mega-drought in the Yangtze River Basin. China. Environmental Research Letters 18 (11), 114006.
Lu, R., Xu, K., Chen, R., Chen, W., Li, F., Lv, C., 2023. Heat waves in summer 2022 and increasing concern regarding heat waves in general. Atmos. Oceanic Sci. Lett. 16 (1), 100290.
Lundberg, S. 2017. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.
Lundberg, S.M. and Lee, S.-I. 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems 30.
Menditto, A., Patriarca, M., Magnusson, B., 2007. Understanding the meaning of accuracy, trueness and precision. Accred. Qual. Assur. 12, 45–47.
Mienye, I.D., Sun, Y., 2022. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Access 10, 99129–99149.
Miralles, D.G., Teuling, A.J., Van Heerwaarden, C.C., Vilà-Guerau de Arellano, J., 2014. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat. Geosci. 7 (5), 345–349.
Mishra, V., Aadhar, S., Mahto, S.S., 2021. Anthropogenic warming and intraseasonal summer monsoon variability amplify the risk of future flash droughts in India. npj Clim. Atmos. Sci. 4 (1), 1.
Mohammadi Foumani, N., Miller, L., Tan, C.W., Webb, G.I., Forestier, G., Salehi, M., 2024. Deep learning for time series classification and extrinsic regression: a current survey. ACM Comput. Surv. 56 (9), 1–45.
Molnar, C., 2020. Interpretable Machine Learning.
Mouginot, J., Rignot, E., Bjørk, A.A., Van den Broeke, M., Millan, R., Morlighem, M., Noël, B., Scheuchl, B., Wood, M., 2019. Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018. Proc. Natl. Acad. Sci. 116 (19), 9239–9244.
Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y., Beroza, G.C., 2020. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nat. Commun. 11 (1), 3952.
Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13 (9), 4349–4383.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10 (3), 282–290.
Naud, C.M., Chen, Y., Rangwala, I., Miller, J.R., 2013. Sensitivity of downward longwave surface radiation to moisture and cloud changes in a high-elevation region. J. Geophys. Res. Atmos. 118 (17), 10,072–010,081.
Nguyen, H., Wheeler, M.C., Hendon, H.H., Lim, E.-P., Otkin, J.A., 2021. The 2019 flash droughts in subtropical eastern Australia and their association with large-scale climate drivers. Weather Clim. Extremes 32, 100321.
Otkin, J.A., Anderson, M.C., Hain, C., Svoboda, M., 2015. Using temporal changes in drought indices to generate probabilistic drought intensification forecasts. J. Hydrometeorol. 16 (1), 88–105.
Otkin, J.A., Svoboda, M., Hunt, E.D., Ford, T.W., Anderson, M.C., Hain, C., Basara, J.B., 2018. Flash droughts: a review and assessment of the challenges imposed by rapid-onset droughts in the United States. Bull. Am. Meteorol. Soc. 99 (5), 911–919.
Otkin, J.A., Woloszyn, M., Wang, H., Svoboda, M., Skumanich, M., Pulwarty, R., Lisonbee, J., Hoell, A., Hobbins, M., Haigh, T., 2022. Getting ahead of flash drought: from early warning to early action. Bull. Am. Meteorol. Soc. 103 (10), E2188–E2202.
Padrón, R.S., Gudmundsson, L., Michel, D., Seneviratne, S.I., 2020. Terrestrial water loss at night: global relevance from observations and climate models. Hydrol. Earth Syst. Sci. 24 (2), 793–807.
Park, S., Im, J., Jang, E., Rhee, J., 2016. Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agric. For. Meteorol. 216, 157–169.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., 2019. Pytorch: an imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems 32.
Pendergrass, A.G., Meehl, G.A., Pulwarty, R., Hobbins, M., Hoell, A., AghaKouchak, A., Bonfils, C.J., Gallant, A.J., Hoerling, M., Hoffmann, D., 2020. Flash droughts present a new challenge for subseasonal-to-seasonal prediction. Nat. Clim. Chang. 10 (3), 191–199.
Pitis, S. 2016 Non-zero initial states for recurrent neural networks.
Plésiat, É., Dunn, R.J., Donat, M.G., Kadow, C., 2024. Artificial intelligence reveals past climate extremes by reconstructing historical records. Nat. Commun. 15 (1), 9191.

Powers, D.M. 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.
Prechelt, L. (2002) Neural Networks: Tricks of the trade, pp. 55-69, Springer.
Prodhan, F.A., Zhang, J., Yao, F., Shi, L., Pangali Sharma, T.P., Zhang, D., Cao, D., Zheng, M., Ahmed, N., Mohana, H.P., 2021. Deep learning for monitoring agricultural drought in South Asia using remote sensing data. Remote Sens. (Basel) 13 (9), 1715.
Qian, Y., Hsu, P.C., Murakami, H., Xiang, B., You, L., 2020. A hybrid dynamical-statistical model for advancing subseasonal tropical cyclone prediction over the western North Pacific. Geophys. Res. Lett. 47 (20), e2020GL090095.
Qing, Y., Wang, S., Ancell, B.C., Yang, Z.-L., 2022. Accelerating flash droughts induced by the joint influence of soil moisture depletion and atmospheric aridity. Nat. Commun. 13 (1), 1139.
Rahmati, M., Amelung, W., Brogi, C., Dari, J., Flammini, A., Bogena, H., Brocca, L., Chen, H., Groh, J., Koster, R.D., 2024. Soil moisture memory: State-of-the-art and the way forward. Rev. Geophys. 62 (2), e2023RG000828.
Raissi, M., Perdikaris, P. and Karniadakis, G.E. 2017. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv preprint arXiv:1711.10561.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, F., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743), 195–204.
Reid, P.C., Hari, R.E., Beaugrand, G., Livingstone, D.M., Marty, C., Straile, D., Barichivich, J., Goberville, E., Adrian, R., Aono, Y., 2016. Global impacts of the 1980s regime shift. Glob. Chang. Biol. 22 (2), 682–703.
Riveros-Iregui, D.A., Emanuel, R.E., Muth, D.J., McGlynn, B.L., Epstein, H.E., Welsch, D. L., Pacific, V.J., Wraith, J.M., 2007. Diurnal hysteresis between soil CO2 and soil temperature is controlled by soil water content. Geophys. Res. Lett. 34 (17).
Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M.D., Muñoz-Marí, J., 2019. Inferring causation from time series in Earth system sciences. Nat. Commun. 10 (1), 2553.
Rushton, K., Eilers, V., Carter, R., 2006. Improved soil moisture balance methodology for recharge estimation. J. Hydrol. 318 (1–4), 379–399.
Samadi-koucheksaraee, A., Ahmadianfar, I., Bozorg-Haddad, O., Asghari-pari, S.A., 2019. Gradient evolution optimization algorithm to optimize reservoir operation systems. Water Resour. Manag. 33 (2), 603–625.
Samadi-Koucheksaraee, A., Chu, X., 2024. Development of a novel modeling framework based on weighted kernel extreme learning machine and ridge regression for streamflow forecasting. Sci. Rep. 14 (1), 30910.
Samadi-Koucheksaraee, A., Shirvani-Hosseini, S., Ahmadianfar, I. and Gharabaghi, B. (2022) Computational intelligence for water and environmental sciences, pp. 3-33, Springer.
Senay, G.B., Budde, M., Brown, J. and Verdin, J. 2008 Mapping flash drought in the US: Southern Great Plains.
Shahid, M., Cong, Z., Zhang, D., 2018. Understanding the impacts of climate change and human activities on streamflow: a case study of the Soan River basin. Pakistan. Theoretical and Applied Climatology 134 (1), 205–219.
Shirvani-Hosseini, S., Samadi-Koucheksaraee, A., Ahmadianfar, I. and Gharabaghi, B. (2022) Computational intelligence for water and environmental sciences, pp. 157-178, Springer.
Small, E.E., Kurc, S.A., 2003. Tight coupling between soil moisture and the surface radiation budget in semiarid environments: Implications for land-atmosphere interactions. Water Resour. Res. 39 (10).
Sugihara, G., May, R., Ye, H., Hsieh, C.-H., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. Science 338 (6106), 496–500.
Sundararajan, M., Taly, A. and Yan, Q. 2017 Axiomatic attribution for deep networks, pp. 3319-3328, PMLR.
Trok, J.T., Barnes, E.A., Davenport, F.V., Diffenbaugh, N.S., 2024. Machine learning–based extreme event attribution. Sci. Adv. 10 (34), eadl3242.
Tufaner, F., Özbeyaz, A., 2020. Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms. Environ. Monit. Assess. 192, 1–14.
Tyagi, S., Zhang, X., Saraswat, D., Sahany, S., Mishra, S.K., Niyogi, D., 2022. Flash drought: Review of concept, prediction and the potential for machine learning, deep learning methods. Earth's Future 10 (11), e2022EF002723.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in Neural Information Processing Systems 30.
Venkatesan, R. and Li, B. (2017) Convolutional neural networks in visual computing: a concise guide, CRC Press.
Wadoux, A.-M.-C., Heuvelink, G.B., De Bruin, S., Brus, D.J., 2021. Spatial cross-validation is not the right way to evaluate map accuracy. Ecol. Model. 457, 109692.
Wang, C., Han, L., Guo, A., Zhao, X., Yan, H., 2020. Impact of meteorological conditions on agricultural production in autumn of 2019. Chin. J. Agrometeorol. 41 (1), 61–64.
Wang, J., Bras, R.L., 2001. Effect of temperature on surface energy balance. Water Resour. Res. 37 (12), 3383–3386.
Wang, Y., Yang, J., Chen, Y., De Maeyer, P., Li, Z., Duan, W., 2018. Detecting the causal effect of soil moisture on precipitation using convergent cross mapping. Sci. Rep. 8 (1), 12171.
Wang, Y., Yuan, X., 2023. High temperature accelerates onset speed of the 2022 unprecedented flash drought over the Yangtze River Basin. Geophys. Res. Lett. 50 (22), e2023GL105375.
Waskom, M.L., 2021. Seaborn: statistical data visualization. Journal of Open Source Software 6 (60), 3021.
Wenke, S. and Fleming, J. 2019. Contextual recurrent neural networks. arXiv preprint arXiv:1902.03455.
Woo, S., Park, J., Lee, J.-Y. and Kweon, I.S. 2018 Cbam: Convolutional block attention module, pp. 3-19.
Xiao, C., Chen, N., Hu, C., Wang, K., Gong, J., Chen, Z., 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. Remote Sens. Environ. 233, 111358.
Xiao, X., Ming, W., Luo, X., Yang, L., Li, M., Yang, P., Ji, X., Li, Y., 2024. Leveraging multisource data for accurate agricultural drought monitoring: a hybrid deep learning model. Agric Water Manag 293, 108692.
Xu, L., Zhang, X., Wu, T., Yu, H., Du, W., Zhang, C., Chen, N., 2024. Global prediction of flash drought using machine learning. Geophys. Res. Lett. 51 (21), e2024GL111134.
Xu, Q.-S., Liang, Y.-Z., 2001. Monte Carlo cross validation. Chemom. Intel. Lab. Syst. 56 (1), 1–11.
Xu, Y., Goodacre, R., 2018. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. Journal of Analysis and Testing 2 (3), 249–262.
Yang, C., Yuan, H., Zhang, F., Xie, M., Wang, Y., Jiang, G.M., 2024. Convective initiation nowcasting in South China using physics-augmented random forest models and geostationary satellites. Earth Space Sci. 11 (7), e2024EA003571.
Yang, S., Li, R., Wu, T., Hu, G., Xiao, Y., Du, Y., Zhu, X., Ni, J., Ma, J., Zhang, Y., 2020. Evaluation of reanalysis soil temperature and soil moisture products in permafrost regions on the Qinghai-Tibetan Plateau. Geoderma 377, 114583.
Yaseen, Z.M., Halder, B., Tan, M.L., Kilinc, H.C., Ahmadianfar, I., Abba, S.I., Heddam, S., Abdul Maulud, K.N., Demir, V., Al-Areeq, A.M., 2025. Climate change impact analysis on seasonal drought and landforms using meteorological and remote-sensing-derived indices. Acta Geophys. 1–33.
Yevjevich, V.M. (1967) An objective approach to definitions and investigations of continental hydrologic droughts, Colorado State University Fort Collins, CO, USA.
Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716.
Yuan, X., Ma, Z., Pan, M., Shi, C., 2015. Microwave remote sensing of short-term droughts during crop growing seasons. Geophys. Res. Lett. 42 (11), 4394–4401.
Yuan, X., Wang, L., Wu, P., Ji, P., Sheffield, J., Zhang, M., 2019. Anthropogenic shift towards higher risk of flash drought over China. Nat. Commun. 10 (1), 4661.
Yuan, X., Wang, Y., Ji, P., Wu, P., Sheffield, J., Otkin, J.A., 2023. A global transition to flash droughts under climate change. Science 380 (6641), 187–191.
Zhang, A., Lipton, Z.C., Li, M. and Smola, A.J. 2021. Dive into deep learning. arXiv preprint arXiv:2106.11342.
Zhao, W.L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., Qiu, G.Y., 2019. Physics-constrained machine learning of evapotranspiration. Geophys. Res. Lett. 46 (24), 14496–14507.
Zheng, Y., Coxon, G., Woods, R., Power, D., Rico-Ramirez, M.A., McJannet, D., Rosolem, R., Li, J., Feng, P., 2024. Evaluation of reanalysis soil moisture products using cosmic ray neutron sensor observations across the globe. Hydrol. Earth Syst. Sci. 28 (9), 1999–2022.
Zhu, Q., Luo, Y., Zhou, D., Xu, Y.-P., Wang, G., Tian, Y., 2021. Drought prediction using in situ and remote sensing products with SVM over the Xiang River Basin, China. Nat. Hazards 105, 2161–2185.