Sustainable Earth Resources Communications

Article

An interpretable eXtreme Gradient Boosting–SHAP framework for spatial estimation of hydrological drought across multiple time scales

Babak Mohammadi[1,*]

[1] Hydrology Research Unit, Swedish Meteorological and Hydrological Institute, Norrköping SE-601 76, Sweden

* Correspondence: babak.mohammadi@smhi.se

Received: May 25, 2026
Revised: June 15, 2026
Accepted: June 18, 2026
Published: June 23, 2026

Cited as:
Mohammadi, B. An interpretable eXtreme Gradient Boosting–SHAP framework for spatial estimation of hydrological drought across multiple time scales. Sustainable Earth Resources Communications, 2026, 2(2): 80-91. https://doi.org/10.46690/serc.2026.02.03

Abstract: Hydrological drought poses significant challenges to water resource management, hydropower production, and ecosystem sustainability, even in traditionally water-abundant regions such as Sweden. Monthly streamflow data spanning from 1961 to 2025 were used to calculate the streamflow drought index (SDI) at multiple time scales (SDI-1, SDI-3, SDI-6, and SDI-12) at a target station (Gävunda krv) in Sweden using eXtreme Gradient Boosting (XGB) models driven by SDI values from seven neighbouring streamflow stations (Grötsjön, Höljes krv, Långhag, Ljusne strömmar krv, Nybro, Rolfsta, and Skallböle krv) as spatial predictors. The XGB models were trained on approximately 85% of the data and tested on the remaining 15%, achieving strong agreement between observed and estimated SDI and consistently low errors in both training and testing phases. Model performance generally improved with increasing accumulation period during the training phase, indicating that longer SDI time scales exhibit stronger spatial coherence among neighbouring stations. Scatter plot and time series analyses confirmed the models' ability to accurately reproduce the temporal dynamics of major drought events, including the extreme 2018 northern European drought, with no evidence of overfitting. SHapley Additive exPlanations (SHAP) analysis was applied to interpret the trained XGB models and, through beeswarm summary plots, to identify which neighbouring stations most strongly influenced drought estimations at Gävunda krv across different time scales and hydrological conditions. The analysis showed that drought at Långhag station was consistently the dominant predictor of drought at Gävunda krv across all four time scales, whereas the second-ranked contributor shifted from Nybro (SDI-1) to Höljes krv (SDI-3 and SDI-6) and then Grötsjön (SDI-12). Rolfsta and Skallböle krv ranked among the least influential stations, most consistently at the longer accumulation periods (SDI-6 and SDI-12). The results demonstrate that the proposed XGB–SHAP framework offers a powerful, accurate, and interpretable tool for spatial hydrological drought estimation in Nordic hydroclimatic settings, with practical implications for drought early warning systems and water resource management in Sweden.

Keywords: Streamflow drought index; water resources; eXtreme Gradient Boosting; SHapley Additive exPlanations; hydrological drought; Sweden

1. Introduction

Drought is widely recognised as one of the most complex and devastating natural hazards, affecting ecosystems, agriculture, water supply, and socio-economic systems across virtually all climatic zones (Wilhite, 2016; Mishra and Singh, 2010; Vicente-Serrano et al., 2020). Unlike sudden-onset disasters such as floods or earthquakes, drought is a creeping phenomenon whose impacts accumulate gradually over time and space,

Yandy Scientific Press

ISSN: 3104-977X

making it particularly difficult to detect, quantify, and manage (Van Loon, 2015; AghaKouchak et al., 2023; Hagenlocher et al., 2019). Droughts are commonly classified into four interconnected types (meteorological, agricultural, hydrological, and socio-economic), each reflecting different stages in the propagation of water deficits through the hydrological cycle (Ma et al., 2022; Gevaert et al., 2018; Zhang et al., 2022). Among these, hydrological drought, characterised by sustained deficits in surface water availability including below-normal streamflow and reduced reservoir storage, is of particular concern for water resource management as it directly reflects the actual supply of usable freshwater within river systems (Van Loon, 2015; Ahmadi and Moradkhani, 2019). Under ongoing climate change, the frequency and severity of drought events are expected to intensify in many regions worldwide, posing escalating risks to water security, food production, and environmental sustainability (Pokhrel et al., 2021; Naumann et al., 2018; Gebrechorkos et al., 2025).

Quantifying and monitoring drought requires the use of standardised drought indices that can objectively characterise drought severity, duration, and spatial extent across different time scales and regions (Liu et al., 2021; Gebrechorkos et al., 2023). Among the most widely used meteorological drought indices are the Standardized Precipitation Index (SPI; McKee et al., 1993), which relies solely on precipitation data, and the Standardized Precipitation Evapotranspiration Index (SPEI; Vicente-Serrano et al., 2010), which additionally incorporates potential evapotranspiration to account for temperature-driven water demand. In this context, Nalbantis (2008) and Nalbantis and Tsakiris (2009) proposed the Streamflow Drought Index (SDI), a simple yet effective hydrological drought index that uses cumulative streamflow volumes computed over various overlapping reference periods. The SDI follows a calculation procedure analogous to the SPI, whereby cumulative streamflow volumes are fitted to a suitable probability distribution and subsequently transformed to a standard normal variable, with negative values indicating drought conditions and positive values representing non-drought states (Nalbantis, 2008; Nalbantis and Tsakiris, 2009). The multi-scalar nature of SDI enables differentiation between short-term hydrological stress (SDI-1 and SDI-3) and long-term sustained water deficits (SDI-6 and SDI-12), providing critical information for both real-time drought monitoring and strategic water resource planning (Tareke and Awoke, 2022; Mliyeh et al., 2025).

Over the past two decades, machine learning (ML) approaches have increasingly complemented and, in many cases, outperformed traditional process-based and statistical models in hydrological prediction, driven by their ability to capture complex nonlinear relationships within high-dimensional datasets without requiring explicit physical parameterisation (Dikshit et al., 2021; Kumar et al., 2023; Xu and Liang, 2021). A wide range of ML algorithms, including artificial neural networks (ANN), support vector machines (SVM), random forests (RF), long short-term memory (LSTM) networks, and gradient boosting methods have been applied to drought forecasting, streamflow prediction, and related water resource problems with increasing success (Mohammadi et al., 2025; Akbarian et al., 2023; Deulkar et al., 2025; Niu and Feng, 2021). Among these, eXtreme Gradient Boosting (XGB) has emerged as a particularly powerful tool in hydrological modelling and water resources studies (Niazkar et al., 2024; Ni et al., 2020; Osman et al., 2021). In the specific context of drought prediction, XGB has been successfully employed to forecast various drought indices at multiple temporal scales, demonstrating high accuracy and robustness across different hydroclimatic regions (Li et al., 2025; Gul et al., 2023; Zhang et al., 2019; Mokhtar et al., 2021; Zhang et al., 2024).

While ML models such as XGB have demonstrated remarkable predictive performance in hydrological applications, their inherent "black-box" nature remains a fundamental barrier to their adoption in operational water resource management and scientific research, where understanding the physical drivers behind predictions is as important as the accuracy of the predictions themselves (Slater et al., 2025). This challenge has given rise to the rapidly growing field of Explainable Artificial Intelligence (XAI), which encompasses a suite of post-hoc interpretation techniques designed to reveal how individual input features contribute to model outputs, thereby bridging the gap between predictive power and physical interpretability (Lundberg and Lee, 2017; Dikshit and Pradhan, 2021; Zounemat-Kermani and Kheimi, 2026). Among the available XAI methods, SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee (2017), has become the most widely adopted framework in hydrological research due to its solid theoretical foundation rooted in cooperative game theory and its ability to provide both local and global feature importance assessments with desirable properties of consistency and local accuracy (Lundberg and Lee, 2017; Lundberg et al., 2020).

Despite this rapid methodological progress, a clear gap remains in the way explainable ML has been applied to drought. Where SHAP has been employed in drought modelling, it has almost invariably been used to rank which variables or which temporal lags govern the estimation, whereas its capacity to rank which locations govern drought at a given gauge, and thereby to expose the spatial architecture of drought coherence within a river network, has remained largely unexploited. This distinction matters because drought propagates through a river system as a spatially organised signal rather than as an isolated point process, and knowledge of which neighbouring catchments carry that signal, and which do not, is directly relevant to drought monitoring, network design, and early warning.

The spatial interpretability of multi-scale hydrological drought estimation in Nordic catchments (i.e., which neighbouring stations drive drought at a target gauge) has received little attention. The novelty of the present study lies in how explainable ML algorithms are designed and interpreted for estimating SDI at a target station solely from SDI values at neighbouring streamflow stations, and applying the XGB–SHAP framework to identify which stations most strongly influence hydrological drought conditions at the target gauge across multiple time scales in Sweden. To address this gap, the present study pursues the following objectives: (i) to compute SDI at four time scales (SDI-1, SDI-3, SDI-6, and SDI-12) for eight streamflow stations in Sweden using monthly streamflow data obtained from the Swedish Meteorological and Hydrological Institute (SMHI); (ii) to develop XGB models that predict SDI values at the target station (Gävunda krv) using SDI values from seven neighbouring stations (Grötsjön, Höljes krv, Långhag, Ljusne strömmar krv, Nybro, Rolfsta, and Skallböle krv) as input features; and (iii) to apply SHAP analysis to interpret the trained XGB models, identifying which stations exert the greatest influence on drought estimations at Gävunda krv across different time scales.

2. Materials and methods

2.1. Study area and data used

The present study focuses on eight streamflow gauging stations distributed across central and west-central Sweden. The selected station names are Grötsjön, Höljes krv, Långhag, Ljusne strömmar krv, Nybro, Rolfsta, Skallböle krv, and Gävunda krv. The monthly streamflow data used in this study were obtained from the Swedish Meteorological and Hydrological Institute (SMHI). Among the eight stations, Gävunda krv was designated as the target station for SDI estimation, while the remaining seven stations (Grötsjön, Höljes krv, Långhag, Ljusne strömmar krv, Nybro, Rolfsta, and Skallböle krv) served as input features for the XGB model. The data were divided into training and testing subsets following an approximately 85%/15% temporal split, with the exact periods varying slightly across SDI time scales, for SDI-1, training covered December 1961 to November 2015 and testing covered December 2015 to May 2025; for SDI-3, training spanned February 1962 to November 2015 and testing from December 2015 to May 2025; for SDI-6, training extended from May 1962 to December 2015 and testing from January 2016 to May 2025; and for SDI-12, training ran from November 1962 to December 2015 and testing from January 2016 to May 2025. The longitude and latitude of the study stations, along with their names and abbreviations, are listed in Table 1. The geographic locations of the eight gauging stations, together with the topography of the study region represented by the digital elevation model (DEM) are shown in Fig. 1.

Table 1. Names, abbreviations and coordinates of the eight streamflow gauging stations used in this study.

No.	Station Name	Abbrev.	Longitude (°E)	Latitude (°N)
1	Gävunda krv	GAV	14.1240	60.7379
2	Grötsjön	GRO	12.4391	61.8203
3	Höljes krv	HOL	12.5438	60.9512
4	Långhag	LAN	15.7132	60.3959
5	Ljusne strömmar krv	LJU	17.0804	61.2123
6	Nybro	NYB	15.5252	61.3623
7	Rolfsta	ROL	16.9374	61.7364
8	Skallböle krv	SKA	16.9619	62.3636

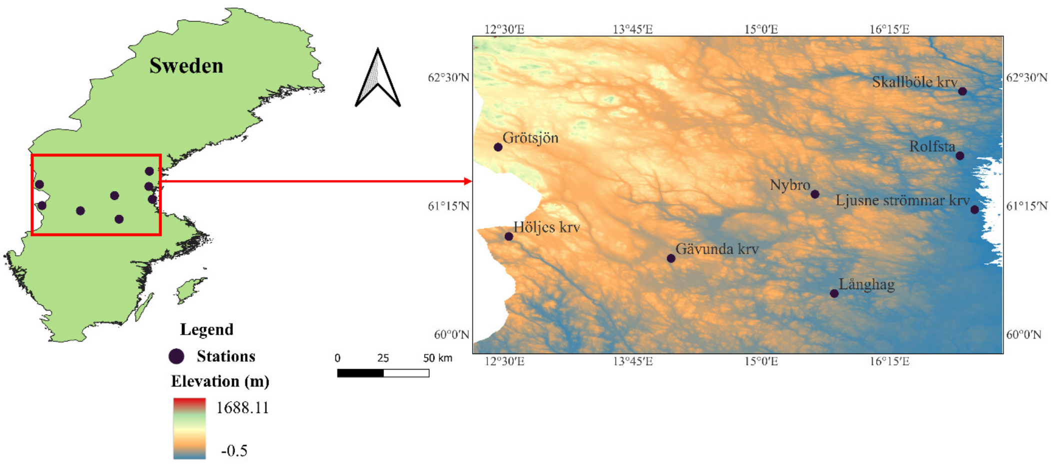

Fig. 1. Locations of the eight streamflow gauging stations over the digital elevation model (DEM) of the study area.

2.2. Methodological framework

The overall methodological framework adopted in this study consists of four sequential stages. In the first stage, monthly streamflow data from all eight stations were collected from SMHI and subjected to quality control and preprocessing. In the second stage, the Streamflow Drought Index (SDI) was computed at four time scales (SDI-1, SDI-3, SDI-6, and SDI-12). In the third stage, for each SDI time scale, an XGB regression model was developed using the SDI values from the seven neighbouring stations as input features and the SDI values at Gävunda krv as the target variable; the model was trained on the training subset and evaluated on the independent testing subset using multiple goodness-of-fit metrics. In the fourth stage, SHAP (SHapley Additive exPlanations) analysis was applied to the developed XGB models to interpret the contribution of each input station to the SDI estimations at Gävunda krv. This integrated framework (combining SDI computation, XGB estimation, and SHAP) enables accurate multi-scale hydrological drought estimation at the target station, and it can also provide a meaningful understanding of which neighbouring stations and under what hydrological conditions they most strongly influence drought at Gävunda krv.

2.3. Streamflow drought index (SDI) calculation

The Streamflow Drought Index (SDI) was computed for each of the eight stations following the procedure of Nalbantis (2008) and Nalbantis and Tsakiris (2009), implemented here in its multi-scalar, probability-based form that is analogous to the Standardized Precipitation Index (McKee et al., 1993). For each time scale k (k = 1, 3, 6, and 12 months), the cumulative streamflow volume ending in month i was obtained as the trailing sum of the monthly streamflow over the preceding k months (Eq. 1), so that each value integrates the antecedent flow conditions of the previous k months.

$$V_{i,k} = \sum_{j=i-k+1}^{i} Q_j \tag{1}$$

where Q_j is the monthly streamflow in month j and $V_{i,k}$ is the cumulative streamflow volume for month i at time scale k.

The cumulative series were then grouped by calendar month (January–December) and, for each month separately, fitted to a set of candidate probability distributions. For each month, the best-fitting distribution was selected with the Kolmogorov–Smirnov (KS) goodness-of-fit test, taking the distribution with the smallest KS statistic among those not rejected at the 0.05 significance level. Using the fitted cumulative distribution function (CDF), each cumulative volume was converted into its non-exceedance probability, which was then mapped to the corresponding quantile of the standard normal distribution through the inverse normal (probit) function, yielding the SDI (Eq. 2):

$$SDI_{i,k} = \Phi^{-1}[F(V_{i,k})] \tag{2}$$

where F is the CDF of the selected distribution and Φ^{-1} denotes the inverse of the standard normal cumulative distribution function. By construction, SDI follows a standard normal distribution with a mean of 0 and a variance of 1. Positive SDI values indicate wetter-than-normal conditions, while negative values indicate hydrological drought, with severity classified according to the thresholds listed in Table 2.

Table 2. Drought-severity classification based on SDI thresholds (Nalbantis (2008) and Nalbantis and Tsakiris (2009)).

Classification	SDI Range
Non-drought	SDI ≥ 0.0
Mild drought	−1.0 < SDI < 0.0
Moderate drought	−1.5 < SDI ≤ −1.0
Severe drought	−2.0 < SDI ≤ −1.5
Extreme drought	SDI ≤ −2.0

2.4. XGB model development and SHAP-based model interpretation

The eXtreme Gradient Boosting (XGB) algorithm (Chen and Guestrin, 2016) was employed to estimate SDI values at the target station (Gävunda krv) using the concurrent SDI values from the seven neighbouring stations as input features. XGB is an advanced ensemble learning method that builds a sequence of decision trees in an additive manner, in which each successive tree is trained to correct the residual errors of the preceding ensemble by minimising a regularised objective function that combines a differentiable loss term (squared error for regression) with penalty terms controlling model complexity (Chen and Guestrin, 2016; Niazkar et al., 2024).

Separate XGB models were developed for each of the four SDI time scales (SDI-1, SDI-3, SDI-6, and SDI-12). Each model received seven input features corresponding to the concurrent SDI values at Grötsjön, Höljes krv, Långhag, Ljusne strömmar krv, Nybro, Rolfsta, and Skallböle krv, and produced a single output representing the estimated SDI at Gävunda krv. The models were implemented in the R programming environment using the xgboost package (Chen and Guestrin, 2016) with a gradient-boosted tree booster, using Root Mean Square Error (RMSE) as the evaluation metric. The hyperparameters were set within the ordinary ranges commonly recommended in the literature for gradient-boosted tree models as follows: a learning rate (eta) of 0.03; a maximum tree depth of 3; a minimum child weight of 5; a minimum loss reduction (gamma) of 0.5; L2 (lambda = 3) and L1 (alpha = 0.2) regularisation terms were included to penalise excessive model complexity; and row and column subsampling ratios of 0.8 each were applied to introduce stochasticity and improve generalisation. Each model was trained for a maximum of 800 boosting rounds with early stopping after 50 consecutive rounds without improvement; also, a fixed random seed (set.seed(123)) was set to ensure reproducibility.

To move beyond the black-box nature of the XGB models and obtain physically meaningful insights into the drivers of SDI estimation, SHapley Additive exPlanations (SHAP) analysis was applied to each trained model following the unified framework introduced by Lundberg and Lee (2017). SHAP is grounded in the concept of Shapley values from cooperative game theory, in which each input feature is treated as a player in a coalition and its contribution to the prediction is quantified by evaluating all possible feature subsets with and without that feature, thereby providing a theoretically fair and locally accurate attribution of each feature's influence on individual predictions (Lundberg and Lee, 2017; Lundberg et al., 2020). The TreeSHAP algorithm (Lundberg et al., 2020), specifically optimised for tree-based ensembles, was used to compute exact Shapley values with polynomial rather than exponential computational complexity, making it particularly efficient for the XGB architecture used here. SHAP values were computed and visualised using the shapviz (Mayer, 2023) and SHAPforxgboost (Liu and Just, 2023) R packages, and the analysis was performed on the entire dataset (training and testing combined) to provide a comprehensive interpretation across all observed hydrological conditions.

2.5. Model performance evaluation

The estimation performance of the XGB models was assessed on both the training and testing periods using four statistical metrics: the Mean Absolute Error (MAE), RMSE, Pearson correlation coefficient (r), and Refined Index of Agreement (dr). These metrics were computed using Equations 3 to 6.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{4}$$

$$r = \frac{\sum_{i=1}^{n}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2 \cdot \sum_{i=1}^{n}(P_i - \overline{P})^2}} \tag{5}$$

$$\mathrm{dr} = \begin{cases} 1 - \dfrac{\sum_{i=1}^{n}|P_i - O_i|}{c.\sum_{i=1}^{n}|O_i - \overline{O}|}, \text{if } \sum_{i=1}^{n}|P_i - O_i| \le c.\sum_{i=1}^{n}|O_i - \overline{O}| \\ \dfrac{c.\sum_{i=1}^{n}|O_i - \overline{O}|}{\sum_{i=1}^{n}|P_i - O_i|} - 1, \text{if } \sum_{i=1}^{n}|P_i - O_i| > c.\sum_{i=1}^{n}|O_i - \overline{O}| \end{cases} \tag{6}$$

where O_i and P_i are observed and estimated data, respectively, $\overline{O}$ and $\overline{P}$ are the means of variables O and P, respectively, n is the sample size, and c is a constant value (set to 2).

3. Results and discussion

3.1. XGB model performance across SDI time scales

Table 3 presents the goodness-of-fit metrics of the XGB models for estimating SDI at Gävunda krv across the four time scales (SDI-1, SDI-3, SDI-6, and SDI-12) during both the training and testing phases. Overall, the results demonstrate that the XGB models achieved strong estimation performance across all SDI time scales, with Pearson correlation coefficients (r) consistently exceeding 0.90 in both training and testing periods, indicating a robust linear agreement between observed and estimated SDI values. During the training phase, model accuracy improved broadly from SDI-1 to SDI-12, with MAE decreasing from 0.326 to 0.211, RMSE decreasing from 0.434 to 0.271, and r increasing from 0.901 to 0.970, suggesting that longer accumulation periods produce smoother and more spatially coherent drought signals that are more readily captured by the inter-station relationships learned by the XGB model. The refined index of agreement (dr) followed a similar improving trend, rising from 0.792 for SDI-1 to 0.865 for SDI-12 during training, further confirming the enhanced model–observation agreement at longer time scales. This pattern is consistent with the hydrological expectation that short-term streamflow variability (SDI-1) is more strongly influenced by localised and stochastic factors such as individual rainfall events and rapid snowmelt pulses, whereas longer-term cumulative streamflow (SDI-6, SDI-12) integrates broader catchment-scale processes that exhibit stronger spatial correlation among neighbouring stations.

The testing phase results, which represent the model's ability to generalise to the independent and unseen period, closely mirrored the training performance with only modest reductions in accuracy, indicating minimal overfitting despite the complexity of the XGB ensemble framework. For SDI-1, the test RMSE (0.427) was virtually identical to the training RMSE (0.434), and the test r (0.915) actually slightly exceeded the training value (0.901), demonstrating excellent generalisation capacity for short-term drought estimation. Across all time scales, the differences between training and testing MAE remained within 0.019 to 0.074, and the dr values decreased by only 0.011 to 0.046 from training to testing, reflecting stable and reliable model behaviour. The SDI-12 model exhibited the highest training performance (r = 0.970, RMSE = 0.271, dr = 0.865) but also showed the largest relative decline during testing (r = 0.946, RMSE = 0.363, dr = 0.819), which can be attributed to the stronger temporal smoothing at the 12-month scale that may amplify the impact of any distributional differences between the training and testing periods, including potential shifts in long-term hydrological regimes or changes in flow regulation patterns between the training (1961–2015) and testing (2016–2025) periods. Nevertheless, even the SDI-12 testing performance remained highly satisfactory, with all metrics indicating strong estimation skill. Collectively, these results confirm that the XGB model effectively leverages the spatial hydrological information embedded in neighbouring station SDI values to accurately reconstruct drought conditions at Gävunda krv across multiple time scales, with particularly robust performance for medium-term drought windows (SDI-3 and SDI-6) that balance between short-term noise and long-term smoothing.

3.2. Scatter plot analysis of observed versus estimated SDI values

Fig. 2 presents scatter plots of observed versus estimated SDI values at Gävunda krv for each of the four time scales (SDI-1, SDI-3, SDI-6, and SDI-12), with training and testing data points distinguished by colour. Across all four panels, the data points cluster tightly along the 1:1 diagonal line, visually confirming the strong agreement between observed and estimated values indicated by the high Pearson correlation coefficients. For SDI-1 (Fig. 2a), the scatter exhibits the widest dispersion around the diagonal, particularly in the moderate-to-severe drought range (SDI values between −1.0 and −2.0), reflecting the inherent challenge of capturing short-term monthly streamflowvariability through spatial information from neighbouring stations alone. As the accumulation period increases from SDI-3 (Fig. 2b) to SDI-6 (Fig. 2c) and SDI-12 (Fig. 2d), the scatter progressively tightens, with data points converging more closely toward the 1:1 line, consistent with the improving correlation coefficients from r = 0.901/0.915 (SDI-1 train/test) to r = 0.970/0.946 (SDI-

12 train/test). Notably, the SDI-6 model (Fig. 2c) displays the most visually symmetric and well-distributed scatter across the full range of drought and wet conditions, spanning approximately from −3.0 to +3.0, suggesting that this intermediate time scale captures the broadest spectrum of hydrological variability while maintaining strong spatial predictability among the stations.

Table 3. Performance evaluation metrics of the XGB models for estimating SDI at Gävunda krv using seven neighbouring stations as input features, assessed during training and testing phases across four time scales (SDI-1, SDI-3, SDI-6, and SDI-12).

Metrics	SDI-1		SDI-3		SDI-6		SDI-12	
	Train	Test	Train	Test	Train	Test	Train	Test
MAE	0.326	0.347	0.246	0.297	0.252	0.271	0.211	0.285
RMSE	0.434	0.427	0.317	0.364	0.316	0.331	0.271	0.363
dr	0.792	0.781	0.845	0.808	0.840	0.823	0.865	0.819
r	0.901	0.915	0.947	0.943	0.957	0.951	0.970	0.946

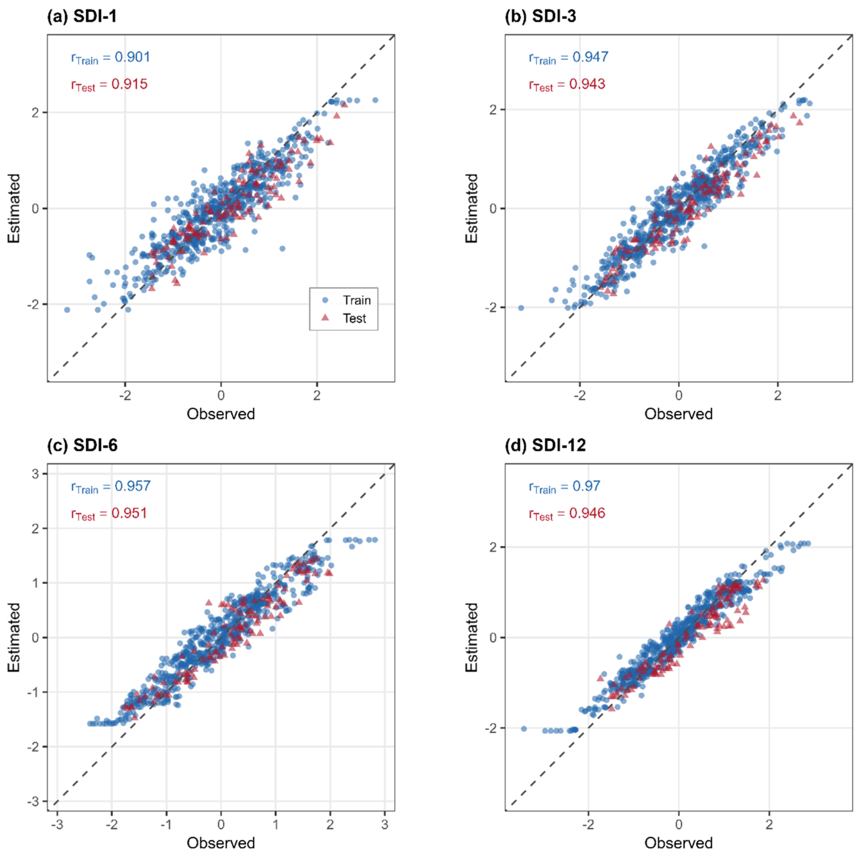

Fig. 2. Scatter plots of observed versus estimated SDI values at Gävunda krv for (**a**) SDI-1, (**b**) SDI-3, (**c**) SDI-6, and (**d**) SDI-12 during training (blue) and testing (red) phases. The Pearson correlation coefficients for both phases are annotated in each panel. The dashed diagonal line represents the 1:1 perfect agreement line.

A particularly encouraging observation across all four panels is that the training (blue) and testing (red) data points overlap substantially and follow the same alignment along the diagonal, providing strong visual evidence that the XGB models did not overfit to the training data and generalise well to the independent testing period. This is especially notable for SDI-1 (Fig. 2a), where the test correlation (r = 0.915) slightly exceeds the training correlation (r = 0.901), and the test points are indistinguishable in their scatter pattern from the training points, indicating that the spatial drought relationships among stations remained temporally stable between the training and testing periods. However, for SDI-12 (Fig. 2d), a subtle increase in scatter is visible among the testing points at extreme negative SDI values (below −2.0), suggesting that the model's ability to reproduce the most severe long-term drought conditions is slightly diminished during the testing period, which may reflect the reduced sample size of extreme events in the shorter testing window or potential non-stationarities in long-term flow regulation patterns between the two periods. Furthermore, across all time scales, the model appears to slightly underestimate the magnitude of the most extreme positive SDI values (wet conditions exceeding +2.0) and extreme negative values (drought conditions below −2.0), a common characteristic of regression-based ML models that tend toward the mean of the training distribution for observations in the tails (Niazkar et al., 2024). Despite this minor limitation at the extremes, the overall scatter plot analysis robustly demonstrates that the XGB models provide reliable and consistent SDI estimations across the full range of hydrological conditions and all four temporal scales examined.

From an operational standpoint, this tendency to underestimate the most extreme values has direct implications for drought early warning, since underestimation of drought severity may lead to the most severe and extreme droughts being classified as less severe than they actually are, potentially delaying or weakening warning signals at precisely the conditions of greatest concern. In practical deployment, this limitation should be taken into account, for example by treating estimated values approaching the severe-drought range as conservative indicators and by complementing the model output with observed streamflow and additional drought indicators when triggering early warning thresholds. Addressing the underestimation more directly, for instance through quantile-based or extreme-value-oriented model formulations that place greater weight on the tails of the distribution, is identified as a valuable direction for future work.

3.3. Temporal evolution of observed and estimated SDI values

Fig. 3 displays the time series of observed and estimated SDI values at Gävunda krv across the entire study period for all four time scales, with the vertical dashed line separating the training and testing phases. Across all panels, the estimated SDI series (shown in red) closely tracks the observed series (shown in blue) throughout both the training and testing periods, successfully reproducing the temporal dynamics of alternating drought and wet episodes spanning more than six decades. The XGB models effectively capture the timing, duration, and relative magnitude of major drought events visible in the historical record, including notable negative SDI excursions during the mid-1970s, mid-1990s, and the severe 2018 drought that is clearly identifiable as a sharp negative spike across all time scales. As expected from the performance metrics, the agreement between observed and estimated series becomes visually tighter from SDI-1 (Fig. 3a) to SDI-12 (Fig. 3d), with the short-term SDI-1 series exhibiting the most rapid and high-frequency oscillations that occasionally deviate from the estimations, while the longer-term SDI-6 and SDI-12 series display smoother temporal patterns that the model reproduces with greater fidelity. Importantly, the transition from the training to testing period shows no visible discontinuity or degradation in the quality of estimations, confirming the temporal robustness and generalisation capacity of the models. The estimated series maintains accurate tracking of both the drought trough during 2018–2019 and the subsequent recovery visible in the testing period across all time scales, further confirming that the spatial relationships among the seven input stations and Gävunda krv remain stable and physically meaningful beyond the training window. A minor limitation is observed at the extreme peaks and troughs, where the estimated values occasionally underestimate the full amplitude of the most severe drought or wettest conditions though the overall temporal phasing and drought onset/termination timing are accurately preserved.

3.4. SHAP-based feature importance and interpretability analysis

Fig. 4 presents the SHAP beeswarm summary plots for the XGB models across the four SDI time scales, where each point represents a single observation, the horizontal position indicates the SHAP value (contribution to the estimation), and the colour denotes the original feature value from low (purple) to high (orange/yellow). The most striking and consistent finding across all four panels is the overwhelming dominance of Långhag as the most influential predictor of SDI at Gävunda krv, with SHAP values typically spanning a range of approximately −1.5 to +1.5 across the four time scales, far exceeding the contribution of any other station. The colour gradient for Långhag displays a clear and physically coherent directional pattern across all time scales: high SDI values at Långhag (orange/yellow points) consistently produce positive SHAP values, pushing the estimated SDI at Gävunda krv toward wetter conditions, while low SDI values at Långhag (purple points) yield strongly negative SHAP values, driving estimations toward drought. This dominant influence likely reflects the shared hydroclimatic characteristics between the Långhag and Gävunda krv catchments, which are likely to belong to the same river system, where both stations appear to be subject to similar precipitation patterns, snowmelt dynamics, and potentially interconnected flow regulation. The second most important station varies across time scales: Nybro ranks second for SDI-1, Höljes krv for SDI-3 and SDI-6, and Grötsjön for SDI-12, indicating that the relative contribution of neighbouring stations to drought estimation shifts with the temporal aggregation window, reflecting different spatial scales of hydrological drought coherence.

The evolution of feature importance rankings across time scales reveals meaningful hydrological insights into the spatial structure of drought propagation in the study region. At the short-term SDI-1 scale, the secondary stations (Nybro, Höljes krv, Ljusne strömmar krv) show moderate but relatively compact SHAP distributions, suggesting that monthly streamflow variability at Gävunda krv is primarily governed by one of the closest stations (Långhag) with limited but detectable contributions from more distant gauges. As the accumulation period increases to SDI-3 and SDI-6, Höljes krv gains prominence as the second-ranked feature, suggesting that seasonal and medium-term drought conditions at Gävunda krv share common large-scale hydroclimatic drivers with this catchment, potentially linked to synoptic-scale precipitation patterns and regional snowpack dynamics that manifest over multi-month periods. At the longest time scale (SDI-12), Grötsjön rises to second position, implying that long-term sustained drought at Gävunda krv is most coherent with catchments that share similar slow-responding hydrological storage characteristics. Conversely, Rolfsta and Skallböle krv rank among the least influential stations, most consistently at the longer accumulation periods (SDI-6 and SDI-12), with their SHAP values tightly clustered near zero, indicating minimal explanatory contribution to drought conditions at Gävunda krv. This finding suggests that these two stations may be hydrologically disconnected from Gävunda krv due to differences in catchment characteristics, flow regulation regimes, or geographic distance.

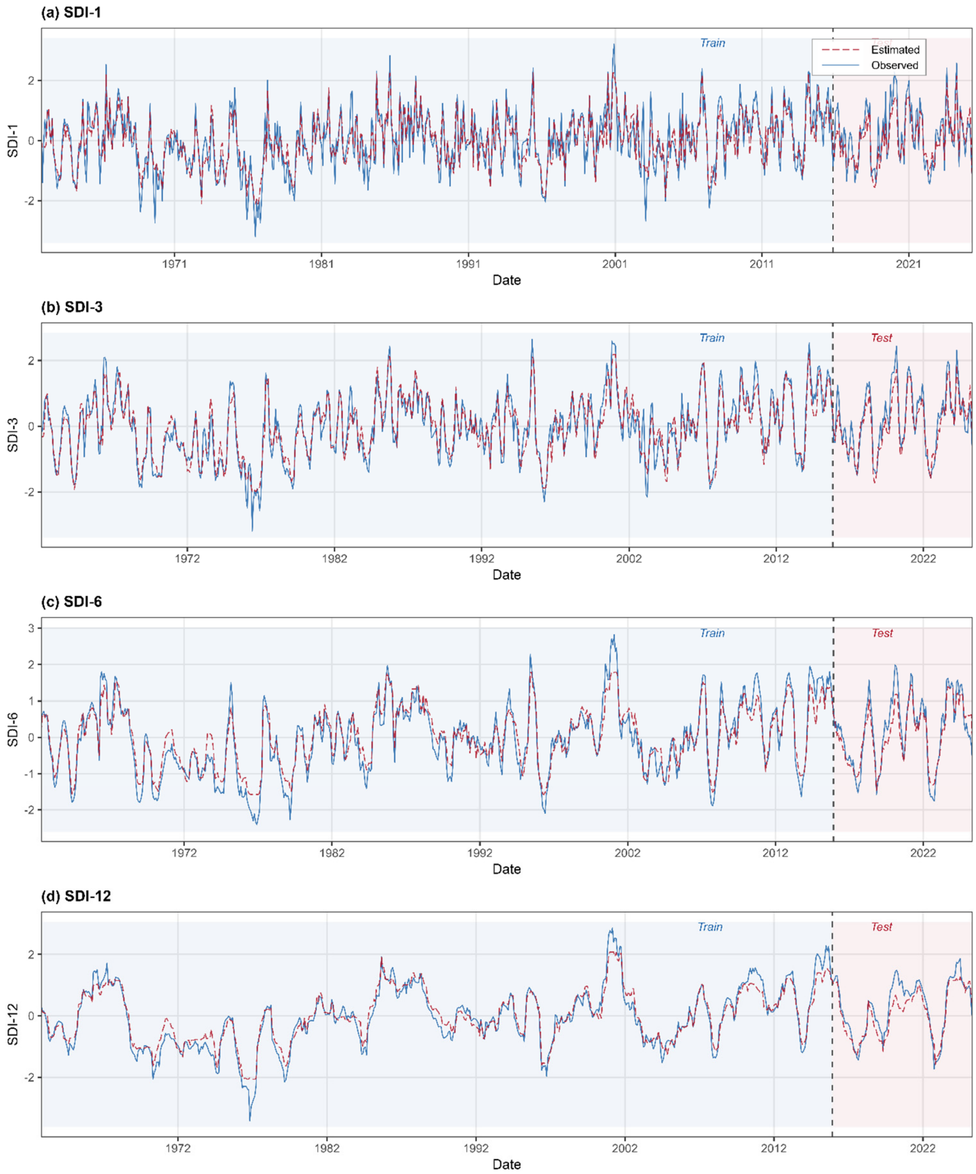

Fig. 3. Time series of observed (blue) and estimated (red) SDI values at Gävunda krv for (**a**) SDI-1, (**b**) SDI-3, (**c**) SDI-6, and (**d**) SDI-12 over the entire study period. The vertical dashed line indicates the boundary between the training and testing phases.

The SHAP rankings obtained here can be interpreted in the light of the spatial configuration of the gauging network. A broad tendency for influence to decline with increasing separation from the target station is apparent, in that stations located relatively close to Gävunda krv generally occupy the upper positions of the SHAP ranking, whereas the more distant gauges exhibit SHAP values clustered near zero across the time scales examined. This pattern is hydrologically coherent, since catchments situated within a limited distance of one another are exposed to broadly similar precipitation fields, comparable snow accumulation and melt timing, and the same synoptic-scale weather systems, and therefore tend to experience largely synchronous streamflow deficits. The relationship is nevertheless not strictly monotonic, and the departures from it are themselves informative. Långhag consistently ranks as the dominant predictor even though it is not the single closest gauge to the target, which indicates that influence is governed not by geographic separation alone but by the degree of shared hydrological behaviour, arising from comparable catchment response times, similar snowmelt-dominated flow regimes, and potentially interconnected regulation within the same river system. Conversely, the consistently negligible contribution of certain gauges is likely to reflect not merely their greater distance but underlying differences in catchment response and flow-regulation regime, which would tend to decouple their drought signal from that of Gävunda krv.

Streamflow gauges located within a limited distance of one another commonly exhibit strong multicollinearity, and correlated input features might affect the distribution of SHAP attributions, since the contribution shared among collinear predictors may be apportioned between them in ways that do not necessarily correspond to physical causation. The overwhelming dominance of Långhag across all four time scales should therefore be interpreted with appropriate caution: part of its high attribution may reflect a strong statistical association with the target rather than a uniquely causal upstream–downstream hydrological linkage between the Långhag and Gävunda krv catchments.

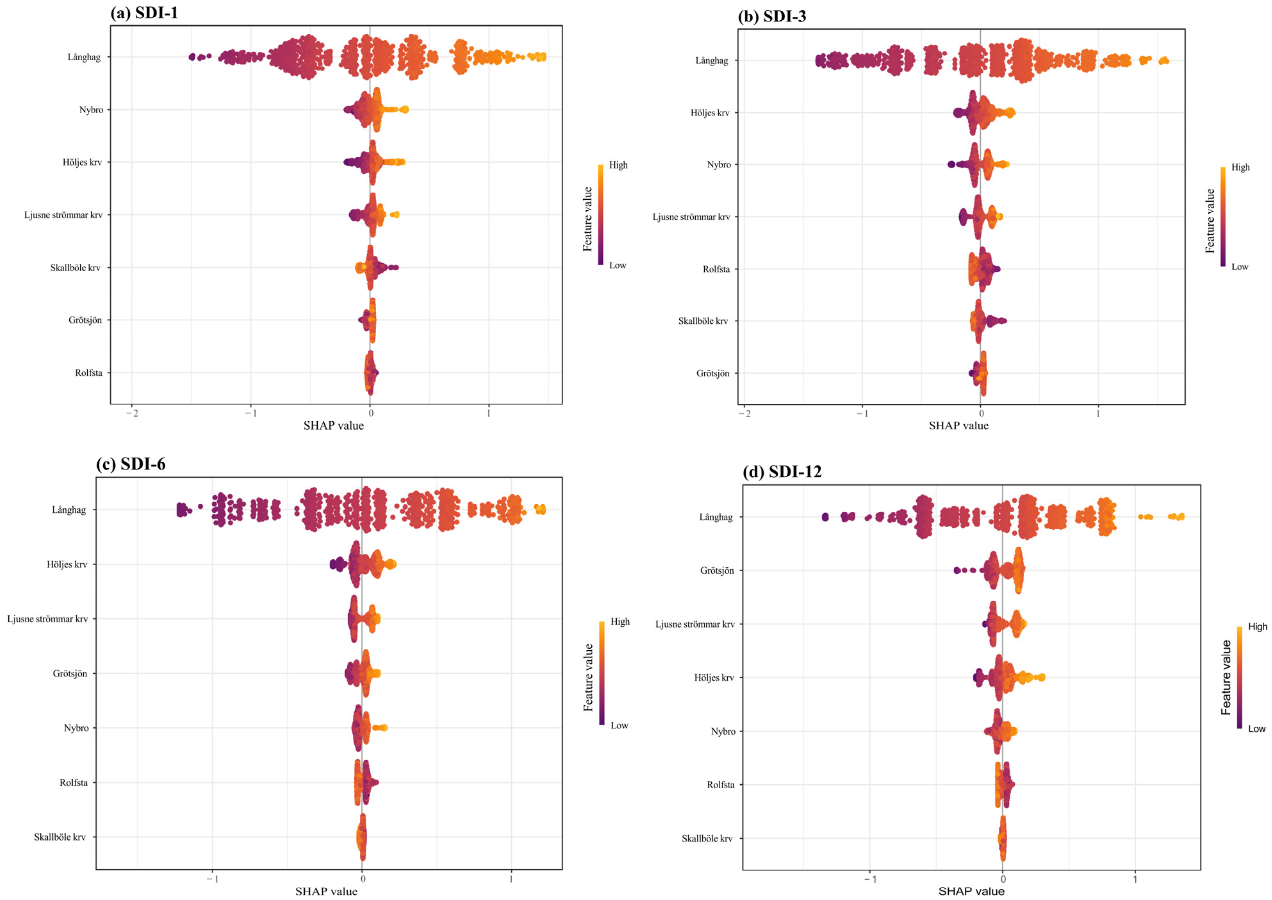

Fig. 4. SHAP beeswarm summary plots showing the contribution of each input station to the XGB estimation of SDI at Gävunda krv for (**a**) SDI-1, (**b**) SDI-3, (**c**) SDI-6, and (**d**) SDI-12. Each point represents a single observation; horizontal position indicates the SHAP value (positive = pushing estimation toward wetter conditions; negative = pushing toward drought); colour represents the original feature value from low (purple) to high (orange/yellow). Features are ranked from top to bottom by mean absolute SHAP value.

The fact that Långhag consistently outranks the one gauge that lies closer to the target suggests that its influence is not explained by proximity-induced collinearity alone and is likely to reflect a genuine hydrological similarity in catchment response and flow regime. A definitive separation of statistical collinearity from physical connectivity would nevertheless require dedicated analysis, which is recommended for consideration in future potential studies.

It is also relevant to distinguish the SHAP-based rankings from a simple correlation analysis among the SDI series. Because SHAP quantifies the marginal contribution of each feature in the presence of all others, rather than its bivariate association with the target, a station that is strongly correlated with Gävunda krv may nonetheless receive a modest SHAP attribution if the information it carries is largely duplicated by another, more informative station. Under such conditions, the boosted trees preferentially exploit the more informative gauge, and the attribution of the redundant one is correspondingly suppressed. The SHAP results therefore identify which stations act as non-redundant carriers of drought information within the network, a property that a correlation matrix alone cannot reveal and that is directly relevant to monitoring-network design. Viewed in this light, the systematic migration of the second-ranked contributor with increasing accumulation period acquires a clear physical meaning. At the monthly scale, the secondary influence is exerted by a comparatively nearby catchment, consistent with the dominance of localised precipitation events and rapid snowmelt pulses in generating short-term flow anomalies. As the accumulation window lengthens to three, six, and twelve months, the secondary influence shifts between catchments, culminating at the twelve-month scale in a markedly more distant gauge (Grötsjön), indicating that sustained multi-month deficits are organised at a larger spatial scale by regional hydroclimatic forcing and are mediated by catchments whose storage characteristics impart a comparably slow hydrological response. The spatial footprint of drought coherence at Gävunda krv therefore appears to widen as the integration period increases, a result that emerges only from the joint application of a multi-scalar drought index and a spatially structured feature space.

4. Conclusions

This study presented an explainable framework for estimating the Streamflow Drought Index (SDI) at multiple time scales (SDI-1, SDI-3, SDI-6, and SDI-12) at Gävunda krv in Sweden using eXtreme Gradient Boosting (XGB) models driven by SDI values from seven neighbouring streamflow stations (Grötsjön, Höljes krv, Långhag, Ljusne strömmar krv, Nybro, Rolfsta, and Skallböle krv), with SHapley Additive exPlanations (SHAP) employed for model interpretation. Monthly streamflow data spanning from 1961 to 2025 were obtained from the Swedish Meteorological and Hydrological Institute (SMHI), and SDI values were computed by fitting probability distributions, with the best fit selected by Kolmogorov–Smirnov testing across several candidate distributions. The XGB models demonstrated strong estimation performance across all four time scales, with Pearson correlation coefficients ranging from 0.901 to 0.970 during training and from 0.915 to 0.951 during testing, while RMSE values ranged from 0.271 to 0.434 during training and from 0.331 to 0.427 during testing. During training, model accuracy improved with increasing accumulation period, confirming that longer SDI time scales capture more spatially coherent drought signals among neighbouring stations. The close agreement between training and testing metrics, further corroborated by scatter plot and time series analyses, demonstrated that the models generalised well to the independent testing period without evidence of overfitting, accurately reproducing the timing, duration, and severity of major historical drought events including the extreme 2018 northern European drought.

The SHAP-based interpretation provided physically meaningful insights into the spatial drivers of hydrological drought estimation at Gävunda krv, revealing which neighbouring stations exerted the greatest influence on drought conditions across different time scales and under varying hydrological states. Specifically, Långhag emerged as the overwhelmingly dominant predictor across all four SDI time scales, with SHAP values extending from approximately −1.5 to +1.5, exhibiting a clear positive directional relationship where high (low) SDI values at Långhag consistently pushed estimations toward wetter (drier) conditions at Gävunda krv. The ranking of secondary stations shifted meaningfully across time scales (from Nybro at SDI-1, to Höljes krv at SDI-3 and SDI-6, and Grötsjön at SDI-12) revealing that the spatial structure of hydrological drought coherence evolves with the temporal aggregation window, with short-term drought driven primarily by the dominant neighbouring station and long-term drought sharing common signals with catchments exhibiting similar slow-responding storage characteristics. Rolfsta and Skallböle krv were identified as consistently low-influence stations, particularly at the longer accumulation periods, suggesting potential hydrological disconnection from the target catchment and providing practical guidance for drought monitoring network optimisation. The application of beeswarm summary plots enabled a transparent understanding of the nonlinear and interactive relationships learned by the XGB models, bridging the gap between estimation accuracy and hydrological interpretability. These findings highlight the value of integrating explainable artificial intelligence methods with machine learning in operational hydrology, as they not only enhance stakeholder trust in model estimations but also generate new knowledge about the spatial propagation and interconnectedness of hydrological drought across river systems in Sweden. The demonstrated ability to estimate drought conditions at a target station using spatially distributed streamflow information from neighbouring gauges has practical implications for drought early warning systems, particularly in situations where real-time data at a specific location may be temporarily unavailable due to station malfunction or where forecasted streamflow at upstream stations could be used to anticipate downstream drought development.

Despite the promising results, several limitations and avenues for future research should be acknowledged. First, the present study focused exclusively on a spatial estimation framework using concurrent SDI values from neighbouring stations, without incorporating temporal lead–lag relationships or meteorological variables (predictors) such as precipitation, temperature, or snow water equivalent, which could further enhance estimation skill and enable true forecasting rather than nowcasting. Second, the influence of hydropower regulation on streamflow at several of the study stations may affect the natural drought signal propagation, and future work should investigate the impact of flow regulation on SDI estimation accuracy by comparing regulated and unregulated catchments. Third, while XGB proved highly effective, future studies could benchmark its performance against alternative machine learning architectures such as Long Short-Term Memory (LSTM) networks, Random Forest, and hybrid deep learning models to identify the optimal approach for multi-scale SDI estimation in Nordic hydroclimatic settings. Fourth, the framework could be extended to a larger network of Swedish stations and applied under climate change projections to assess how spatial drought relationships may evolve under future warming scenarios. Finally, the integration of the proposed XGB–SHAP framework into operational drought monitoring platforms could provide real-time, interpretable, and multi-scale hydrological drought assessments across Sweden, supporting proactive water resource management and climate adaptation strategies in an increasingly drought-vulnerable Scandinavian landscape.

Funding

This research received no external funding.

Data Availability Statement

The streamflow data used in this study were obtained from SMHI's open data portal (https://www.smhi.se/data).

Conflict of interest

Given the role as Editorial Board Member, Babak Mohammadi had no involvement in the peer review of this paper and had no access to information regarding its peer-review process. Full responsibility for the editorial process of this paper was delegated to another editor of the journal.

Use of AI and AI-assisted Technologies

During the preparation of this work, the author used Claude (Opus 4.6, Anthropic) to refine the writing style, improve grammatical accuracy, and enhance the clarity and coherence of expression throughout the manuscript. Following the use of this tool, the author carefully reviewed, verified, and edited all generated content to ensure accuracy, consistency, and alignment with the research findings.

Open Access

References

AghaKouchak, A., Huning, L. S., Sadegh, M., et al. Toward impact-based monitoring of drought and its cascading hazards. Nature Reviews Earth & Environment, 2023, 4(8): 582-595.

Ahmadi, B., Moradkhani, H. Revisiting hydrological drought propagation and recovery considering water quantity and quality. Hydrological Processes, 2019, 33(10): 1492-1505.

Akbarian, M., Saghafian, B., Golian, S. Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran. Journal of Hydrology, 2023, 620: 129480.

Chen, T., Guestrin, C. Xgboost: A scalable tree boosting system. Presented at Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, San Francisco, California, USA, 2016.

Deulkar, A. M., Dixit, P. R., Londhe, S. N., et al. Comparative assessment of artificial neural networks (ANNs), long short term memory network (LSTM) and hydrologic engineering Centre-Hydrologic modelling system (HEC-HMS) for runoff modelling. Water Resources Management, 2025, 39(5): 2049-2068.

Dikshit, A., Pradhan, B. Interpretable and explainable AI (XAI) model for spatial drought prediction. Science of the Total Environment, 2021, 801, 149797.

Dikshit, A., Pradhan, B., Huete, A. An improved SPEI drought forecasting approach using the long short-term memory neural network. Journal of environmental management, 2021, 283, 111979.

Gebrechorkos, S. H., Peng, J., Dyer, E., et al. Global high-resolution drought indices for 1981–2022. Earth System Science Data, 2023, 15: 5449–5466.

Gebrechorkos, S. H., Sheffield, J., Vicente-Serrano, S. M., et al. Warming accelerates global drought severity. Nature, 2025, 642(8068): 628-635.

Gevaert, A. I., Veldkamp, T. I., Ward, P. J. The effect of climate type on timescales of drought propagation in an ensemble of global hydrological models. Hydrology and Earth System Sciences, 2018, 22(9): 4649-4665.

Gul, E., Staiou, E., Safari, M. J. S., et al. Enhancing meteorological drought modeling accuracy using hybrid boost regression models: A case study from the Aegean region, Türkiye. Sustainability, 2023, 15(15): 11568.

Hagenlocher, M., Meza, I., Anderson, C. C., et al. Drought vulnerability and risk assessments: state of the art, persistent gaps, and research agenda. Environmental Research Letters, 2019, 14(8): 083002.

Kumar, V., Kedam, N., Sharma, K. V., et al. Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. Water, 2023, 15(14): 2572.

Li, M., Yao, Y., Feng, Z., et al. Hydrological drought prediction and its influencing features analysis based on a machine learning model. Natural Hazards and Earth System Sciences, 2025, 25(11): 4299-4316.

Liu, C., Yang, C., Yang, Q., et al. Spatiotemporal drought analysis by the standardized precipitation index (SPI) and standardized precipitation evapotranspiration index (SPEI) in Sichuan Province, China. Scientific reports, 2021, 11(1): 1280.

Liu, Y., Just, A. (2023). SHAPforxgboost: SHAP Plots for 'XGBoost'. R package version 0.2.0. Available online: https://cran.r-project.org/package=SHAPforxgboost (accessed on 25 June 2026).

Lundberg, S. M., Erion, G., Chen, H., et al. From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2020, 2(1): 56-67.

Lundberg, S. M., Lee, S. I. A unified approach to interpreting model predictions. Presented at 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017.

Ma, L., Huang, Q., Huang, S., et al. Propagation dynamics and causes of hydrological drought in response to meteorological drought at seasonal timescales. Hydrology Research, 2022, 53(1): 193-205.

Mayer, M. (2023). shapviz: SHAP visualizations. R package version 0.10.3. Available online: https://CRAN.R-project.org/package=shapviz (accessed on 25 June 2026).

McKee, T. B., Doesken, N. J., Kleist, J. The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology, Anaheim, California, USA, 17-22 January 1993.

Mishra, A. K., Singh, V. P. A review of drought concepts. Journal of hydrology, 2010, 391(1-2): 202-216.

Mliyeh, M. M., Aqnouy, M., Laaraj, M., et al. Assessing drought dynamics in a semi-arid basin: a multi-index approach using hydrological and remote-sensing indicators. Environmental Sciences Europe, 2025, 37(1): 1-19.

Mohammadi, B., Abdallah, M., Oucheikh, R., et al. Enhancing streamflow drought prediction: integrating wavelet decomposition with deep learning and quantile regression neural network models. Earth Science Informatics, 2025, 18(2): 232.

Mokhtar, A., Jalali, M., He, H., et al. Estimation of SPEI meteorological drought using machine learning algorithms. IEEE Access, 2021, 9: 65503-65523.

Nalbantis, I. Evaluation of a hydrological drought index. European water, 2008, 23(24): 67-77.

Nalbantis, I., Tsakiris, G. Assessment of hydrological drought revisited. Water resources management, 2009, 23(5): 881-897.

Naumann, G., Alfieri, L., Wyser, K., et al. Global changes in drought conditions under different levels of warming. Geophysical Research Letters, 2018, 45(7): 3285-3296.

Ni, L., Wang, D., Wu, J., et al. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. Journal of Hydrology, 2020, 586: 124901.

Niazkar, M., Menapace, A., Brentan, B., et al. Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023). Environmental Modelling & Software, 2024, 174: 105971.

Niu, W. J., Feng, Z. K. Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. Sustainable Cities and Society, 2021, 64, 102562.

Osman, A. I. A., Ahmed, A. N., Chow, M. F., et al. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain Shams Engineering Journal, 2021, 12(2): 1545-1556.

Pokhrel, Y., Felfelani, F., Satoh, Y., et al. Global terrestrial water storage and drought severity under climate change. Nature Climate Change, 2021, 11(3): 226-233.

Slater, L., Blougouras, G., Deng, L., et al. Challenges and opportunities of ML and explainable AI in large-sample hydrology. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 2025, 383(2302): 20240287.

Tareke, K. A., Awoke, A. G. Hydrological drought analysis using streamflow drought index (SDI) in Ethiopia. Advances in Meteorology, 2022, 2022(1): 7067951.

Van Loon, A. F. Hydrological drought explained. Wiley Interdisciplinary Reviews: Water, 2015, 2(4): 359-392.

Vicente-Serrano, S. M., Beguería, S., López-Moreno, J. I. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. Journal of climate, 2010, 23(7): 1696-1718.

Vicente-Serrano, S. M., Quiring, S. M., Peña-Gallardo, M., et al. A review of environmental droughts: Increased risk under global warming?. Earth-Science Reviews, 2020, 201, 102953.

Wilhite, D. A. Drought as a natural hazard: concepts and definitions. In Droughts, Routledge, pp. 3-18, 2016.

Xu, T., Liang, F. Machine learning for hydrologic sciences: An introductory overview. WIREs Water, 2021, 8(5): e1533.

Zhang, H., Loaiciga, H. A., Sauter, T. A novel fusion-based methodology for drought forecasting. Remote Sensing, 2024, 16(5): 828.

Zhang, R., Chen, Z. Y., Xu, L. J., et al. Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. Science of the Total Environment, 2019, 665: 338-346.

Zhang, X., Hao, Z., Singh, V. P., et al. Drought propagation under global warming: Characteristics, approaches, processes, and controlling factors. Science of the Total Environment, 2022, 838: 156021.

Zounemat-Kermani, M., Kheimi, M. Explainable Artificial Intelligence in Hydrology: A Review. Water Resources Management, 2026, 40(3): 106.