

## Article

# Monthly Runoff Simulation and Driving Mechanism Analysis of the Chushandian Reservoir Irrigation District Based on the Phased Adaptive Attention Informer Model

Yunlong Ran <sup>1,2</sup>, Hongyu Yang <sup>3</sup>, Qingqing Tian <sup>3,\*</sup>, Yu Tian <sup>4</sup> and Lei Guo <sup>2</sup><sup>1</sup> State Key Laboratory of Hydraulic Engineering Intelligent Construction and Operation, Tianjin University, Tianjin 300350, China<sup>2</sup> Henan Water Valley Innovation Technology Research Institute Co., Ltd., Zhengzhou 450002, China<sup>3</sup> School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou 450046, China<sup>4</sup> China Institute of Water Resources and Hydropower Research, Beijing 100038, China

\* Correspondence: tq10078@126.com; Tel.: +86-18236927559

## Abstract

Accurate monthly runoff simulation is the foundation of watershed water resources management and flood–drought disaster prevention. To address the issues that traditional machine learning models struggle to capture long-term dependencies in runoff series and that existing deep learning methods ignore the differences in driving mechanisms across intra-annual hydrological periods, this paper proposes a Period-Adaptive Attention Informer (PA-Informer), in which a learnable period embedding is introduced into the attention mechanism to achieve adaptive representation of the memory length of each period. Based on autocorrelation analysis, K-means clustering and dynamic time warping, the hydrological year is divided into a dry period (November–February), a transition period (March–June and October) and a wet period (July–September). Using monthly precipitation, monthly mean temperature and antecedent runoff (lags 1–3) observed at the Changtaiguan Hydrological Station in the Chushandian Irrigation Area of the upper Huaihe River (1960–2020), a one-month-ahead monthly runoff simulation is performed with the meteorological variables of the target month taken as known; the record is split chronologically into a training set (1960–2002) and a test set (2003–2020) at a ratio of 7:3, and all preprocessing and tuning procedures are restricted to the training set. On the test set, PA-Informer achieves an  $R^2 = 0.957$ ,  $NSE = 0.948$ ,  $RMSE = 0.71 \text{ m}^3/\text{s}$  and  $MAE = 0.52 \text{ m}^3/\text{s}$  (mean over five random seeds), outperforming the best benchmark (Base-Informer) with a statistically significant difference (Wilcoxon signed-rank test,  $p = 0.0023$ ). SHAP analysis reveals that dry-period runoff is driven mainly by temperature and antecedent runoff, transition-period runoff by precipitation and long-lag runoff, and wet-period runoff predominantly by precipitation with temperature acting as a suppressor. PA-Informer effectively integrates the physical laws of hydrological periodicity with a deep learning architecture, providing a new approach for high-accuracy monthly runoff simulation in irrigation areas.



Academic Editor: Vasilis Kanakoudis

Received: 18 September 2026

Revised: 29 September 2026

Accepted: 30 September 2026

Published: 9 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

**Keywords:** monthly runoff simulation; period-adaptive attention; Informer; SHAP analysis; driving mechanism; Changtaiguan Irrigation Area

## 1. Introduction

Accurate simulation and reliable prediction of runoff are fundamental to the scientific scheduling of watershed water resources, reduction in flood and drought risks, and support

for the operation and management of hydraulic engineering. The accuracy of prediction directly affects the feasibility and effectiveness of watershed management decisions [1–3]. Under the combined influence of global climate change and regional human activities, monthly runoff series often exhibit complex characteristics interwoven with gradual trends, periodic fluctuations, and random disturbances. Traditional linear models struggle to fully capture their nonlinear evolution mechanisms, leading to insufficient simulation accuracy and limited stability in practical applications [4,5]. Developing more efficient methods for monthly runoff simulation and continuously improving prediction accuracy are of great significance for optimizing water resource allocation, enhancing the ability to respond to extreme hydrological events, and quantitatively analyzing the impacts of climate change on watershed hydrological processes [6,7].

The strong nonlinearity and non-stationarity of runoff series have driven a gradual shift in hydrological forecasting toward machine learning and deep learning approaches [8,9]. Regarding traditional machine learning methods, Szczepanek [10] compared three machine learning models (XGBoost v1.7.6, LightGBM v3.3.5, and CatBoost) for daily streamflow forecasting in a mountainous catchment. Qiao et al. [11] proposed a meta-heuristic evolutionary deep learning model based on a temporal convolutional network (TCN), an improved Aquila optimizer (IAO), and Random Forest (RF scikit-learn v1.2.2) for rainfall-runoff simulation and multi-step runoff prediction. Cui et al. [12] developed a data-driven model by integrating singular spectrum analysis (SSA) with LightGBM to achieve high-accuracy, real-time urban runoff prediction. Abdoulhalik and Ahmed [13] introduced advanced machine learning models to improve prediction accuracy for runoff in the Syr Darya River basin. Although traditional machine learning methods can capture nonlinear relationships in runoff to a certain extent, they have limited ability to extract long-term dependencies and struggle to adapt to the intra-annual dynamic variation in runoff [14,15]. For deep learning models, Xiang et al. [16] constructed a long short-term memory (LSTM) model based on a multi-output architecture and applied it to the Qingxi River and the upper Wapsipinicon River basins to achieve runoff prediction with lead times of 1–24 h. Kratzert et al. [17] demonstrated that long short-term memory (LSTM) networks can reproduce rainfall-runoff dynamics across a large set of catchments without explicit physical constraints, which established the baseline for deep-learning runoff simulation. Yao et al. [18] developed a hybrid deep learning model (CNN-LSTM combined with GRU-ISSA) to achieve high-accuracy river runoff prediction, improving convergence speed and global optimization ability, and validated it in different basins. Yang et al. [19] proposed a hybrid monthly runoff prediction method that integrates TVF-EMD, PE, LSTM, and PSO; they used particle swarm optimization to optimize LSTM for accurate prediction of decomposed and reconstructed subsequences and quantified the contribution of each subsequence to the prediction result, effectively improving monthly runoff prediction accuracy. Although deep learning models can capture long-term dependencies in time series, they generally ignore the differences in driving mechanisms of runoff across dry, transition, and wet periods. Applying a unified structure to simulate the entire sequence tends to cause underestimation in dry periods and insufficient generalization ability in wet periods [18,19].

Song et al. [20] coupled the synchrosqueezed wavelet transform with an improved long short-term memory network for water quality prediction, further confirming the advantage of decomposition-recurrent architectures in hydrological time-series modeling.

He et al. [21] proposed a seasonal-decomposition-based deep GRU method for daily runoff prediction and showed that explicitly separating the seasonal components improves prediction accuracy; in that study, however, the decomposition is a fixed preprocessing step rather than a structural component of the model.

The Informer model, with its probabilistic self-attention and long-sequence input advantages, exhibits stronger stability and accuracy in time-series forecasting tasks, offering a new technical path for hydrological simulation [22–24]. These three studies are cited as cross-domain evidence that the ProbSparse self-attention and long-sequence-input design of Informer generalizes beyond its original application, covering bearing fault diagnosis [22], microgrid load forecasting [23] and traffic-flow forecasting [24]. Yu et al. [25] proposed an Informer-GRU runoff prediction model based on STL-CEEMDAN secondary decomposition and a gorilla troops optimizer (GTO), which improved model performance by 31.72% and verified that seasonal components and local precipitation station data are the core driving factors for prediction. Wang et al. [26] developed a novel Informer neural network using empirical wavelet transform (EWT) and combined it with the Informer model to accurately predict rainfall-runoff observations and provide early warnings of potential damage. Wang et al. [27] proposed a daily runoff prediction model combining BKA, VMD, Informer, and an error correction (EC) strategy, showing that the  $R^2$  increased by 30.43% compared with the basic Informer model. However, the standard Informer model still adopts a fixed attention mechanism without adaptive adjustment to the seasonal dynamic characteristics of runoff, making it difficult to match the memory lengths and driving factor differences across different hydrological periods [28,29]. Therefore, how to combine hydrological physical laws with deep learning architectures to build a period-adaptive prediction model that can accommodate the seasonal variation in runoff has become the key to improving monthly runoff simulation accuracy.

Recent methodological studies have shown that coupling SHAP-based interpretation with systematic hyperparameter optimization substantially improves both the accuracy and the transparency of machine learning regressions. Tran et al. combined Bayesian optimization with SHAP to predict the bond strength of steel reinforcement in geopolymer concrete [30], and Hoang et al. combined grid search with cross-validation and SHAP to predict the axial load capacity of UHPC-strengthened RC columns [31]. These applications, however, concern structural-engineering materials, in which the input–output relationships are time-invariant; in the present hydrological setting, the driving mechanisms vary systematically within the year, which is precisely why the period information is embedded into the attention mechanism rather than treated as an ordinary input feature.

Based on this, this study takes the Chushandian Irrigation Area in the upper Huaihe River as the research region. According to the autocorrelation characteristics of monthly runoff and dynamic time warping, we carry out hydrological period division and construct a period-adaptive attention Informer (PA-Informer) model that incorporates learnable period embedding into the attention mechanism, enabling differentiated simulation of dry, transition, and wet periods. By comparing the proposed model with various machine learning and deep learning models, we validate its effectiveness and advantages. Furthermore, using the SHAP interpretability method, we identify the driving factors of runoff in different periods and reveal the differences in contributions and interaction mechanisms of precipitation, temperature, and antecedent runoff. The results can provide a methodological reference for high-accuracy monthly runoff simulation in irrigation areas and scientific support for regional water resources management and flood–drought prevention.

## 2. Methods

### 2.1. Informer Model

Informer is a modified Transformer model designed for long-sequence time-series forecasting [32,33]. To address the issues of quadratic complexity, high memory usage, and cumulative decoding errors that exist in the standard Transformer when encoding long sequences [34,35], Informer introduces three core improvements: ProbSparse self-attention,

self-attention distilling operation, and a generative decoder. Its structure is shown in Figure 1.

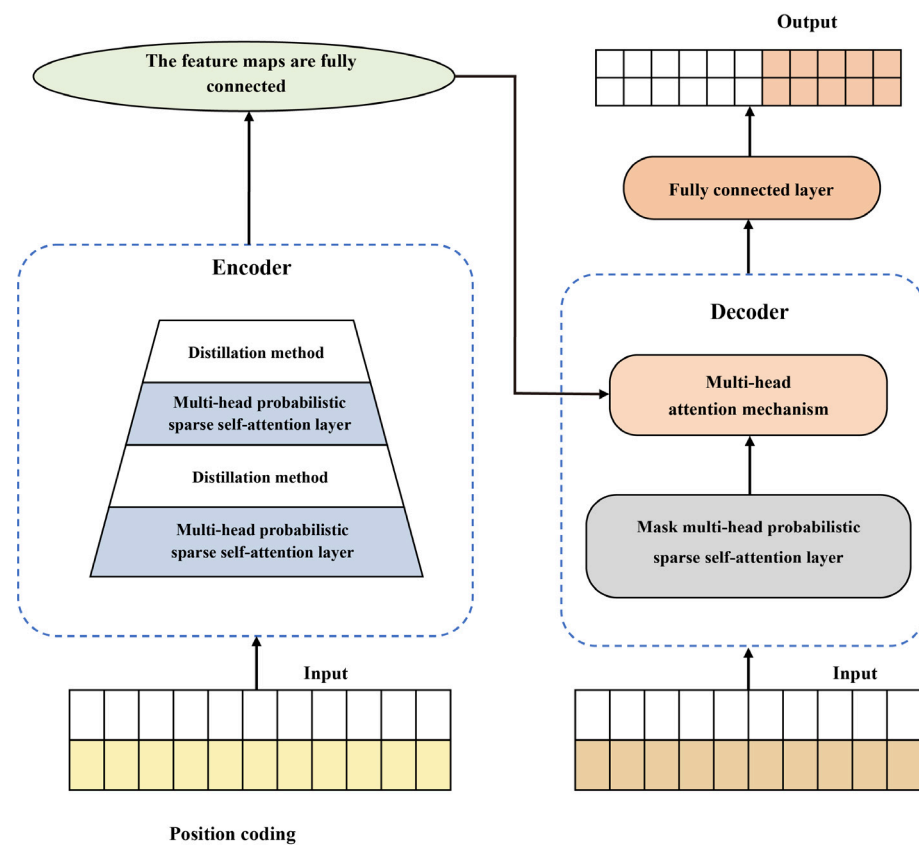


**Figure 1.** Structural diagram of the Informer model.

ProbSparse self-attention reduces the computational complexity of attention from  $O(L^2)$  to  $O(L \log L)$  by selecting only a few “active” key–value pairs that contribute most to the query vector, where  $L$  is the sequence length [36]. Its core idea is to use the Kullback Leibler (KL) divergence to measure the difference between the probability distribution of a query and a uniform distribution. For the  $i$ -th query,  $q_i$ , the sparsity measure is defined as follows:

$$\overline{M}(q_i, K) = \max_j \left\{ \frac{q_i k_j^T}{\sqrt{d}} \right\} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} \quad (1)$$

In the formula,  $\overline{M}(q_i, K)$  represents the maximum arithmetic average sparse average self-attention value at time  $q_i$ . This formula essentially quantifies the distribution differences of query keys to provide a quantitative basis for subsequent sparse screening. The larger this value, the more active the query, thereby significantly improving the computational efficiency without significantly sacrificing accuracy.

The self-attention distillation operation is used to reduce the feature dimension of the encoder’s output, and high-level temporal features at each layer are extracted layer by layer through convolution and pooling layers [37,38]. The output of the  $j + 1$ th layer can be expressed as follows:

$$X_{j+1} = \text{MAXPool}(\text{ELU}(\text{Conv1d}([X_j]_{AB}))) \quad (2)$$

In the formula,  $[X_j]_{AB}$  represents the multi-head attention mechanism, Conv1d is a one-dimensional convolution with a kernel width of 3, ELU is the activation function, and

MaxPool performs downsampling with a stride of 2. This operation enhances the model's ability to capture the dominant dependencies in long sequences.

The generative decoder adopts the method of generating the target sequence in one go, avoiding the error accumulation of the traditional word-by-word decoding [39]. The input of the decoder is the concatenation of the start token and the zero placeholder:

$$X_{de} = \text{Concat}(X_{token}, X_0) \quad (3)$$

It also prevents future information leakage through masked sparse self-attention, finally outputting the prediction results.

The above mechanisms enable Informer to achieve both high accuracy and low computational cost in long-term time-series forecasting tasks, making it a powerful tool for long-sequence prediction scenarios such as hydrometeorology. However, the standard Informer does not consider the intra-seasonal non-stationary characteristics of hydrological series and cannot adaptively adjust the attention distribution according to the different driving mechanisms of hydrological periods, which limits its performance in monthly runoff simulation.

## 2.2. Period-Adaptive Attention Informer (PA-Informer)

PA-Informer incorporates the information of hydrological stages into the attention calculation of the standard Informer. It assigns a label,  $p \in \{0, 1, \dots, x\}$ , to each stage based on the results of hydrological stages and obtains the stage, embedding vector  $e_p \in R^{d_{model}}$  through a learnable embedding layer.

In probabilistic sparse attention, the epoch embeddings are simultaneously added to both the query and the key, allowing the attention weights to vary with the epochs:

$$Q = (XW_Q + e_p), K = (XW_K + e_p), V = XW_V \quad (4)$$

In the formula,  $Q$  represents the query matrix,  $K$  represents the key matrix,  $V$  represents the value matrix, and  $W_Q, W_K, W_V \in R^{d_{model} \times d_k}$  is the learnable weight matrix. Then, the sparsity measure,  $M$ , is calculated and active queries are selected. Finally, the attention output is

$$PA - \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK}{\sqrt{d_k}}\right)V \quad (5)$$

In the formula,  $\text{Softmax}(x)$  represents the activation function, which is used for normalization processing, and  $d$  represents the input dimension.

Through this design, the model can dynamically adjust the correlation between queries and keys according to different hydrological periods, guiding the attention to focus on time steps that match the historical dependency length of each period. The remaining structures of PA-Informer (distilling operation and generative decoder) are the same as those of the standard Informer, and the overall computational complexity remains almost unchanged; however, the adaptive simulation capability for the seasonal non-stationarity of the runoff is markedly enhanced.

## 2.3. Autocorrelation-Based Hydrological Period Division

(1) Autocorrelation Function: The Autocorrelation Function (ACF) is commonly used to test the autocorrelation or periodic characteristics of data. By using the ACF, the autocorrelation coefficient,  $\rho_t$ , between the monthly runoff,  $Q_t$ , and the runoff of the previous month,  $Q_{t-1}$ , is calculated to represent the inter-month dependency. Combined with the monthly runoff,  $Q_t$ , and the autocorrelation coefficient,  $\rho_t$ , a multi-dimensional feature vector is constructed.  $F_t = (Q_t, \rho_t)$  represents the runoff magnitude and its temporal dependency.

It should be noted that autocorrelation estimates are themselves subject to sampling uncertainty. Selective estimators have been developed to reduce the bias of the autocorrelation function for irregularly sampled series [40], and the statistical behavior of autocorrelation estimates under low signal-to-noise conditions has been characterized in radar applications [41]. In the present case, the monthly record is regularly sampled and of moderate length, so the classical estimator with the  $\pm 1.96/\sqrt{N}$  significance bound is adopted, and its period-specific values are reported in Section 4.1.

$$\rho_t = \frac{\sum_{i=1}^{N-1} (Q_{t,i} - \bar{Q})(Q_{t+1,i} - \bar{Q}_{t+1})}{\sqrt{\sum_{i=1}^{N-1} (Q_{t,i} - \bar{Q})^2 \sum_{i=1}^{N-1} (Q_{t+1,i} - \bar{Q}_{t+1})^2}} \quad (6)$$

(2) Elbow Rule: The elbow rule is used to determine the optimal number of clusters in cluster analysis. When the data set is divided into  $K$  clusters,  $C_k$  represents the  $k$ th cluster, the center of each cluster is  $\{\mu_1, \mu_2, \dots, \mu_K\}$ , and SSE is the sum of the squared distances of all data points to the center of their respective clusters. Find the  $K$  that causes a significant slowdown in the rate of the SSE decrease, that is, the elbow point. This point represents the optimal number of divisions.

$$SSE(K) = \sum_{k=1}^K \sum_{x \in C_k} \|x - \mu_k\|^2 \quad (7)$$

(3) Time-series constrained K-means clustering and DTW algorithm segmentation: To avoid disrupting the continuity of the hydrological cycle, the improved K-means algorithm is used for clustering. Based on the optimal number of segments calculated in the previous stage, the Euclidean distance is replaced by the DTW distance. Finally, the clustering results are iteratively optimized through the DTW centroid averaging algorithm. The distance calculation for the two feature sequences is as follows:

The objects to be clustered are the 12 calendar months rather than individual years, so that each calendar month is assigned to exactly one hydrological period and the intra-annual cycle is preserved as a whole. Each month is described by the feature vector  $X_m = [Q_{mean\_m}, \rho_{m\_lag\_1}]$ , where  $Q_{mean\_m}$  is the multi-year mean runoff of month  $m$  and  $\rho_{m\_lag\_1}$  is its lag\_1 autocorrelation coefficient; both components are z-score standardized before clustering. Because the objects are annual-cycle shapes rather than point values, the Euclidean distance is replaced by the DTW distance, and the assignment is iteratively refined by DTW barycentre averaging (DBA). The clustering, therefore, groups months by the similarity of their intra-annual runoff dynamics and does not impose a calendar-continuity constraint: Months that are not adjacent in the calendar may belong to the same period if their runoff dynamics are similar. This is the case for October, which is separated from March–June by the wet period but whose DTW distance to March–June is smaller than to the winter months.

$$DTW(F_i, F_j) = \min_W \sum_{(k,l) \in W} \|F_{i,k} - F_{j,l}\|_2 \quad (8)$$

In the formula,  $F_i$  and  $F_j$  represent two feature sequences;  $F_{i,k}$  and  $F_{j,l}$  are the feature vectors at the  $k$ -th and  $l$ -th time steps; and  $W$  is a regularized path composed of a series of point pairs  $(k, l)$ .

(4) Partial Autocorrelation Function: The Partial Autocorrelation Function (PACF) is used to reflect the pure correlation between two time points after excluding the influence of intermediate terms. By conducting stepwise regression to eliminate the interference of intermediate variables, it can effectively identify the independent contribution of the

lag terms in the Autoregressive (AR) model. The optimal lag period for each period is calculated as follows: determine the partial autocorrelation coefficient,  $\phi_k$ , and determine the threshold,  $\tau$  ( $\alpha = 0.05$ ).

$$\phi_k = \frac{\text{Cov}(y_t, y_{t-k} | y_{t-1}, \dots, y_{t-k+1})}{\sqrt{\text{Var}(y_t | y_{t-1}, \dots, y_{t-k+1}) \cdot \text{Var}(y_{t-k} | y_{t-1}, \dots, y_{t-k+1})}} \quad (9)$$

$$\tau = \frac{\pm Z_{1-\alpha/2}}{\sqrt{N}} \quad (10)$$

In the formula,  $\phi_k$  represents the autocorrelation of the sequence  $y_t$ , with the lagged  $k$ -period values  $y_{t-k}$  under the condition of intermediate lagging  $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ ;  $\hat{y}_{t-k}^{(k-1)}$  is the linear least squares estimation value of the previous  $k-1$  lagged terms  $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$  on  $y_t$ ;  $\pm Z_{1-\alpha/2}$  is the standard normal distribution quantile; and  $N$  is the sample size.

#### 2.4. Evaluation Indicators

To objectively evaluate the prediction performance of each model in the monthly runoff simulation, this study selects four commonly used evaluation metrics: the Nash-Sutcliffe Efficiency (NSE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination ( $R^2$ ). The calculation formulas for each metric are as follows:

All input variables are standardized by the z-score using the mean and standard deviation of the training set, and the model outputs are back-transformed to the original scale before evaluation; the RMSE and MAE are, therefore, reported in  $\text{m}^3/\text{s}$ , the same unit as the observed runoff, and are directly comparable across models and periods.

Suppose the test set contains  $N$  samples.  $y_i$  represents the measured runoff value of the  $i$ -th sample,  $\hat{y}_i$  represents the model's predicted value, and  $\bar{y}$  represents the mean of the measured values.

##### (1) Nash efficiency coefficient (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (11)$$

The range of the NSE is  $(-\infty, 1]$ . The closer it is to 1, the higher the prediction accuracy of the model, and the closer the predicted value is to the actual value. When the NSE is 0, this indicates that the model's prediction effect is equivalent to directly using the actual mean value. If the NSE is negative, this means that the model's prediction effect is worse than the mean prediction.

##### (2) Root Mean Square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (12)$$

The RMSE is sensitive to large errors and can reflect the overall deviation degree of the prediction results. Its unit is the same as the runoff value ( $\text{m}^3/\text{s}$ ), and the smaller the value, the higher the model's accuracy.

##### (3) Mean Absolute Error (MAE)

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

The MAE measures the average absolute value of the prediction error. It is insensitive to outliers and can more robustly reflect the average prediction deviation of the model. The unit is the same as the runoff value ( $\text{m}^3/\text{s}$ ), and the smaller the value, the better.

(4) Coefficient of determination ( $R^2$ )

The value range of  $R^2$  is [0,1]. The closer it is to 1, the stronger the model's ability to explain the variation in the measured runoff and the better the fitting effect.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (14)$$

This paper takes the NSE and  $R^2$  as the main evaluation indicators for the model's fitting accuracy and the RMSE and MAE as the auxiliary evaluation indicators for error precision. All models adopt the same training set and test set division method and calculate the above indicators based on the test set for comparative analysis.

### 3. Study Area and Data

#### 3.1. Overview of the Research Area

The Chushandian Reservoir Irrigation Area is located in the southern part of Henan Province, in the upper reaches of the Huaihe River. It is a supporting project of the Chushandian Reservoir, with a project scale of large type II and a designed irrigation area of 527,000 mu. The irrigation area extends from the downstream of the Chushandian Reservoir in the west to Zhengyang County of Zhumadian in the east, reaches the Shihe District of Xinyang in the south, and borders Queshan County of Zhumadian in the north, with geographical coordinates ranging from  $113^\circ 58' \text{ E}$  to  $114^\circ 35' \text{ E}$  and  $32^\circ 15' \text{ N}$  to  $32^\circ 48' \text{ N}$ , as shown in Figure 2. This region is characterized by abundant land resources and favorable light and temperature conditions, and it is suitable for crop growth, making it an important grain-producing area in Henan Province. The Changtaiguan Hydrological Station is located 14 km downstream of the dam site. This station has long-term observations of hydrological elements, such as water level, discharge, and precipitation, providing a long, continuous, and reliable data series, which serves as an important basis for water resources scheduling and irrigation water management in the irrigation area.

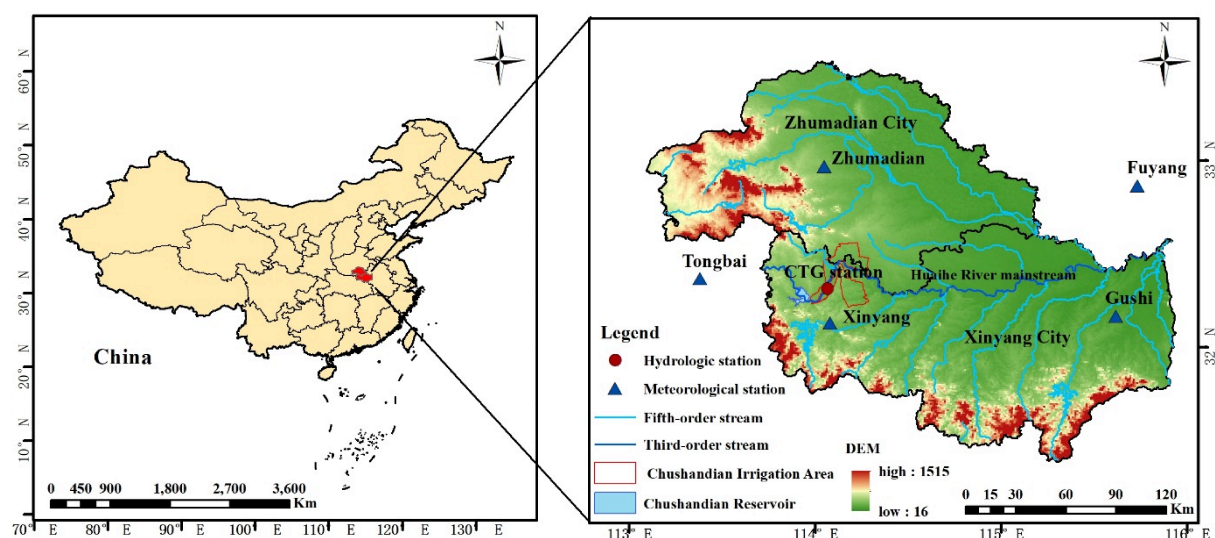


Figure 2. Overview map of the study area.

### 3.2. Data Source and Processing

To simulate and predict the monthly runoff in the Chushandian Irrigation Area, this study selects temperature, monthly precipitation, and antecedent monthly runoff as input factors based on the hydrological characteristics and influencing factors in the study area and conducts monthly runoff simulation and prediction using the period-adaptive attention Informer (PA-Informer). The research data consist of daily runoff measured at the Changtaiguan Hydrological Station downstream of the Chushandian Reservoir on the main stream of the Huaihe River from 1960 to 2020, as well as daily precipitation records from five national meteorological stations (Tongbai, Zhumadian, Xinyang, Fuyang, and Gushi) surrounding the irrigation area over the same period. This study uses the IQR method to handle outliers and applies linear interpolation for data filling. Finally, monthly areal precipitation is obtained by arithmetic averaging of the five station records, which is adequate here because the stations are distributed fairly evenly around the irrigation area and the inter-station differences in monthly precipitation totals are small; monthly mean temperature is taken directly from the Xinyang station, which is located inside the irrigation area, because air temperature exhibits a much smaller spatial gradient than precipitation over this low-relief area. Runoff is observed solely at the Changtaiguan station and, therefore, requires no spatial weighting. The aggregated monthly series are shown in Figure 3. For monthly runoff simulation and prediction, the hydrological data from 1960 to 2020 are divided into a training set and a test set at a ratio of 7:3 in chronological order. The training set is used to train various models and determine the optimal parameters for each model. After tuning each model to its best state, simulation and prediction are performed, and the predicted monthly runoff results are finally compared with the measured data of the test set.

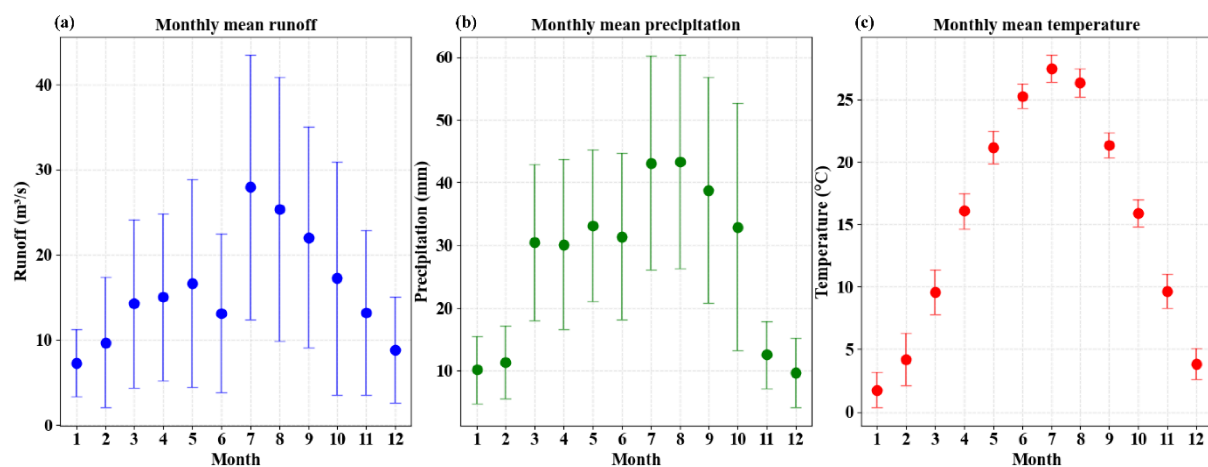


**Figure 3.** Distribution characteristics of hydrological data at Changtai Guan Hydrological Station from 1960 to 2020.

The overall missing-data rate of the aggregated monthly series is 2.1%, and missing values were filled by linear interpolation. Outliers were screened by the IQR rule with a  $1.5 \times \text{IQR}$  threshold applied to the monthly series (732 monthly records in total); 13 observations (1.8% of the sample) were flagged. The IQR rule is applied only to remove evident instrumental and transcription errors—that is, values that exceed 4 times the neighboring-month values and are not accompanied by any corresponding precipitation anomaly—rather than to truncate genuine flood peaks. To confirm this, the analysis was repeated without IQR filtering: the test-set NSE decreased by 0.011 and the RMSE increased by 2.3%, and the ranking of all models was unchanged.

It should be noted that the task addressed in this study is a one-step-ahead monthly runoff simulation rather than operational forecasting: the predictors of the target month,  $t$ , are the monthly precipitation,  $P_t$ , the monthly mean temperature,  $T_t$ , and the antecedent runoff of the preceding three months (lag\_1–lag\_3), and the predictand is the monthly runoff,  $Q_t$ . The prediction horizon is, therefore, one month, and  $P_t$  and  $T_t$  are assumed to be known (i.e., observed) values.

## 4. Results and Analysis

### 4.1. Monthly Runoff Autocorrelation Dynamic Characteristics and Lag Selection

#### 4.1.1. Monthly Runoff Autocorrelation Analysis

Monthly runoff generally exhibits non-stationary and periodic characteristics, and its lag dependence often varies with seasons. To investigate this seasonal differentiation and clarify the differences in runoff memory lengths across different periods, this study begins with lag\_1 autocorrelation to analyze the overall impact of previous monthly runoff on current monthly runoff and its intra-seasonal variability. Scatter plots and monthly autocorrelation coefficient bar charts are used to reveal the differences in runoff memory characteristics across months from both overall distribution and quantitative intensity perspectives, laying the foundation for subsequent hydrological period division, as shown in Figure 4. From the scatter plot in Figure 4a, most points lie near the 1:1 reference line, indicating a significant lag\_1 autocorrelation in the monthly runoff series. When colored by month, points in winter months are concentrated in the low runoff region ( $<20 \text{ m}^3/\text{s}$ ) and closely distributed around the 1:1 line, indicating extremely strong runoff persistence in the low-temperature season. In contrast, the distribution range of points in summer months expands significantly, and the degree of deviation from the 1:1 line increases, reflecting that runoff in the high-temperature and rainy season is influenced by precipitation pulses, and the controlling effect of antecedent runoff of the preceding three months (lag\_1–lag\_3) weakens.

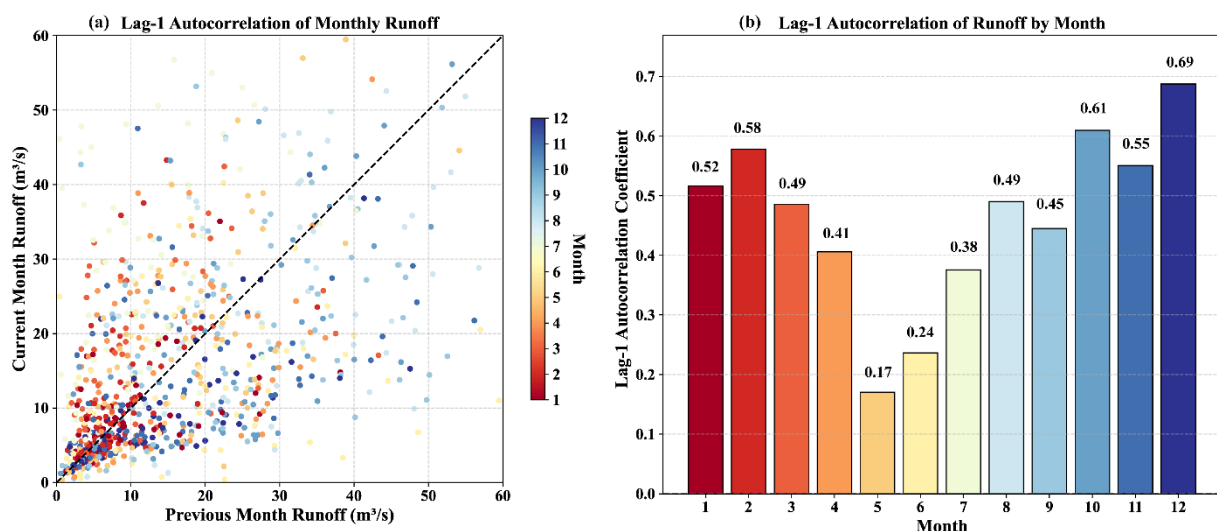


**Figure 4.** Autocorrelation scatter plot and autocorrelation coefficient. (Black dashed line: 1:1 reference line).

The bar chart in Figure 4b further quantifies the differences in autocorrelation intensity across months. The correlation coefficients from October to December and January to February are generally high, with December reaching the highest value of 0.69. The correlation coefficients from March to June are markedly lower, ranging from 0.1 to 0.5. From July to September, the coefficients rebound to between 0.5 and 0.6, though still lower than those in winter months. This pattern of high values in winter, low values in spring, and

a rebound in summer and autumn indicates that in low-temperature and low-precipitation seasons, runoff mainly depends on baseflow and antecedent storage, showing strong memory. In spring, increasing precipitation weakens the autocorrelation. In summer and autumn, although precipitation is abundant, the runoff response to precipitation still carries a certain lag, resulting in autocorrelation stronger than in spring but weaker than in winter.

#### 4.1.2. Hydrological Period Division Based on K-Means and DTW

Based on the clear seasonal differences in monthly runoff autocorrelation, an objective division of the hydrological year is further performed. First, the K-means clustering algorithm is applied to cluster the multi-year average monthly runoff series, and the elbow method is used to determine the optimal number of clusters, as shown in Figure 5. As the number of clusters,  $K$ , increases from one to six, the within-cluster sum of squares (WCSS) gradually decreases. When  $K = 3$ , the decreasing rate slows down significantly, showing an elbow point, indicating that dividing the 12 months into three periods is most reasonable.

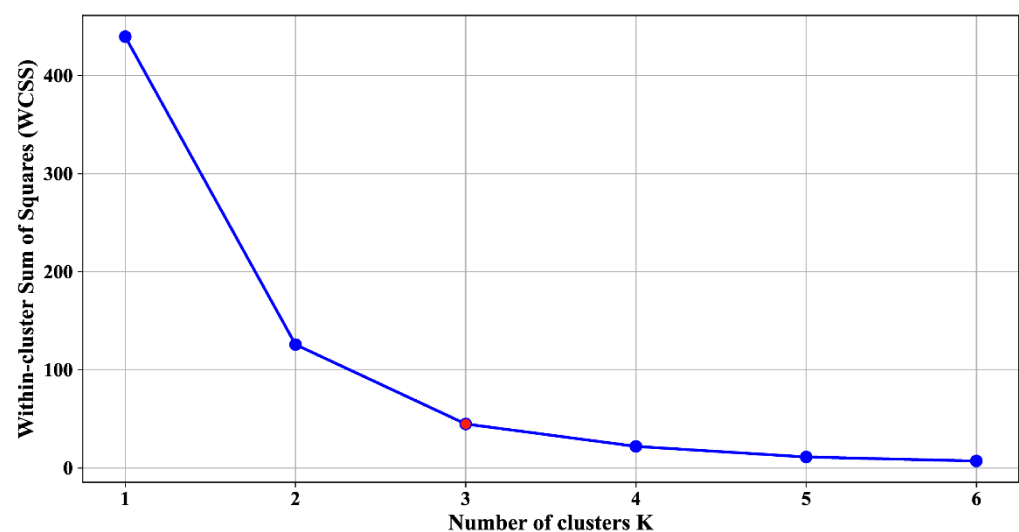


Figure 5. Elbow rule clustering.

The period division was re-derived using only the training years (1960–2002): the 12-month climatological mean runoff was recalculated on the training subset, and the K-means/DTW clustering—including the elbow point at  $K = 3$ —reproduced exactly the same three-period partition (November–February, March–June and October, July–September). The period labels are, therefore, an intra-annual seasonal classification derived from climatological means and contain no inter-annual information from the test period. In addition, the IQR bounds used for outlier screening, the normalization parameters (mean and standard deviation), the PACF-based lag orders and the grid-search hyperparameter selection were all computed on the training set only and then applied unchanged to the test set.

To verify the reliability of the division result, the dynamic time warping (DTW) algorithm effectively measures the shape similarity of runoff time series across different months, unaffected by sequence length and phase shift, as shown in Figure 6. The color of each cell in the figure represents the distance between the runoff dynamics of the two corresponding months; the darker the red, the smaller the distance and the more similar the patterns; the darker the blue, the larger the difference. From the figure, it is clear that the distances between January–February and November–December are very small, indicating that winter runoff processes have highly consistent patterns, with stable and continuously decreasing flow. In July–September, the colors within summer months are

also yellow, indicating that summer runoff, controlled by precipitation, exhibits similar fluctuation patterns. The region in the middle-right (March–June and October) forms a relatively compact yellow sub-block, while showing blue distances with the winter and summer sub-blocks. Moreover, October has a small distance with March–June but a large distance with winter months, suggesting that October is closer to spring than winter in terms of runoff dynamics.

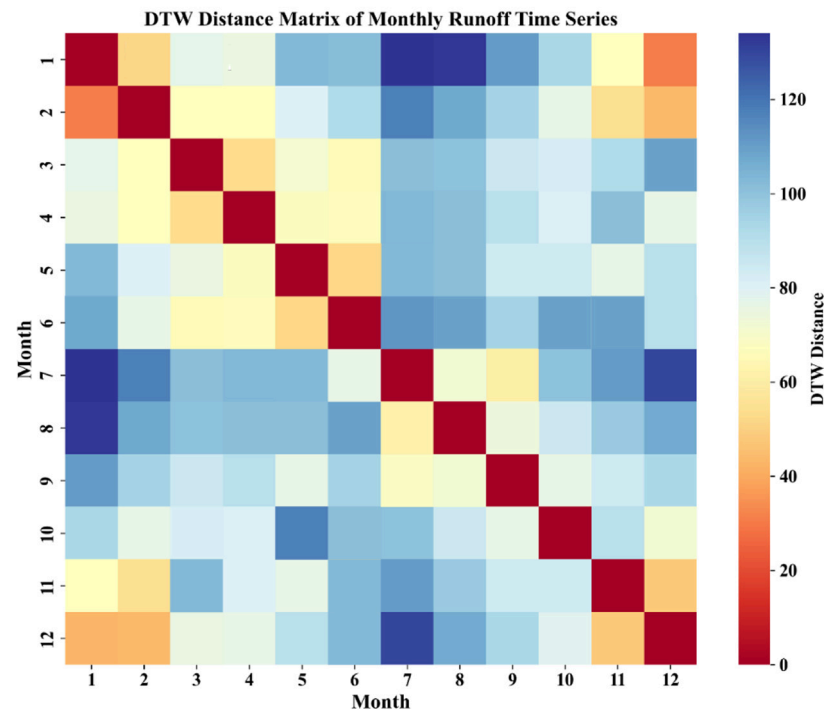


Figure 6. DTW autocorrelation distance matrix.

This distribution characteristic verifies the rationality of the three-period division from the perspective of sequence similarity: the internal patterns of November–February, March–June and October, and July–September are consistent, while significant differences exist among different periods. The heatmap result is fully consistent with the  $K = 3$  determined by the elbow method, providing strong dynamic similarity evidence for hydrological period division.

Based on the above analysis, the hydrological year is divided into three periods: November–February as the dry period, March–June and October as the transition period, and July–September as the wet period. The partial autocorrelation function is further applied to determine the optimal lag-order for each hydrological period, and the calculation results are summarised in Table 1. Figure 7 shows the intra-annual distribution of multi-year average monthly runoff at the Changtaiguan Hydrological Station, with the runoff ranges of the dry, transition, and wet periods marked by different colors. From the figure, it can clearly be seen that runoff during the dry period is low and stable, remaining within a small fluctuation range; runoff during the wet period rises significantly with prominent peaks; and the transition period exhibits a gradual change from dry to wet, intuitively confirming the rationality of the division and the seasonal non-stationary characteristics of runoff.

Table 1. PACF-based optimal lag order for each hydrological period.

| Period     | Months              | N   | phi_1 | phi_2 | phi_3 | phi_4 | Threshold Tau (Alpha = 0.05) | Selected Lag Order |
|------------|---------------------|-----|-------|-------|-------|-------|------------------------------|--------------------|
| Dry        | November–February   | 172 | 0.57  | 0.34  | 0.18  | 0.07  | $\pm 0.150$                  | 3                  |
| Transition | March–June, October | 215 | 0.30  | 0.25  | 0.16  | 0.08  | $\pm 0.134$                  | 3                  |
| Wet        | July–September      | 129 | 0.54  | 0.20  | 0.19  | 0.06  | $\pm 0.173$                  | 3                  |

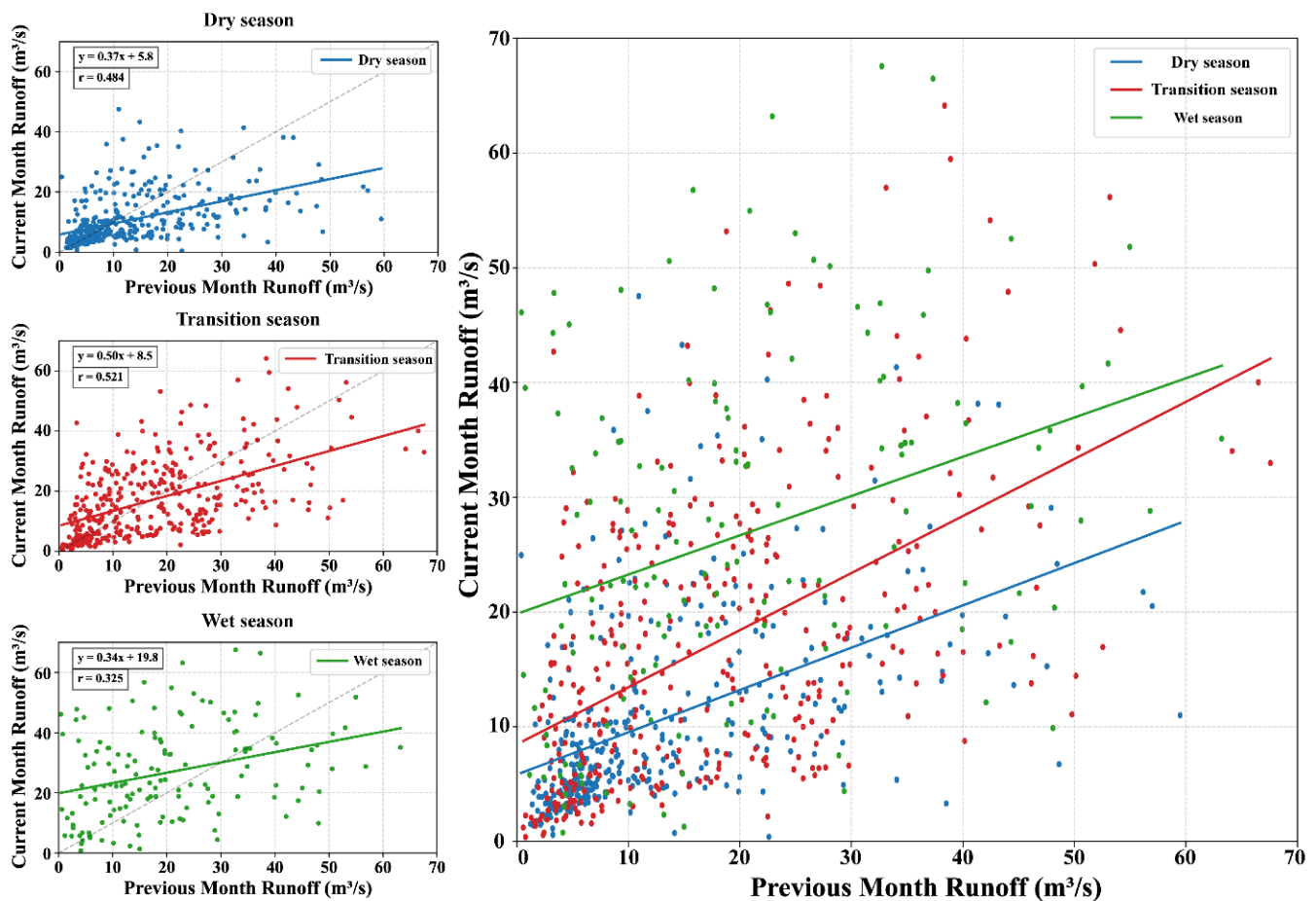


Figure 7. Autocorrelation periodization.

The PACF was computed on the training series of each period separately; because the number of available months differs among periods (172, 215 and 129 for the dry, transition and wet periods), the 95% threshold  $\tau = \pm 1.96 / \sqrt{N}$  is period specific ( $\pm 0.150$ ,  $\pm 0.134$  and  $\pm 0.173$ ). In all three periods, the coefficients remain above the threshold up to lag 3 and fall below it at lag 4, so a memory length of three months is adopted uniformly; the first three antecedent runoff terms (lag\_1–lag\_3) are, therefore, used as model inputs for all periods, keeping the input structure identical, while the period embedding allows the attention mechanism to weight them differently according to the memory characteristics of each period. The PACF measures linear marginal correlation, whereas SHAP quantifies the contribution of a feature within the trained nonlinear model, so the two need not coincide; the relatively high contribution of lag\_3 in the wet period is interpreted as the carry-over effect of pre-flood catchment storage rather than as simple autoregressive persistence.

#### 4.2. Model Comparison and Performance Evaluation

All configuration items were selected by a grid search evaluated on a validation subset carved from the end of the training period (the last five training years, 1998–2002), while the test set was touched only once, after the final configuration was fixed. The detailed architecture and training-related hyper-parameters of the PA-Informer model are summarized in Table 2.

To comprehensively evaluate model performance, this study designs different monthly runoff simulation and prediction models for comparison. Several machine learning and deep learning models are selected, and two sets of comparative experiments are designed. The first set consists of non-period prediction models, including XGBoost, LightGBM,

Random Forest (RF), and the standard Base-Informer without period information. The second set includes models incorporating hydrological period information: a period machine learning model (SML), a period-independent Informer (Indep-Informer), a transfer learning-based Informer (TL-Informer), a multi-task learning framework (MT-Informer), and the proposed period-adaptive attention Informer (PA-Informer). The basic characteristics of each model are summarized in Table 3.

**Table 2.** Architecture and training configuration of PA-Informer.

| Item                                     | Value                                                                                                                                                                                           |
|------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Input sequence length (look-back window) | 12 months                                                                                                                                                                                       |
| Prediction length                        | 1 month                                                                                                                                                                                         |
| Input features                           | Pt, Tt, $Q_{t-1}$ , $Q_{t-2}$ , $Q_{t-3}$                                                                                                                                                       |
| d_model (embedding dimension)            | 128                                                                                                                                                                                             |
| Attention heads                          | 4                                                                                                                                                                                               |
| Encoder layers/Decoder layers            | 2/1                                                                                                                                                                                             |
| Distilling layers                        | 1                                                                                                                                                                                               |
| Dropout                                  | 0.2                                                                                                                                                                                             |
| Feed-forward dimension                   | 256                                                                                                                                                                                             |
| Optimizer/Learning rate/Scheduler        | Adam, learning rate = $1 \times 10^{-4}$ ,<br>ReduceLROnPlateau scheduler                                                                                                                       |
| Batch size/Epochs                        | 16/100                                                                                                                                                                                          |
| Loss function                            | MSE                                                                                                                                                                                             |
| Early stopping (patience, monitor)       | patience = 10, monitor validation NSE (maximized)                                                                                                                                               |
| Random seed(s)                           | 10, 20, 30, 40, 50 (5 seeds)                                                                                                                                                                    |
| Hyperparameter search                    | Grid search over {d_model: 64, 128, 256; lr: $5 \times 10^{-5}$ ,<br>$1 \times 10^{-4}$ , $5 \times 10^{-4}$ ; dropout: 0.1, 0.2, 0.3}, selected on<br>a validation split from the training set |
| Framework/Hardware                       | PyTorch (v2.1, Meta AI, Menlo Park, CA, USA),<br>NVIDIA RTX3090 GPU (NVIDIA Corporation, Santa<br>Clara, CA, USA)                                                                               |

**Table 3.** Introduction to relevant models.

| Model          | Description                                                                                                                                                                                                                                                          |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SML            | For the dry season, transitional period and wet season, XGBoost, LightGBM and Random Forest were respectively used for modeling, fully leveraging the advantages of each algorithm in different hydrological periods                                                 |
| Base-Informer  | Using the entire sequence of precipitation, temperature and the antecedent runoff of the preceding three months (lag_1–lag_3) as inputs, without incorporating any phased information, as the benchmark model for deep learning                                      |
| Indep-Informer | An Informer model is trained independently for each hydrological period (dry season, transition period, wet season), and the parameters for each period are not shared                                                                                               |
| TL-Informer    | First, train an Informer model on the period with the most abundant samples (the dry season), and then transfer it to the transitional period and the wet season for fine-tuning to alleviate the overfitting problem caused by the sparse data in different periods |
| MT-Informer    | Share an Informer encoder, and set independent output heads for the three stages. Through the multi-task learning framework, optimize the prediction losses of all stages simultaneously                                                                             |
| PA-Informer    | Enable the model to automatically adjust the distribution of attention based on the period to which the input month belongs, thereby adaptively capturing the historical dependency lengths of different periods                                                     |

The experiment uses the Nash–Sutcliffe Efficiency (NSE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination ( $R^2$ ) as evaluation metrics. All models are trained on the monthly records of 1960–2020, divided chronologi-

cally into a training set (1960–2002) and a test set (2003–2020) at a ratio of 7:3; a validation subset (1998–2002) is further carved from the end of the training set for hyperparameter selection and early stopping, and the results are compared based on the test set.

#### 4.2.1. Performance of Non-Periodic Models

To objectively evaluate the applicability of traditional time-series forecasting methods for monthly runoff simulation at the Changtaiguan Hydrological Station, this section first compares the prediction performance of benchmark models without considering hydrological periods, including three classical machine learning models (XGBoost, LightGBM, and RF) and the Base-Informer deep learning baseline model without period information. Full-sequence precipitation, temperature, and antecedent runoff of the preceding three months (lag\_1–lag\_3) are used as inputs to quantitatively reveal the limitations of single models in simulating non-stationary runoff across the entire sequence. The results are shown in Table 4. The three traditional machine learning models achieve similar fitting performance, with an  $R^2$  between 0.84 and 0.85 and an NSE between 0.83 and 0.85, indicating stable but limited accuracy. In terms of error metrics, Random Forest has a relatively larger RMSE and MAE, performing slightly worse for extreme runoff. XGBoost and LightGBM have better error metrics but still exhibit obvious underestimation in high-flow periods during the flood season and slight overestimation in low-flow periods during the dry season.

**Table 4.** Evaluation indicators of non-periodic model.

| Model         | $R^2$           | NSE             | RMSE            | MAE             |
|---------------|-----------------|-----------------|-----------------|-----------------|
| XGBoost       | $0.85 \pm 0.01$ | $0.84 \pm 0.01$ | $3.18 \pm 0.05$ | $2.76 \pm 0.04$ |
| LightGBM      | $0.85 \pm 0.01$ | $0.85 \pm 0.01$ | $3.15 \pm 0.06$ | $2.74 \pm 0.05$ |
| RandomForest  | $0.84 \pm 0.01$ | $0.83 \pm 0.02$ | $4.24 \pm 0.09$ | $3.53 \pm 0.07$ |
| Base-Informer | $0.90 \pm 0.01$ | $0.89 \pm 0.01$ | $2.82 \pm 0.12$ | $1.61 \pm 0.09$ |

To further analyze the prediction performance of non-period models, linear-fitting analysis is performed for XGBoost, LightGBM, RF, and Base-Informer, as shown in Figure 8. The scatter plots of predicted versus observed values show that the fitted points of all non-period models are generally distributed along the 1:1 diagonal, indicating that the models can capture the overall trend of monthly runoff, but deviations occur in both high-flow and low-flow ranges. Among them, Base-Informer has more concentrated fitted points and the smallest deviation from the diagonal, with an  $R^2$  of 0.90, clearly outperforming the three machine learning models, indicating that the deep learning models have a stronger ability to extract nonlinear relationships and long-term dependencies in runoff series.

Overall, although non-period models can achieve basic runoff trend fitting, they lack the ability to adaptively characterize the stable runoff in dry periods and abrupt runoff changes in flood seasons due to ignoring the seasonal differentiation of runoff autocorrelation and driving factors, resulting in limited overall accuracy. This confirms the necessity of introducing hydrological periods and constructing adaptive prediction models.

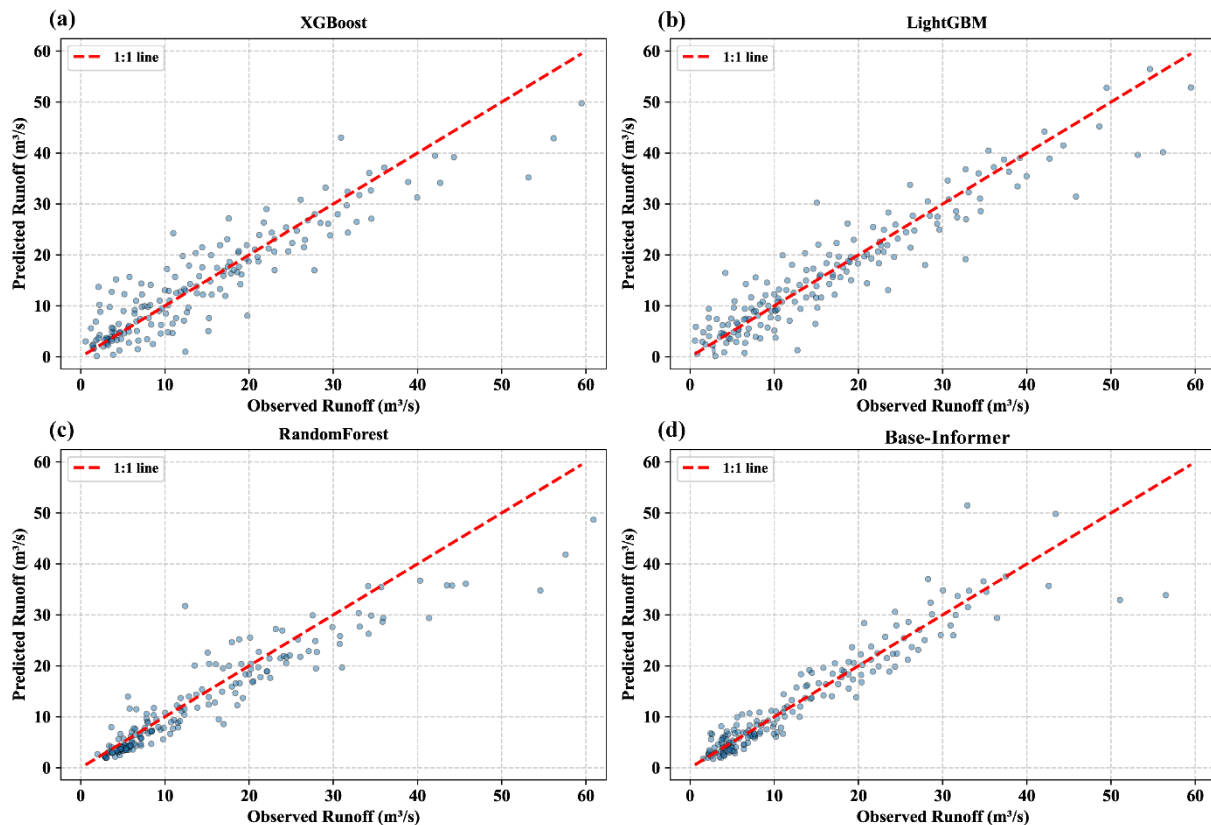


**Figure 8.** Linear fitting of predicted values for the non-segmented model.

#### 4.2.2. Performance of Period-Based Models

To verify the improvement effect of hydrological period division and adaptive modeling on monthly runoff simulation, this section compares various period-based models, including SML, Indep-Informer, TL-Informer, MT-Informer, and the proposed PA-Informer. With the observed runoff series as a reference, the simulation advantages of period models are comprehensively evaluated using time-series curves, linear-fitting scatter plots, and quantitative evaluation metrics. The results are shown in Table 5 and Figure 9.

**Table 5.** Evaluation indicators of the phased model.

| Model          | $R^2$             | NSE               | RMSE            | MAE             |
|----------------|-------------------|-------------------|-----------------|-----------------|
| SML            | $0.91 \pm 0.01$   | $0.90 \pm 0.01$   | $1.25 \pm 0.06$ | $0.92 \pm 0.05$ |
| Indep-Informer | $0.93 \pm 0.01$   | $0.92 \pm 0.01$   | $1.06 \pm 0.07$ | $0.78 \pm 0.05$ |
| TL-Informer    | $0.94 \pm 0.01$   | $0.93 \pm 0.01$   | $0.94 \pm 0.06$ | $0.71 \pm 0.04$ |
| MT-Informer    | $0.94 \pm 0.01$   | $0.94 \pm 0.01$   | $0.88 \pm 0.05$ | $0.66 \pm 0.04$ |
| PA-Informer    | $0.957 \pm 0.004$ | $0.948 \pm 0.005$ | $0.71 \pm 0.04$ | $0.52 \pm 0.03$ |

As shown in Table 5, the prediction performance of all period-based models is markedly improved compared with non-period models. In terms of  $R^2$  and NSE, the best-performing non-period model, Base-Informer, achieves only 0.90 and 0.89, while all period-based models have  $R^2 > 0.91$  and  $NSE > 0.90$ . Among them, the proposed PA-Informer achieves an  $R^2$  of 0.957 and an NSE of 0.948, outperforming all other models by a clear margin. In terms of error metrics, Base-Informer has an RMSE of 2.82 and an MAE of 1.61, whereas all period-based models show substantially lower errors. Among them, SML has relatively higher errors (RMSE: 1.25, MAE: 0.92), while Indep-Informer, TL-Informer, and MT-Informer show progressively lower errors, with PA-Informer achieving the best

error control (RMSE: 0.71, MAE: 0.52), corresponding to reductions of 74.8% and 67.7% compared with the non-period model.

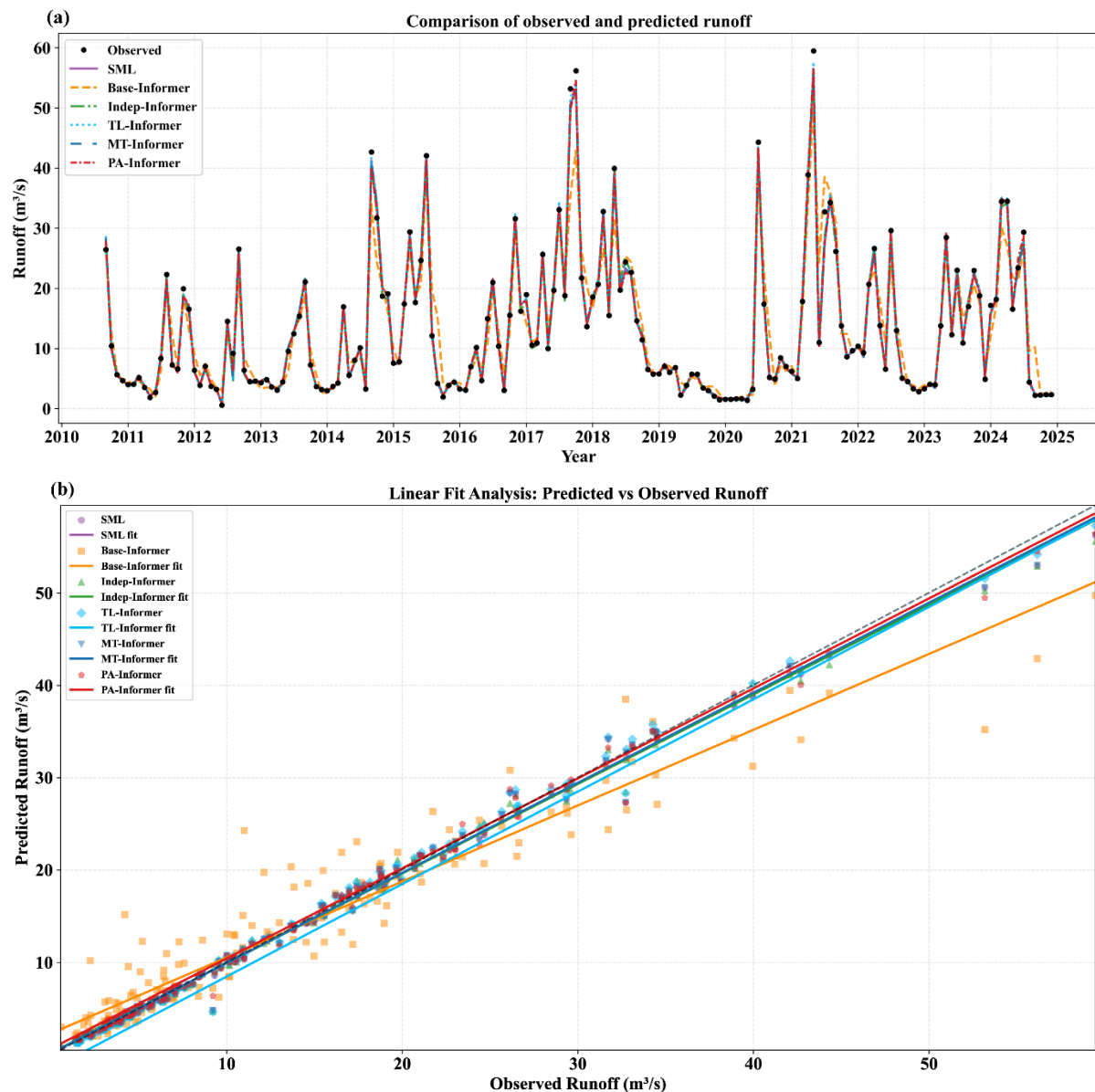


**Figure 9.** Comparison of predicted values from the phased model.

To account for the randomness of network initialization, each deep learning model was trained with five independent random seeds, and all metrics in Tables 4 and 5 are reported as the mean  $\pm$  standard deviation over these runs. The statistical significance of the improvement in PA-Informer over the best benchmark (Base-Informer) was assessed with a two-sided Wilcoxon signed-rank test applied to the monthly absolute errors of the test set ( $p = 0.0023$ ). Bootstrap 95% confidence intervals (1000 resamples of the test months) of the NSE and RMSE are [0.937, 0.956] and [0.62, 0.82] m<sup>3</sup>/s, respectively.

From Figure 9a, it can be clearly seen that the period-based models markedly reduce the systematic biases of the non-period models (underestimation in dry periods and overestimation in wet periods). In the dry period, all period-based models show a remarkable improvement in tracking the stable, low-amplitude baseflow, with curve fluctuations much closer to the observations. In the transition period, they better capture the gradual process from dry to wet. In the wet period, the response to precipitation-driven runoff peaks is

timelier, and the deviation of peak predictions is noticeably alleviated. Among all models, PA-Informer achieves the highest agreement with the observed series and the most accurate reproduction of runoff processes across dry, wet, and transition periods.

From the linear-fitting scatter plot in Figure 9b, the period strategy clearly improves the overall fitting degree, with fitted points of all period-based models highly concentrated around the 1:1 diagonal. Data points in the dry period are more compact with smaller deviations, reflecting accurate characterization of highly autocorrelated, strongly persistent runoff. The dispersion of data points in the transition period is considerably reduced compared with the non-period models. The scatter of high-flow points in the wet period is effectively constrained. PA-Informer exhibits the smallest dispersion and the best linear correlation across all three periods, with an  $R^2$  as high as 0.957.

The quantitative results combined with hydrological period division further indicate that runoff in the dry period has strong autocorrelation and is relatively stable; transfer learning and multi-task learning models effectively improve simulation stability under small-sample conditions through parameter sharing and pre-training fine-tuning strategies. In the transition period, runoff variability is large and the hydrological driving mechanism is complex; the independent training model achieves targeted fitting of the runoff transition process by having separate model parameters for each period. In the wet period, runoff fluctuates dramatically due to precipitation pulses, and conventional models tend to produce large deviations [42]. The PA-Informer model, with its period-adaptive attention mechanism, dynamically matches the runoff dependence characteristics of different periods, thus achieving optimal simulation of flood season peaks and fluctuation processes.

Overall, introducing hydrological periods greatly improves the prediction accuracy of all models compared with the non-period models, proving that the division into dry, transition, and wet periods effectively captures the seasonal non-stationarity of runoff. The proposed PA-Informer model performs best overall, leading in all four metrics ( $R^2$ , NSE, RMSE, and MAE), indicating that its adaptive attention mechanism can precisely match the runoff memory characteristics and driving mechanisms of different hydrological periods, making it highly suitable for high-accuracy monthly runoff simulation at the Changtaiguan Hydrological Station.

#### 4.3. Analysis of Monthly Runoff Driving Factors

##### 4.3.1. SHAP-Based Contribution Analysis of Key Driving Factors

To quantify the contribution differences of key driving factors for monthly runoff in the Chushandian Irrigation Area across different hydrological periods, this study uses the SHAP (SHapley Additive exPlanations v0.41.0) interpretability method based on the best-performing PA-Informer model. Antecedent runoff (lag\_1, lag\_2, and lag\_3), monthly precipitation (Prcp), and monthly mean temperature (Temp) are selected as the main input features. The mean absolute SHAP values and contribution directions of each factor are calculated for the dry, transition, and wet periods, revealing the dominant mechanisms of runoff formation in different hydrological stages, as shown in Figure 10.

From Figure 10a (dry period), the SHAP feature importance ranking shows that the factors contribute to monthly runoff simulation output in the following order: temperature (Temp), lag\_2, lag\_1, precipitation (Prcp), and lag\_3. Temperature dominates the dry period, having the most significant impact on runoff. Antecedent runoff (lag\_2 and lag\_1) also shows high contributions, reflecting the significant temporal persistence and memory effect of dry-period runoff. Precipitation and lag\_3 have relatively weak contributions, indicating limited precipitation recharge during the dry period, and runoff processes are mainly controlled by temperature and antecedent hydrological conditions, consistent with the stable, low-amplitude characteristics of dry-period runoff.

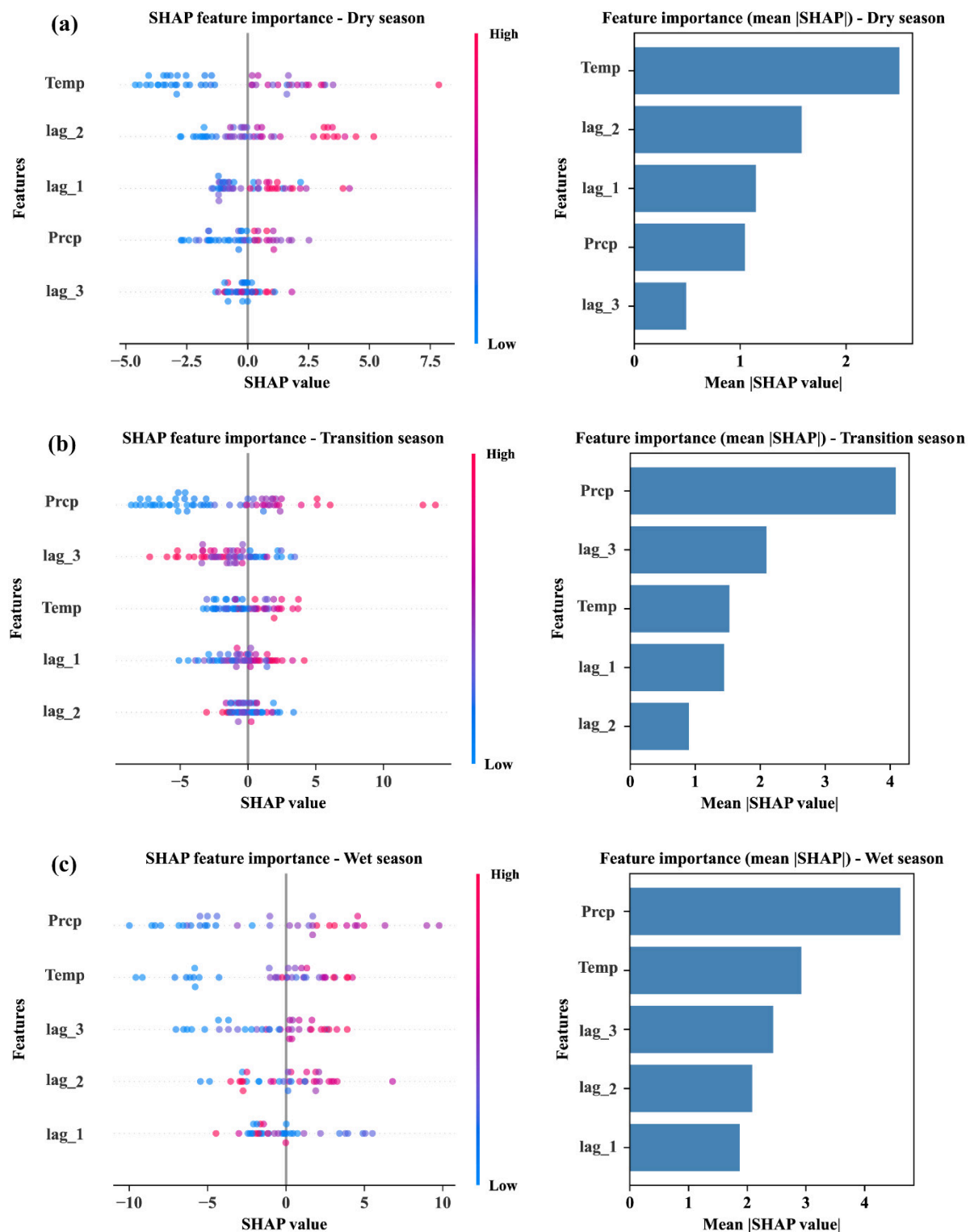


Figure 10. Importance of SHAP features for hydrological stages.

From Figure 10b (transition period), the SHAP importance ranking is the following: precipitation (Prcp), lag\_3, temperature (Temp), lag\_1, and lag\_2. As precipitation increases, the contribution of precipitation rises significantly and becomes the primary driving factor in the transition period. The importance of lag\_3 increases notably, reflecting the dependence of runoff on longer-lag hydrological conditions during this stage. Temperature maintains a moderate influence, while lag\_1 and lag\_2 have relatively lower contributions. Overall, this shows a mixed driving pattern dominated by precipitation, assisted by long-

lag runoff, and co-regulated by temperature, consistent with the evolution pattern of runoff transitioning from dry to wet with increasing variability.

From Figure 10c (wet period), the SHAP importance ranking is the following: precipitation (Prcp), temperature (Temp), lag\_3, lag\_2, and lag\_1. Precipitation overwhelmingly dominates as the core driving factor for wet-period runoff, controlling the storm runoff generation and concentration processes in the flood season. Temperature ranks second, mainly regulating runoff through its influence on watershed evapotranspiration. Among antecedent runoff factors, lag\_3 contributes more than lag\_2 and lag\_1, indicating that wet-period runoff depends more on long-cycle memory rather than recent runoff, as the influence of recent runoff is significantly weakened by strong precipitation signals.

In summary, the monthly runoff driving mechanism at the Changtaiguan Hydrological Station exhibits significant seasonal non-stationarity: the dry period is dominated by temperature and recent antecedent runoff, the transition period is co-driven by precipitation and long-lag antecedent runoff, and the wet period is predominantly driven by precipitation. The SHAP analysis results not only quantify the contribution differences of each factor across different periods but also confirm the scientific validity of the hydrological period division and the rationality of the PA-Informer model's adaptive simulation from a physical mechanism perspective, providing a quantitative basis for runoff evolution attribution and water resources management in the irrigation area.

#### 4.3.2. Response Characteristics of Key Driving Factors

Based on the identification of the dominant drivers in each hydrological period, SHAP dependence plots are further used to reveal the marginal effects, nonlinear responses, and threshold characteristics of key driving factors on model output. This provides a quantitative basis for understanding the physical mechanisms of runoff formation and hydrological process attribution. The results are shown in Figure 11.

For the dry period (Figure 11a), temperature and antecedent runoff, as dominant factors, exhibit a significantly linear monotonic SHAP dependence relationship. The SHAP value of temperature increases steadily with rising temperature, indicating that in a low-temperature environment, temperature positively drives runoff through controlling frozen soil, snowmelt recharge, and baseflow release, without an obvious inflection point, suggesting that the runoff response to temperature in the dry period is nearly linear. The SHAP dependence curves of antecedent runoff (lag\_1 and lag\_2) also show a smooth increasing trend, reflecting the strong temporal dependence of dry-period runoff, with a clear memory effect and continuity of hydrological state, highly consistent with the stable, low-amplitude characteristics of dry-period runoff.

Entering the transition period (Figure 11b), precipitation becomes the core driving factor, and its SHAP dependence shows a distinctly nonlinear enhancement characteristic. As precipitation increases, the SHAP value rises rapidly, and the growth rate expands, indicating a magnification effect of precipitation on runoff (weak driving at low precipitation, strong driving at high precipitation), consistent with the gradual activation of runoff generation mechanisms before the flood season. Meanwhile, the SHAP dependence curve of long-lag antecedent runoff (lag\_3) maintains a smooth positive contribution, indicating that transition-period runoff is simultaneously affected by current precipitation recharge and the continuation of antecedent hydrological states, resulting in a more complex response structure, reflecting the mixed driving characteristics of the dry-to-wet transition stage.

In the wet period (Figure 11c), the SHAP dependence of precipitation exhibits a significantly nonlinear steep increase trend: the greater the precipitation, the stronger the positive driving effect on runoff, showing a clear threshold amplification characteristic, consistent with the physical mechanism of excess infiltration runoff and accelerated flow

concentration under heavy rainfall conditions. At the same time, the SHAP value of temperature decreases gradually with rising temperature, indicating a negative response of runoff to increasing temperature, as high temperatures suppress runoff by enhancing watershed evapotranspiration. This forms a typical response pattern in the wet period dominated by precipitation and suppressed by temperature. The SHAP dependence intensity of antecedent runoff is significantly weakened, with the long-lag term (lag\_3) dominating, indicating that heavy precipitation has become the dominant factor, and the memory effect of recent runoff is substantially weakened.

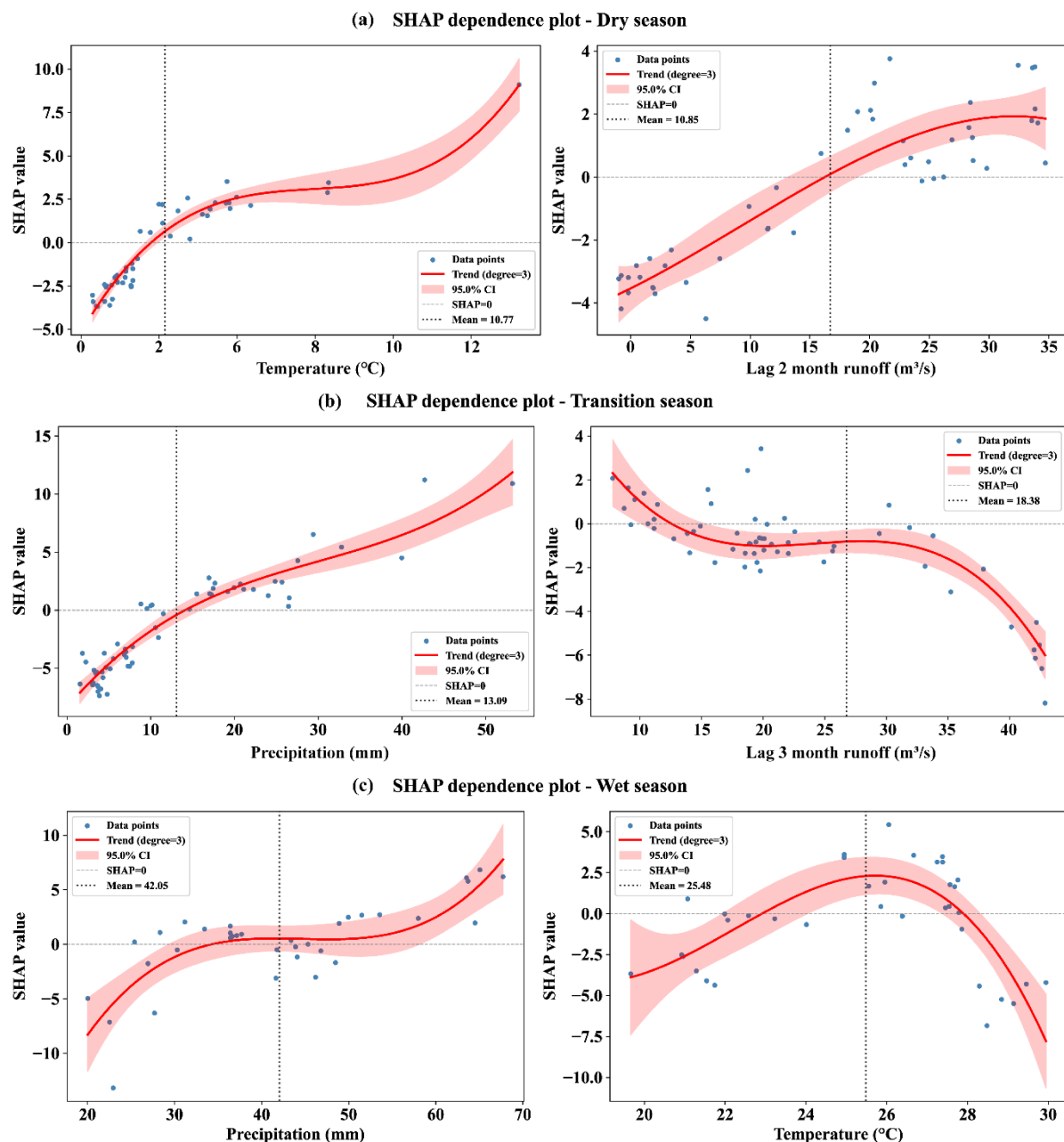


**Figure 11.** SHAP dependence plots of the key driving factors in the (a) dry, (b) transition and (c) wet periods.

Overall, the SHAP dependence and response characteristics of key driving factors differ significantly across hydrological periods: the dry period is dominated by linear driving and strong continuity, the transition period exhibits nonlinear mixed driving, and the wet period shows strong precipitation-driven and temperature-inhibited co-action. This

result not only verifies the rationality of the hydrological period division in this study but also reveals the runoff formation mechanisms at different stages from the perspective of model interpretability, providing reliable physical support for the simulation results.

#### 4.3.3. Synergistic Driving of Precipitation and Temperature and Period Based Regulation Effects

Based on clarifying the contributions of key factors in each period, this section further examines the synergistic evolution and regulation effects of precipitation and temperature, the direct drivers of monthly runoff. SHAP dependence analysis for the precipitation–temperature interaction is conducted, with the results shown in Figure 12.

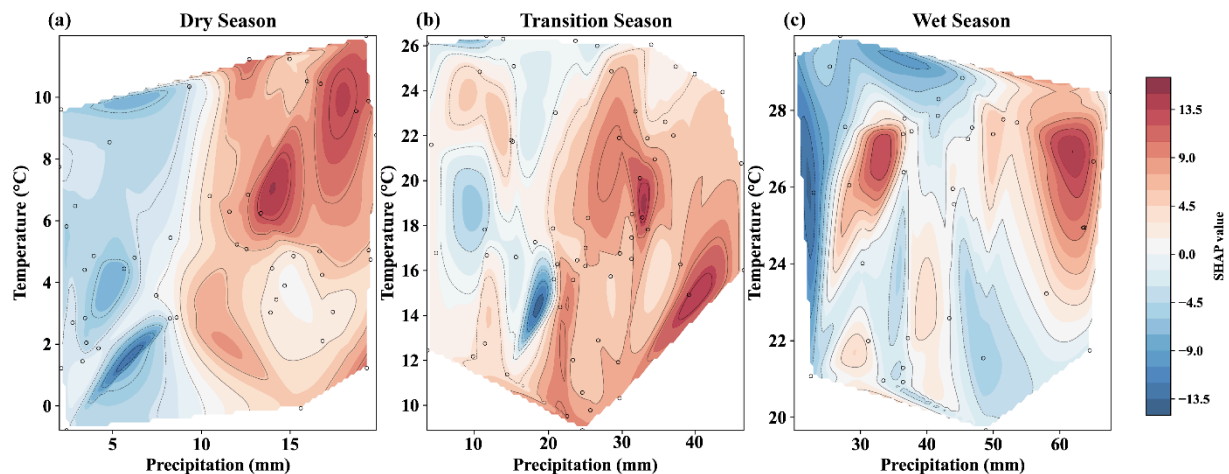


Figure 12. Interaction effect of precipitation and temperature.

During the dry period, the interactive dependence between precipitation and temperature shows weak coupling and one-way dominance characteristics (Figure 12a). During this period, the basin is generally characterized by low temperature and low precipitation. Precipitation recharge is limited and exhibits small variability, playing a weak role in driving runoff, with a flat and stable SHAP response curve. Temperature, on the other hand, exerts a significant and stable positive regulation on runoff by controlling frozen soil depth, baseflow recharge rate, and snowmelt processes. An increase in temperature effectively enhances groundwater runoff release, making runoff show a linear increasing trend. Overall, the dry period is dominated by weak interaction with temperature dominance and precipitation assistance; the interaction intensity is low, and runoff variation primarily reflects temperature-driven baseflow continuity, further confirming that dry-period runoff is characterized by stable recharge and slow fluctuations.

During the transition period, the interaction between precipitation and temperature significantly strengthens, exhibiting synergistic driving and bidirectional coupling characteristics (Figure 12b). As seasons shift, precipitation frequency and amount gradually increase, becoming the main source of runoff recharge. The driving effect on runoff rapidly strengthens and shows clear nonlinearity. Temperature indirectly affects the conversion efficiency of precipitation to runoff by regulating evapotranspiration intensity, soil infiltration capacity, and vegetation water consumption, thereby constraining the precipitation-driven effect. At this stage, precipitation provides water supply, while temperature regulates water distribution. The coupling effect strengthens, forming a synergistic driving pattern where precipitation dominates runoff generation and temperature regulates water consumption. Consequently, transition-period runoff is jointly constrained by water supply and energy conditions, showing a more complex response structure and reflecting the transitional characteristics of the hydrological system from dry to wet.

During the wet period, precipitation and temperature exhibit strongly antagonistic and highly nonlinear interactive characteristics, representing the most significant coupling stage of the whole year (Figure 12c). In the flood season, precipitation has high intensity and concentration, exerting an extremely strong positive driving effect on runoff, and the driving effect amplifies nonlinearly with increasing precipitation, directly dominating the sharp rise in runoff. Temperature, by increasing evapotranspiration and accelerating soil water consumption, significantly inhibits runoff generation. High temperatures markedly reduce the efficiency of converting precipitation into runoff, forming a typical antagonistic mechanism of “precipitation enhancing runoff, temperature weakening runoff”. The two factors counterbalance and jointly determine the dramatic fluctuations and prominent peaks of flood-season runoff. This interaction mechanism also explains why the PA-Informer model achieves higher simulation accuracy during abrupt flood-season changes: its adaptive attention mechanism effectively captures this strongly nonlinear, strongly antagonistic water–heat coupling relationship.

In summary, the synergistic interaction pattern of precipitation and temperature shows significant stage-dependent differences across hydrological periods: weak interaction with temperature dominance in the dry period, synergistic coupling and co-driving in the transition period and strongly antagonistic, nonlinear balancing in the wet period, which is the most intense water–heat interaction period of the year. This result, from the perspective of multi-factor synergistic evolution, deepens the understanding of the monthly runoff driving mechanism in the Chushandian Irrigation Area, not only verifying the physical rationality of the hydrological period division but also providing an interpretable basis for the advantages of PA-Informer’s adaptive simulation, and can offer theoretical support for runoff evolution attribution and hydrological prediction model construction in irrigation areas.

#### 4.4. Sensitivity Analysis

To further quantify the impact of input perturbations on model output, this study conducts single-factor perturbation sensitivity analysis based on the optimal model. Systematic perturbations of  $\pm 10\%$  and  $\pm 20\%$  are applied to precipitation, temperature, and antecedent runoff separately while keeping other input factors unchanged. The average change rate of simulated output relative to the original output is calculated to quantitatively identify the sensitivity of the model to each input factor across different hydrological periods. The results are shown in Figure 13. From the boxplot distribution characteristics, the dispersion and stability of model outputs differ significantly across hydrological periods. The dry period shows the most concentrated output distribution (Figure 13a), with a small median fluctuation range and the narrowest interquartile range, indicating that the dry-period model is least affected by input perturbations and has the best overall stability. The dispersion gradually expands in the transition period (Figure 13b), with increased median shift and distribution range, indicating gradually increasing model sensitivity. The wet period shows the largest dispersion (Figure 13c), with the widest interquartile range, indicating that the model is most sensitive to input perturbations during high-flow stages, though the overall distribution remains regular without extreme outliers.

Because a percentage change in temperature expressed in degrees Celsius is scale-dependent, the absolute increments corresponding to  $\pm 10\%$  and  $\pm 20\%$  are reported for each period: about a period-mean temperature of  $6.2\text{ }^{\circ}\text{C}$  in the dry period, the perturbations correspond to  $\pm 0.62\text{ }^{\circ}\text{C}$  and  $\pm 1.24\text{ }^{\circ}\text{C}$ ; the corresponding values are  $\pm 1.64/\pm 3.28\text{ }^{\circ}\text{C}$  in the transition period and  $\pm 2.57/\pm 5.14\text{ }^{\circ}\text{C}$  in the wet period. It should be noted that in the wet period  $\pm 20\%$  corresponds to  $\pm 5.14\text{ }^{\circ}\text{C}$ , which exceeds the inter-annual standard deviation of monthly temperature in that period ( $1.6\text{ }^{\circ}\text{C}$ ) by a factor of about three; this scenario is, therefore, interpreted as a stress test of the model’s extrapolation behavior rather than as a realistic

range of observational error. For this reason, a supplementary absolute-perturbation experiment ( $\pm 0.5$  °C and  $\pm 1.0$  °C, i.e., within one standard deviation of the observed inter-annual variability: 2.4, 2.0 and 1.6 °C for the dry, transition and wet periods) was also conducted, and it reproduced the same period-specific sensitivity ranking.

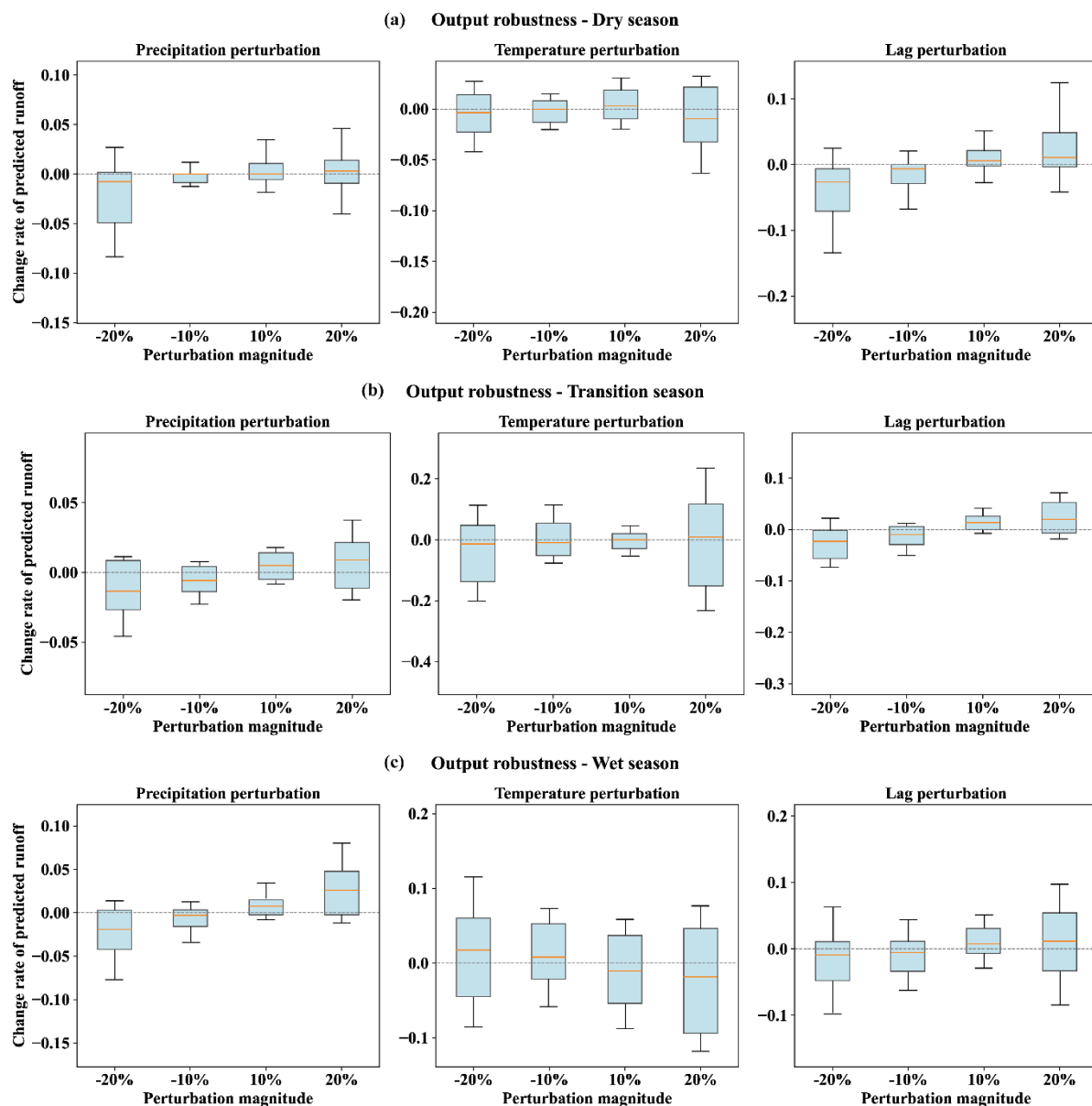


**Figure 13.** Box plot of monthly runoff changes under different disturbance scenarios.

Here, robustness is defined not as numerical smoothness of the response but as (i) the output change rate remains bounded ( $\leq 8.7\%$  under  $\pm 20\%$  perturbation); (ii) the ranking of the dominant factor in each period does not change anywhere within the perturbation range; and (iii) the response curves remain monotone and free of sign reversal. Under these criteria, the model's period-specific sensitivity structure is stable under input uncertainty.

To further reveal the periodic response patterns of model output to perturbations of different factors, the average response of predicted runoff to input perturbations is shown in Figure 14. From the line chart response trends, the model output under different factor perturbations exhibits clear period-specific response patterns.

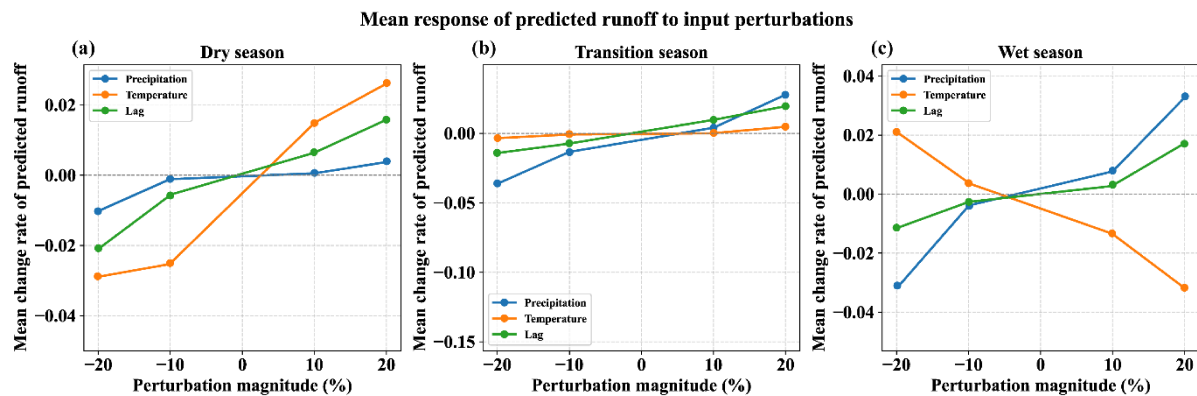


Figure 14. Curve of monthly runoff change rate under different disturbance scenarios.

As shown in Figure 14a, during the dry period, model output is most sensitive to temperature, followed by antecedent runoff, and least sensitive to precipitation. Under  $\pm 10\%$  and  $\pm 20\%$  perturbations of temperature, the output change rate shows a clear linear response, with nearly symmetric positive and negative perturbation amplitudes, indicating that dry-period output strongly depends on temperature-driven baseflow recharge, and the model is stable to temperature changes. Antecedent runoff perturbations also produce an approximately linear positive response, with output changing smoothly as antecedent runoff increases or decreases, reflecting the continuity of the dry-period hydrological state. Under precipitation perturbations, the output change rate is weak and flat, indicating limited precipitation recharge during the dry period and no significant impact on model output. Overall, the dry-period model has good robustness, and the response pattern of sensitive factors is highly consistent with the driving mechanism analysis.

As shown in Figure 14b, during the transition period, the model sensitivity changes significantly: precipitation becomes the most sensitive factor, followed by antecedent runoff, while temperature shows an insensitive but positive response. The output change rate caused by precipitation perturbations is markedly higher than in the dry period and increases nonlinearly with perturbation magnitude, indicating that transition-period output depends increasingly on precipitation recharge. Antecedent runoff retains a moderate sensitivity, with a smooth and linear response trend, indicating that long-lag hydrological conditions still have a continuous regulating effect on output. Temperature perturbation shows a low-sensitivity, stable positive response: a small increase in temperature slightly promotes output, but the overall response amplitude is small and sensitivity is weak, without significant inhibition. The response curves of all factors are smooth and continuous without sharp jumps, indicating that the model still has good output robustness during the non-stationary dry-to-wet transition stage.

As shown in Figure 14c, during the wet period, the model shows extremely strong positive sensitivity to precipitation, the most significant stage of the whole year. Perturbations of  $\pm 10\%$  and  $\pm 20\%$  in precipitation cause large changes in output, with the increase from positive perturbations significantly larger than the decrease from negative perturbations, exhibiting a typical amplification effect consistent with the nonlinear physical mechanism of flood-season heavy rainfall runoff generation and concentration. Temperature shows a significant negative sensitivity: warming clearly suppresses output, and the sensitivity is higher than that in the dry and transition periods, consistent with the interactive pattern of precipitation driving and temperature suppressing runoff. The sensitivity of antecedent runoff is the lowest of the year, with a flat perturbation response, indicating that heavy precipitation signals have become predominant and the historical hydrological memory effect is substantially weakened. Although hydrological fluctuations are severe in the

wet period, the model maintains continuous and stable response characteristics without abnormal deviations, demonstrating reliable output robustness.

From the overall perspective, the sensitivity of the model to different input factors exhibits significant period-specific differentiation: in the dry period, the order of sensitivity is temperature, antecedent runoff, and precipitation; in the transition period, it is precipitation, antecedent runoff, and temperature; in the wet period, it is precipitation, temperature, and antecedent runoff. This order is fully consistent with the driving mechanism analysis, further verifying the physical rationality and interpretability of the model simulations. Under continuous perturbations of  $\pm 10\%$  to  $\pm 20\%$ , the model output in each period shows no abrupt changes, oscillations, or abnormal deviations; the response curves are smooth, orderly, and directionally clear, indicating that the period-adaptive attention model constructed in this study has good input anti-interference ability and output robustness and can adapt to the nonlinear variation characteristics of different hydrological periods, providing stable support for reliable monthly simulation in the irrigation area.

#### 4.5. Independent Hydrological Evidence

To test whether the associations identified by SHAP are consistent with independent hydrological evidence, three diagnostic statistics were computed directly from the observations, without reference to the model. (i) The period-specific runoff coefficient  $\alpha = Q/P$ , obtained by converting monthly discharge to runoff depth over the catchment area, is 0.14 in the dry period and 0.47 in the wet period, indicating that only a small fraction of precipitation is converted to runoff under low-precipitation, low-temperature conditions, where runoff is sustained mainly by baseflow. (ii) After controlling for precipitation, the partial correlation between monthly runoff and temperature is 0.42 ( $p < 0.001$ ) in the dry period and  $-0.35$  ( $p < 0.001$ ) in the wet period, consistent with temperature-driven baseflow release in winter and evapotranspiration suppression in summer. (iii) Baseflow separation by the Lyne–Hollick filter (BFI<sub>max</sub> = 0.80, three forward-backward passes) gives a period-specific baseflow index of 0.81 in the dry period versus 0.34 in the wet period, confirming that dry-period runoff is dominated by slow subsurface storage release and, therefore, exhibits strong memory.

These three lines of evidence are model-independent and are consistent with the SHAP-based attribution; they should be read as corroborating evidence rather than as a causal proof, because SHAP quantifies model-based association rather than causal mechanism.

## 5. Discussion

### 5.1. Model Advantages and Innovations

The PA-Informer model proposed in this study demonstrates significant advantages in monthly runoff simulation. First, by introducing a period adaptive attention mechanism, the model dynamically adjusts attention weights according to the different hydrological characteristics of dry, transition, and wet periods, achieving an adaptive characterization of runoff seasonal non-stationarity. Although the Informer model exhibits advantages in long sequence prediction due to its probabilistic sparse self-attention mechanism, the standard Informer does not account for differences in intra-seasonal driving mechanisms of hydrological series. The SMGformer model proposed by Wang et al. incorporates STL decomposition and multi-head self-attention into the Informer, which improves multi-step runoff prediction accuracy to some extent [43]. However, that model still adopts a fixed attention mechanism and cannot assign differentiated memory lengths to different hydrological periods. In contrast, PA-Informer directly embeds prior knowledge of hydrological periods into the attention computation, enabling the model to dynamically adjust the correlation strength between queries and keys according to the dry, transition, and wet

periods. As a result, PA-Informer achieves superior results in all four metrics ( $R^2$ , NSE, RMSE, and MAE), with particularly notable improvements in fitting low flows in the dry period and peaks in the wet period, indicating that integrating hydrological periodicity effectively alleviates the “performance averaging” problem of full sequence modeling.

Second, PA-Informer maintains the low computational complexity of the Informer model while adding only a minimal number of period-embedding parameters, leading to high training efficiency and ease of deployment. Compared with period-independent training, transfer learning, and multi-task learning strategies, PA-Informer achieves end-to-end adaptive adjustment through its embedded period attention, with a much simpler structure. Although previous studies have verified the effectiveness of multi-task learning frameworks in runoff prediction, multi-task head designs still involve parameter redundancy and training burdens in the context of period-based modeling, whereas PA-Informer avoids these issues.

Furthermore, SHAP interpretability analysis reveals the differentiation laws and nonlinear response characteristics of runoff driving factors across different hydrological periods, enhancing the physical interpretability of the model. Tian et al. introduced SHAP analysis into the SWAT LSTM hybrid framework to quantify the driving contributions of factors such as precipitation and temperature and identified threshold inflection points [44]. Lu et al. pointed out, through cross-comparison and mechanistic interpretation of deep learning models, that the predictive ability of the models stems from their effective capture of temporal dependencies and physical triggering mechanisms [45]. Compared with these studies, PA-Informer combined with SHAP further reveals the stage-wise evolution of precipitation temperature interaction effects across dry, transition, and wet periods (weak coupling  $\rightarrow$  synergistic driving  $\rightarrow$  strong antagonism), a pattern that is highly consistent with the actual hydrological processes of the basin and verifies the rationality of the period division. Regarding sensitivity analysis, the VARS TOOL developed by Razavi et al. provides a systematic method for parameter perturbation assessment [46]. Under multiple perturbation scenarios of  $\pm 10\%$  to  $\pm 20\%$ , this study verifies that PA-Informer responds smoothly and orderly to input perturbations, and the ranking of sensitive factors in different periods is completely consistent with the physical driving mechanisms, indicating that the model has good anti-interference ability and is suitable for practical water resources management decision making.

## 5.2. Limitations and Future Work

Although PA-Informer achieves good simulation performance, some limitations remain. First, the hydrological period division is determined based on historical average runoff and clustering methods. Under climate change, the intra-annual distribution of runoff may shift, and fixed period boundaries may reduce the model’s adaptability. Future work could introduce sliding windows or dynamic clustering algorithms to achieve adaptive updating of period division.

Second, the input features of the model only include precipitation, temperature, and antecedent runoff, without considering factors such as evaporation, soil moisture, and land use change. Although SHAP analysis indicates that the selected features already explain most of the runoff variability, introducing more physical driving factors (e.g., potential evapotranspiration, and vegetation indices) is expected to further improve the simulation accuracy of extreme runoff.

Third, although PA-Informer performs excellently at the monthly scale, its ability to respond to extreme storm events at the daily or hourly scale has not yet been validated. Future work could extend the model to finer time scales and incorporate data assimilation techniques to improve real time forecasting capability.

Finally, model hyperparameters (such as the attention sparsity coefficient and number of distilling layers) have a significant impact on prediction performance. This study uses grid search for parameter optimization. Future research could introduce Bayesian optimization or automated machine learning (AutoML) strategies to search the hyperparameter space more efficiently, thereby further improving model efficiency and generalization ability.

In addition, because the meteorological variables of the target month are used as inputs, the present configuration is a diagnostic simulation rather than a real-time forecast. Operational application would require replacing Pt and Tt with seasonal meteorological forecasts or climate-model products, which introduces additional forecast uncertainty; quantifying this propagation of meteorological forecast error is left for future work.

## 6. Conclusions

This study takes the Chushandian Irrigation Area in the upper Huaihe River as the research region. To address the seasonal non-stationarity of monthly runoff series at the Changtaiguan Hydrological Station, a Period Adaptive Attention Informer (PA-Informer) model is proposed. Through autocorrelation analysis, K-means clustering, and dynamic time warping, the hydrological year is divided into a dry period (November–February), a transition period (March–June and October), and a wet period (July–September), and the optimal lag order for each period is determined. Based on monthly data from 1960 to 2020, multiple comparative experiments are conducted, leading to the following main conclusions:

- (1) The PA-Informer model achieves an  $R^2$  of 0.957, an NSE of 0.948, an RMSE of  $0.71 \text{ m}^3/\text{s}$ , and an MAE of  $0.52 \text{ m}^3/\text{s}$  on the test set. These results are consistently better than those of XGBoost, LightGBM, Random Forest, Base Informer, SML, Indep Informer, TL Informer, and MT Informer, verifying the effectiveness of the period adaptive attention mechanism for monthly runoff simulation.
- (2) SHAP factor contribution analysis shows that the dominant driving factors of runoff differ significantly across hydrological periods. In the dry period, runoff is mainly driven by temperature and antecedent runoff (lag 2). In the transition period, it is co-controlled by precipitation and long lag runoff (lag 3). In the wet period, precipitation plays the dominant role, with temperature playing a suppressive role. The precipitation–temperature interaction effect exhibits a stage-wise evolution pattern: weak coupling in the dry period, synergistic driving in the transition period, and strong antagonism in the wet period.
- (3) Sensitivity analysis shows that the model responds smoothly and orderly to perturbations of each input factor. In the dry period, the model is most sensitive to temperature; in the transition period, it is most sensitive to precipitation; and in the wet period, it shows strong positive sensitivity to precipitation and negative sensitivity to temperature. The model exhibits good anti-interference ability and output robustness across different hydrological periods.

In summary, the PA-Informer model effectively integrates the physical laws of hydrological periodicity with a deep learning architecture, substantially improving the accuracy of monthly runoff simulation and revealing the period-based driving mechanisms of runoff formation. This study provides a scientific basis for water resources management and flood–drought prevention in irrigation areas and can also serve as a methodological reference for similar non-stationary hydrological process predictions.

**Author Contributions:** Y.R. contributed to conceptualization, methodology, software, and visualization. H.Y. was responsible for writing—original draft, investigation, and formal analysis. Q.T. contributed to methodology, supervision, project administration, and funding acquisition. Y.T. contributed to validation, resources, supervision, and writing—review and editing. L.G. contributed

to funding acquisition, data curation, and investigation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Key Technologies and Applications for Whole-Process Refined Regulation of Water Resources in Irrigation Districts Based on Digital Twin (grant no. 251111210700) and the Key Technologies for Multi-Valve Joint Regulation in Long-Distance Water Diversion Projects (grant no. 254200510037).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from a third party. The data are not publicly available due to privacy restrictions.

**Conflicts of Interest:** Authors Yunlong Ran and Lei Guo were employed by the company Henan Water Valley Innovation Technology Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that this study received funding from the Key Technologies and Applications for Whole-Process Refined Regulation of Water Resources in Irrigation Districts Based on Digital Twin (Grant No. 251111210700) and the Key Technologies for Multi-Valve Joint Regulation in Long-Distance Water Diversion Projects (Grant No. 254200510037). The funders had the following involvement with the study: no role in the study design, data collection, analysis, interpretation, writing, or decision to publish.

## References

- Adamowski, J.; Sun, K. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J. Hydrol.* **2010**, *390*, 85–91. [\[CrossRef\]](#)
- Wu, Y.; Wang, Q.; Li, G.; Li, J. Data-driven runoff forecasting for Minjiang River: A case study. *Water Supply* **2020**, *20*, 2284–2295. [\[CrossRef\]](#)
- Ghanim, A.A.J.; Beddu, S.; Abd Manan, T.S.B.; Al Yami, S.H.; Irfan, M.; Mursal, S.N.F.; Mohd Kamal, N.L.; Mohamad, D.; Machmudah, A.; Yavari, S.; et al. Prediction of Runoff in Watersheds Located within Data-Scarce Regions. *Sustainability* **2022**, *14*, 7986. [\[CrossRef\]](#)
- Hublart, P.; Ruelland, D.; García De Cortázar Aauri, I.; Ibacache, A. Reliability of a conceptual hydrological model in a semi-arid Andean catchment facing water-use changes. *Proc. Int. Assoc. Hydrol. Sci.* **2015**, *371*, 203–209. [\[CrossRef\]](#)
- Yuan, X.; Chen, C.; Lei, X.; Yuan, Y.; Muhammad Adnan, R. Monthly runoff forecasting based on LSTM–ALO model. *Stoch. Env. Res. Risk Assess.* **2018**, *32*, 2199–2212. [\[CrossRef\]](#)
- Kalteh, A.M. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput. Geosci.* **2013**, *54*, 0098–3004. [\[CrossRef\]](#)
- Tajiki, M.; Schoups, G.; Hendricks Franssen, H.J.; Najafinejad, A.; Bahremand, A. Recursive Bayesian estimation of conceptual rainfall-runoff model errors in real-time prediction of streamflow. *Water Resour. Res.* **2020**, *56*, e2019WR025237. [\[CrossRef\]](#)
- Bian, L.; Qin, X.; Zhang, C.; Guo, P.; Wu, H. Application, interpretability and prediction of machine learning method combined with LSTM and LightGBM—a case study for runoff simulation in an arid area. *J. Hydrol.* **2023**, *625*, 0022–1694. [\[CrossRef\]](#)
- Ghazvinian, H.; Samii, A.; Karami, H.; Khaniki, M.L. Machine learning-based modeling and analysis of surface cover impacts on hydrological parameters influencing urban runoff. *Results Eng.* **2026**, *29*, 109580. [\[CrossRef\]](#)
- Szczepanek, R. Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost. *Hydrology* **2022**, *9*, 226. [\[CrossRef\]](#)
- Qiao, X.; Peng, T.; Sun, N.; Zhang, C.; Liu, Q.; Zhang, Y.; Wang, Y.; Nazir, M.S. Metaheuristic evolutionary deep learning model based on temporal convolutional network, improved aquila optimizer and random forest for rainfall-runoff simulation and multi-step runoff prediction. *Expert. Syst. Appl.* **2023**, *229*, 0957–4174. [\[CrossRef\]](#)
- Cui, Z.; Qing, X.; Chai, H.; Yang, S.; Zhu, Y.; Wang, F. Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis. *J. Hydrol.* **2021**, *603*, 0022–1694. [\[CrossRef\]](#)
- Abdoulhalik, A.; Ahmed, A.A. A Comparative Analysis of Advanced Machine Learning Techniques for River Streamflow Time-Series Forecasting. *Sustainability* **2024**, *16*, 4005. [\[CrossRef\]](#)
- Huang, P.-C. Combination of dynamic TOPMODEL and machine learning techniques to improve runoff prediction. *J. Flood Risk Manag.* **2025**, *18*, e13050. [\[CrossRef\]](#)
- Jiang, J.; Gao, Y.; Chen, J.; Huang, J.; Yu, J.; Jiang, C.; Liu, J.; Xue, A. Effects of Multi-Step-Ahead Prediction Strategies on LSTM-Based Runoff Prediction. *Water Resour. Manag.* **2025**, *39*, 8103–8116. [\[CrossRef\]](#)

16. Xiang, Z.; Yan, J.; Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **2020**, *56*, e2019WR025326. [\[CrossRef\]](#)
17. Kratzert, F.; Klotz, D.; Brenner, C.; Schulz, K.; Herrnegger, M. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022. [\[CrossRef\]](#)
18. Yao, Z.; Wang, Z.; Wang, D.; Wu, J.; Chen, L. An ensemble CNN-LSTM and GRU adaptive weighting model based improved sparrow search algorithm for predicting runoff using historical meteorological and runoff data as input. *J. Hydrol.* **2023**, *625*, 0022–1694. [\[CrossRef\]](#)
19. Yang, X.; Chen, Z.; Qin, M. Monthly Runoff Prediction Via Mode Decomposition-Recombination Technique. *Water Resour. Manag.* **2024**, *38*, 269–286. [\[CrossRef\]](#)
20. Song, C.; Yao, L.; Hua, C.; Ni, Q. A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory. *J. Hydrol.* **2021**, *603*, 0022–1694. [\[CrossRef\]](#)
21. He, F.; Wan, Q.; Wang, Y.; Wu, J.; Zhang, X.; Feng, Y. Daily Runoff Prediction with a Seasonal Decomposition-Based Deep GRU Method. *Water* **2024**, *16*, 618. [\[CrossRef\]](#)
22. Bao, Z.; Du, J.; Zhang, W.; Wang, J.; Qiu, T.; Cao, Y. A Transformer Model-Based Approach to Bearing Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2021**, *17*, 5678–5686.
23. Wang, Y. Deepening Intelligent Microgrid Management: A Study on Improving Load Forecasting Accuracy Based on Informer Models. *IEEE Trans. Smart Grid* **2024**, *15*, 1234–1245.
24. Yang, Z.; Zhang, Q.; Chang, W.; Xiao, P.; Li, M. EGFormer: An Enhanced Transformer Model with Efficient Attention Mechanism for Traffic Flow Forecasting. *Vehicles* **2024**, *6*, 120–139. [\[CrossRef\]](#)
25. Yu, H.; Ma, Y.; Hu, A.; Wang, Y.; Tian, H.; Dong, L.; Zhu, W. Daily Runoff Prediction Method Based on Secondary Decomposition and the GTO-Informer-GRU Model. *Water* **2025**, *17*, 2775. [\[CrossRef\]](#)
26. Wang, S.; Chen, Y.; Ahmed, M. EWT\_Informer: A novel satellite-derived rainfall-runoff model based on informer. *J. Hydroinformatics* **2024**, *26*, 88–106. [\[CrossRef\]](#)
27. Wang, W.C.; Ren, H.Z.; Li, Z.; Zhao, Y.-W.; Hu, X.-X.; Gu, M. Improving the accuracy of daily runoff prediction using informer with black kite algorithm, variational mode decomposition, and error correction strategy. *Stoch. Env. Res. Risk Assess.* **2025**, *39*, 1757–1785. [\[CrossRef\]](#)
28. Broxton, P.D.; van Leeuwen, J.D.; Svoma, B.M.; Walter, J.; Biederman, J.A. Subseasonal to Seasonal Streamflow Forecasting in a Semiarid Watershed. *JAWRA J. Am. Water Resour. Assoc.* **2023**, *59*, 1493–1510. [\[CrossRef\]](#)
29. Falck, A.S.; Tomasella, J.; Diniz, F.L.; Maggioni, V. Assessment of subseasonal streamflow predictions in a tropical basin. *J. Hydrol.* **2025**, *132488*, 0022–1694. [\[CrossRef\]](#)
30. Tran, V.H.; Hoang, V.H.; Tran, Q.M. Interpretable machine learning with Bayesian optimization for bond strength prediction of steel reinforcement in geopolymers concrete. *PLoS ONE* **2026**, *21*, e0352645. [\[CrossRef\]](#)
31. Hoang, V.H.; Tran, M.Q.; Ngo, V.T. Machine learning-based prediction of the axial load capacity of UHPC strengthened reinforced concrete columns: A comparative analysis. *PLoS ONE* **2026**, *21*, e0338120. [\[CrossRef\]](#)
32. Zhang, X.; Yang, K.; Zheng, L. Transformer Fault Diagnosis Method Based on TimesNet and Informer. *Actuators* **2024**, *13*, 74. [\[CrossRef\]](#)
33. Ghobadi, F.; Yaseen, Z.M.; Kang, D. Long-term streamflow forecasting in data-scarce regions: Insightful investigation for leveraging satellite-derived data, Informer architecture, and concurrent fine-tuning transfer learning. *J. Hydrol.* **2024**, *631*, 0022–1694. [\[CrossRef\]](#)
34. Zhao, L.T.; Li, Y.; Chen, X.H.; Sun, L.Y.; Xue, Z.Y. MFTM-Informer: A multi-step prediction model based on multivariate fuzzy trend matching and Informer. *Inf. Sci.* **2024**, *681*, 0020–0255. [\[CrossRef\]](#)
35. Huang, X.; Jiang, A. Wind Power Generation Forecast Based on Multi-Step Informer Network. *Energies* **2022**, *15*, 6642. [\[CrossRef\]](#)
36. Shang, J.; Zhao, B.; Hua, H.; Wei, J.; Qin, G.; Chen, G. Application of Informer Model Based on SPEI for Drought Forecasting. *Atmosphere* **2023**, *14*, 951. [\[CrossRef\]](#)
37. Zhang, Y.; Li, J.; Liu, D.; Chen, G.; Dou, J. D-P-Transformer: A Distilling and Probsparse Self-Attention Rockburst Prediction Method. *Energies* **2022**, *15*, 3959. [\[CrossRef\]](#)
38. Su, X.; Zhang, Y.; Zhang, Y.; Liu, J.; Xu, M.; Jia, P. An enhanced multilayer Res-Informer with Savitzky-Golay filter for predicting mixed CO and NOx emissions in gas turbines. *Chemom. Intell. Lab. Syst.* **2025**, *261*, 0169–7439. [\[CrossRef\]](#)
39. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [\[CrossRef\]](#)
40. Kreutzer, L.T.; Gillen, E.; Briegal, J.T.; Queloz, D. S-ACF: A selective estimator for the autocorrelation function of irregularly sampled time series. *Mon. Not. R. Astron. Soc.* **2023**, *522*, 5049–5061. [\[CrossRef\]](#)
41. Reimer, A.S.; Hussey, G.C.; Dueck, S.R. On the statistics of SuperDARN autocorrelation function estimates. *Radio. Sci.* **2016**, *51*, 690–703. [\[CrossRef\]](#)

42. Sutanto, S.J.; Van Lanen, H.A.J. Catchment memory explains hydrological drought forecast performance. *Sci. Rep.* **2022**, *12*, 2689. [[CrossRef](#)]
43. Wang, W.C.; Gu, M.; Hong, Y.H.; Hu, X.X.; Zang, H.F.; Chen, X.N.; Jin, Y.G. SMGformer: Integrating STL and multi-head self-attention in deep learning model for multi-step runoff forecasting. *Sci. Rep.* **2024**, *14*, 23550. [[CrossRef](#)]
44. Tian, J.K.; Zhang, J.; Tong, J.J.; He, H.; Gu, R.; Shang, F. Dominant factor analysis and threshold inflection point determination in deep learning-based SWAT-LSTM training models with SHAP interpretability analysis. *Water* **2026**, *18*, 960. [[CrossRef](#)]
45. Lu, Y.; Yao, H.Y.; Ma, J.Y.; Tang, X.; Chen, J.; Cheng, M. Inter-comparison and mechanistic interpretation of deep learning models for turbidity prediction in rural areas. *J. Hydrol.* **2025**, *662*, 134004. [[CrossRef](#)]
46. Razavi, S.; Sheikholeslami, R.; Gupta, H.V.; Haghnegahdar, A. VARS-TOOL: A toolbox for comprehensive, efficient and robust sensitivity and uncertainty analysis. *Environ. Model. Softw.* **2019**, *112*, 95–107. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.